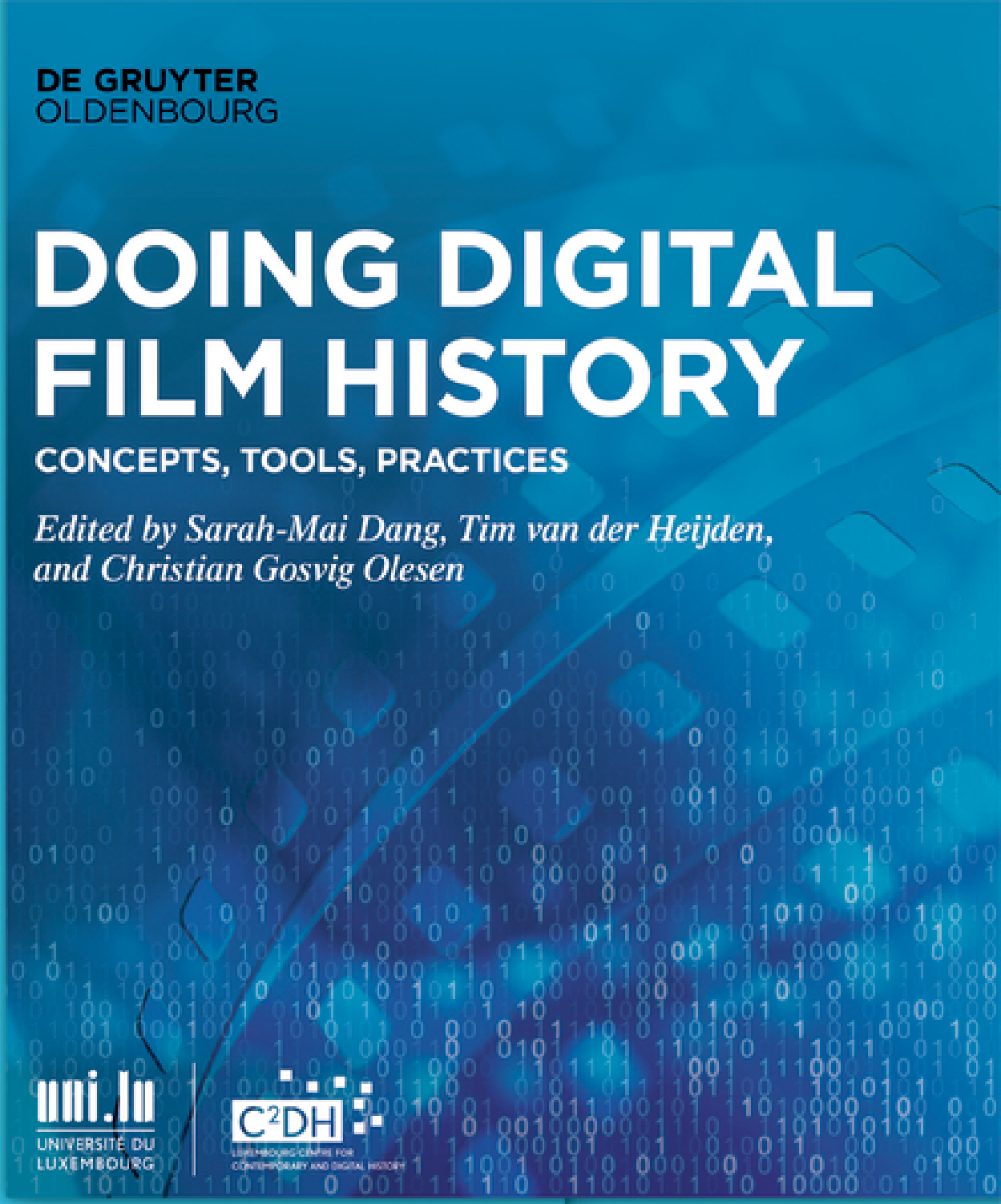
DE GRUYTER
OLDENBOURG

DOING DIGITAL FILM HISTORY

CONCEPTS, TOOLS, PRACTICES

Edited by Sarah-Mai Dang, Tim van der Heijden, and Christian Gosvig Olesen



STUDIES IN DIGITAL HISTORY AND HERMENEUTICS

DE
G

Doing Digital Film History

Studies in Digital History and Hermeneutics

Edited by
Andreas Fickers, Valérie Schafer, Sean Takats,
and Gerben Zaagsma

Volume 11

Doing Digital Film History

Concepts, Tools, Practices

Edited by
Sarah-Mai Dang, Tim van der Heijden,
and Christian Gosvig Olesen

DE GRUYTER
OLDENBOURG

This publication is the result of the international research network “New Directions in Film Historiography: Digital Tools and Methods in Film and Media Studies” (2019–2023), funded by the DFG – Deutsche Forschungsgemeinschaft (German Research Foundation), project number 423266924.

Published with the kind support of the DFG – Deutsche Forschungsgemeinschaft (German Research Foundation).

ISBN 978-3-11-108149-6
e-ISBN (PDF) 978-3-11-108248-6
e-ISBN (EPUB) 978-3-11-108305-6
ISSN 2629-4540
DOI https://doi.org/10.1515/9783111082486

Library of Congress Control Number: 2024942637

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the internet at http://dnb.dnb.de.

Cover image: Antonio Bordunovi/iStock/Getty Images Plus; gonin/iStock/Getty Images Plus
Typesetting: Integra Software Services Pvt. Ltd.
Printing and binding: CPI books GmbH, Leck

www.degruyter.com
Questions about General Product Safety Regulation:
productsafety@degruyterbrill.com

Acknowledgments

The present edited volume is the result of the international research network "New Directions in Film Historiography: Digital Tools and Methods in Film and Media Studies" (2019–2023), funded by the DFG – Deutsche Forschungsgemeinschaft (German Research Foundation), project number 423266924, and its closing conference "Doing Digital Film History," held at Philipps-Universität Marburg in November 2022. For funding the network's research, and in particular the several international meetings leading up to the closing conference and the present volume, we wish to first express our deep gratitude to DFG. We would also like to thank the anonymous reviewer of the grant proposal who recognized the value of the network for reflecting and developing computational methods and analysis in the context of film and media historiography.

In addition to the authors whose work is represented in this volume, all of whom contributed immensely to the network's inspiring activities and discussions, we would like to extend our deep gratitude to network members, Feng-Mei Heberer, Lies Van de Vijver, Niels-Oliver Walkowski; collaborators, Federica Bardelli, Camille-Blot-Wellens, Nadieh Bremer, Johanna Drucker, Andreas Fickers, Dirk Förstner, Jörg Frieß, Oliver Hanley, Julia Lück, Michelle Mantsio, Peter Niedermüller, Andy O'Dwyer, Petra Rauschenbach, Thelma Ross, Martina Schories, Deb Verhoeven, Ingo Zechner; guest presenters, Maggie Hennefeld, Laura Horak, Mila Oiva, Frauke Pirk, Elif Rongen-Kaynakçi; panel chairs, Clara Auclair, Isadora Campregher Paiva, Bregt Lameris, Kai Matuszkiewicz, Nanne van Noord; student assistants, Anne Hart, Judith Kempnich; and the organizing team of the conference, Marlene Leonie Biebricher, Pauline Junginger, Yolanda Well Rull, who all offered invaluable food for thought and support along the way. Our thanks goes also to the attendees at the final conference for their generosity and encouragement of our endeavors during the discussions. The richness of these discussions inspired us to develop and broaden the scope of this volume in productive ways.

We sincerely thank Andreas Fickers, director of the Luxembourg Centre for Contemporary and Digital History (C^2DH) at the University of Luxembourg, for accepting the book into the Studies in Digital History and Hermeneutics series and for his support. We thank our publisher, De Gruyter Oldenbourg, in particular Rabea Rittgerodt, Verena Deutsch, and Annika Padoan, for their guidance on the production of this volume, as well as our reviewers for helping us improve it. A special acknowledgement goes to Elizabeth Rankin, for rigorously and patiently reviewing the entire manuscript in the editing process with an impeccable attention to detail.

https://doi.org/10.1515/9783111082486-202

Finally, we would like to express our gratitude to the contributors to this volume, all of whom generously shared their insights and creatively responded to the potential and challenges of doing digital film history.

Sarah-Mai Dang
Tim van der Heijden
Christian Gosvig Olesen

Contents

Sarah-Mai Dang, Tim van der Heijden, and Christian Gosvig Olesen

Doing Digital Film History: An Introduction

Introduction

How has the digital turn shaped the practices of "doing film history" in both research and teaching? That is the main question of this book. While computational approaches have been used by film historians since the 1960s and 1970s, the arrival and use of digital tools and methods in recent decades has fundamentally changed the ways we search, analyze, interpret, present, and so think and write about film history. This is at the levels of both "close" and "distant" – or "scalable" – reading and viewing, and quantitative and qualitative methodologies, as well as those approaches in between. And it extends from digital film archival practices and data-driven search in both small and large film historical collections to the visualization and "distant viewing" of film historical materials, as well as their dissemination on digital platforms. This book addresses these fundamental issues for researching and teaching film history in the digital age.

The chapters are the outcome of the international conference "Doing Digital Film History," organized at Philipps-Universität Marburg in November 2022.[1] The conference was the closing event of the international research network "New Directions in Film Historiography: Digital Tools and Methods in Film and Media Studies" (2019–2023),[2] funded by the German Research Foundation (DFG). The aim of the network has been to collaboratively explore how digital technologies shape our understanding of film and cinema history from a media studies perspective. In order to analyze the epistemic, conceptual, and methodological frameworks of digital film historiography, the network members brought together theory and practice, while drawing on approaches not only from film and media studies but also from various other disciplines, including history, library and archive studies, critical data studies, and computer science. By means of this interdisciplinary approach, it addressed the challenges and possibilities of digital technologies for the field of film history, which we believe can best be understood in its far-reaching dimensions when both applied and critically reflected upon. The interdisciplinary

1 See the "Doing Digital Film History" conference website: https://www.uni-marburg.de/en/fb09/institutes/media-studies/research/research-projects/digitalfilmhistoriography/events/conference_program, accessed November 5, 2023.

2 See the network website: https://www.uni-marburg.de/en/fb09/institutes/media-studies/research/research-projects/digitalfilmhistoriography, accessed November 5, 2023.

https://doi.org/10.1515/9783111082486-001

approach of the network and questions addressed at the conference about digital film historiography are central to this book.

While film historians have increasingly embraced the new possibilities brought by digital tools and technologies, their practical, epistemological, and methodological implications need further exploration. For example, what are the consequences of the digitization of film historical sources for the study of film and media history? What new questions can be asked? How do digital tools afford different or complementary methodological approaches for analyzing and interpreting film historical data, as well as new forms of presenting and sharing/publishing film historical scholarship? What are the effects of the digital turn on teaching film and media history? By specifically focusing on the concepts, tools, and practices and their interrelations in digital film historical research and teaching, the book addresses the following questions:

- What *concepts* have been (re)introduced, and how have they established new ways of thinking about film history?
- What new *tools* and *methods* have emerged in film historiography, and how have they shaped different ways of searching, analyzing, visualizing, and interpreting films and other film historical sources at scale?
- What *practices* have emerged in digital film research and teaching, and what are their methodological and epistemological implications for our historiographical narration and historical understanding?

By addressing these central issues, this volume aims to contribute to the discussion on doing digital film history.

Positioning the Book in the Field

Digital film historiography is a relatively new and highly dynamic emerging field. While recent studies in film history, media studies, and digital humanities have provided valuable insights and approaches, this book systematically reflects on the implications of the digital turn for doing film historical research and teaching. Our book builds on many foundational publications that address crucial methodological, theoretical, and practical issues in this field. While it reflects on various topics and questions discussed in previous monographs and collections of academic essays that focus on the relation between digital humanities and media

studies[3] or media history[4] more generally, it shifts the focus to research that explicitly deals with film history or film historical data. Aligning with studies on the intersection of film studies and digital humanities, with a focus on moving image analysis[5] or audiovisual data,[6] for instance, our book takes into account the need for approaches that combine quantitative and qualitative methodologies. As previous film historical studies show, statistical analysis of empirical data and data visualization offer a promising approach in this regard.[7]

As many scholars have demonstrated, a plurality of approaches is needed to understand the transdisciplinary complexity of practices in film history.[8] The goal of this volume is to draw attention to the specificities of digital developments in this field, including both the advances and the challenges, while critically self-reflecting on film historiographical practices and premises. Providing additional critical case studies and methodological explorations, this book complements current analysis on the impact of digitization of archival artifacts in film heritage institutions, such as films, magazines, or newspapers and other historical records,

3 Jentery Sayers, ed., *The Routledge Companion to Media Studies and Digital Humanities* (New York: Routledge, 2018).

4 Charles R. Acland and Eric Hoyt, eds., *The Arclight Guidebook to Media History and the Digital Humanities* (Falmer: REFRAME Books, 2016), accessed April 3, 2024, http://projectarclight.org/book/.

5 See the DHQ special issue "Digital Humanities & Film Studies: Analyzing the Modalities of Moving Images," ed. Manuel Burghardt, Adelheid Heftberger, Johannes Pause, Niels-Oliver Walkowski, Matthias Zeppelzauer; Manuel Burghardt et al., "Film and Video Analysis in the Digital Humanities – An Interdisciplinary Dialog," *Digital Humanities Quarterly* 14, no. 4 (2020), accessed April 3, 2024, http://digitalhumanities.org/dhq/vol/14/4/000532/000532.html.

6 See the DHQ special issue "AudioVisual Data in DH," ed. Taylor Arnold, Jasmijn Van Gorp, Stefania Scagliola, and Lauren Tilton, "Introduction: Special Issue on AudioVisual Data in DH," *Digital Humanities Quarterly* 15, no. 1 (2021), accessed April 3, 2024, https://www.digitalhumanities.org/dhq/vol/15/1/000541/000541.html.

7 See, for instance, Adelheid Heftberger, *Digital Humanities and Film Studies: Visualising Dziga Vertov's Work*, Quantitative Methods in the Humanities and Social Sciences (Cham: Springer International Publishing, 2018), https://doi.org/10.1007/978-3-030-02864-0; Yuri Tsivian, "Cinemetrics: Part of the Humanities' Cyberinfrastructure," in *Digital Tools in Media Studies: Analysis and Review, An Overview*, ed. Michael Ross, Manfred Grauer, and Bernd Freisleben (Bielefeld: Transcript, 2009), 93–100. See also the Cinemetrics website: https://cinemetrics.uchicago.edu/, accessed April 3, 2024.

8 Malte Hagener and Yvonne Zimmermann, eds., *How Film Histories Were Made: Materials, Methods, Discourses* (Amsterdam: Amsterdam University Press, 2023); Christian Gosvig Olesen, "Digital Film Historiography – A Bibliography," *Film History in the Making* (blog), August 19, 2020, accessed April 3, 2024, https://filmhistoryinthemaking.com/digital-film-historiography-a-bibliography/.

as prominently done by the Media History Digital Library,[9] and the growing digitalization of film historiography.[10]

Legacies of Digital Film History

While the present volume attends to the emergence of digital tools in film historical research primarily in the past couple of decades, and highlights the newness of the approaches discussed, it is equally invested in acknowledging the legacies embedded in older historical studies and the tools and the data that they process. Too often, we feel, has the emergence of digital tools, archives, and data been depicted as a dramatic rupture from previous practices, when in many cases we may learn from antecedents in film historical research that tend to be neglected. For instance, recent scholarship has highlighted how deeply embedded combined approaches of qualitative and quantitative film analysis are in decades-long experimentation with the visualization of film structures, which developed at an intersection of computational linguistics, literary studies, and film stylometry.[11] Likewise, we tend to forget today how far ahead (some) film archivists were thinking already in the 1980s, in envisioning film collections as datafied entities that could make use of OCR, video technologies, and innovative data formats such as MARC (Machine Readable Cataloging) in novel ways,[12] to offer video-based access and facilitate new ways of retrieving and linking collection items. All are features that are key to the projects discussed in this volume.

By placing digital film historical scholarship in a broader historical frame, we may, in addition to highlighting digital tools' multiple origin points, also achieve a better understanding of the situatedness of the data we work with in digital film history. In particular, we seek to highlight how the work of processing data, which we encounter in archives and research data repositories, and subsequently prepare

9 See the Media History Project website: https://mediahistoryproject.org/, accessed November 5, 2023.

10 Christian Gosvig Olesen, *Visualizing Film History: Film Archives and Digital Scholarship* (Bloomington: Indiana University Press, 2025); Sarah-Mai Dang, "Digital Tools & Big Data: Zu gegenwärtigen Herausforderungen für die Film-und Medienwissenschaft am Beispiel der feministischen Filmgeschichtsschreibung," *MEDIENwissenschaft: Rezensionen / Reviews* 35, nos. 2–3 (2018): 142–156, http://dx.doi.org/10.25969/mediarep/4393.

11 Heftberger, *Digital Humanities and Film Studies*.

12 Roger Smither, *Evaluating Computer Cataloguing Systems: A Guide for Film Archivists* (Brussels: FIAF, 1988); Roger Smither, "Formats and Standards: A Film Archive Perspective on Exchanging Computerized Data," *The American Archivist* 50, no. 3 (1987): 324–337.

for use in digital projects through cleaning and curating, invites us to reflect on how historical interpretation processes condition our work today, and inevitably involve a negotiation between past and present epistemological outlooks.[13]

Digital Hermeneutics in Doing Digital Film History

On a conceptual level, this book draws on the framework of "digital hermeneutics," defined as a framework for the critical and self-reflexive use of digital technologies for doing historical research in the digital age.[14] In the edited volume *Digital History and Hermeneutics*, Andreas Fickers et al. distinguish between four research practices and their related types of criticism in digital historical scholarship (see Figure 1):[15]

1. *Algorithmic criticism*: how digital search practices have been shaped by the logics and strategies of algorithms.
2. *Digital source criticism*: how processes of digitization, data management, and curation have transformed historical "sources" to "data" structured in databases.
3. *Tool criticism*: how processes of analysis have been shaped by digital tools and technologies.
4. *Interface and simulation criticism*: how practices of visualization, interpretation, and publication of data have been enabled through digital interfaces and simulation.

How can we apply the digital hermeneutics framework to digital film historiography? Each of the four criticisms reflects a certain practice and related set of skills that have become part of digital film historical research and teaching. Practices of search, for instance, relate to the use of filmographic data in digital film archives and collections, including their curatorial challenges. Practices of digitization and data management relate to the use of film historical sources and databases in digital film research and teaching. Practices of analysis address the use of digital tools and technologies for the analysis of historical films and their metadata. Practices of

13 Johanna Drucker, *The Digital Humanities Coursebook: An Introduction to Digital Methods for Research and Scholarship*, 1st ed. (Abingdon, Oxon; New York: Routledge, 2021).

14 See https://www.c2dh.uni.lu/events/digital-hermeneutics-history-theory-and-practice, accessed November 5, 2023.

15 Andreas Fickers, Juliane Tatarinov, and Tim van der Heijden, "Digital History and Hermeneutics – Between Theory and Practice: An Introduction," in *Digital History and Hermeneutics*, ed. Andreas Fickers and Juliane Tatarinov (Berlin: De Gruyter, 2022), 1–20, https://doi.org/10.1515/9783110723991-001.

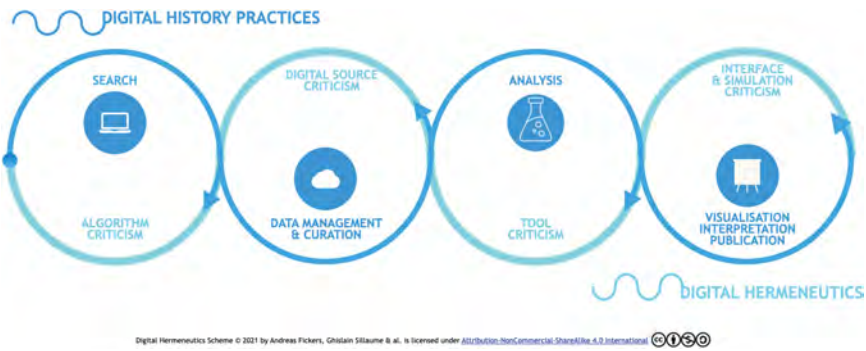


Figure 1: Digital hermeneutics scheme © Andreas Fickers, Ghislain Sillaume et al., 2021.[16]

data visualization relate to the way film historical data have been presented in visual forms, such as graphs and networks, which enables new insights and forms of interpretation of, for instance, the distribution and circulation of historical films.

Organization of This Book

By foregrounding these *practices* of doing digital film history in both research and teaching, we did not only take inspiration from the digital hermeneutics framework as a conceptual lens to reflect on the impact of "the digital" on film historical scholarship. We also used the four distinct practices and their digital criticism as a way of structuring the chapters of this book into four parts: I. SEARCH, II. DATA, III. ANALYSIS, and IV. VISUALIZATION.

Naturally, this differentiation is only a heuristic one. In reality, both in research projects and teaching implementations, these four practices are highly interrelated. Furthermore, as the visualization of the digital hermeneutics framework also exemplifies, they are based on an iterative process in which they inform rather than exclude each other. Nevertheless, we felt that clustering the chapters based on their shared practices and related criticisms would make it easier for the reader to see how they contribute to the imperative question of how the digital has transformed the practices of doing film historical scholarship. Below, we offer an overview of each section's content.

16 Cited in Fickers et al., "Digital History and Hermeneutics," 10.

Prologue

In the prologue, entitled **The DH Dilemma: Knowing More & Knowing for Sure vs. Never Knowing At All**, film scholar and pioneer of feminist film scholarship Jane M. Gaines reflects broadly on various digital humanities projects in the United States and the question of what new forms of doing digital film history have emerged in the past few decades. She reviews the current debates around digital humanities from the perspective of the methodological transition from paper documents and photochemical prints to the digitization of materials and the computerization of historical investigation. She takes up the question not only of knowledge "acquisition" but of assumptions about knowledge as produced by machines.

I Search

The section on search practices in digital film research and education opens with the chapter **Finding Female Film Editors in Wikidata: How to Query and Visualize Filmographic Records**, written by film archivist and scholar Adelheid Heftberger and data analyst Paul Duchesne. It examines the underappreciated contributions of female film workers in the film production industry, focusing on editors. Drawing from recent research and initiatives like the BFI Filmography project and Wikidata, the authors advocate for a more inclusive approach to film historiography, considering the episodic nature of women's careers, and gender discrimination's enduring impact. Specifically, they analyze the representations of female film editors on Wikidata, aiming to uncover patterns and addressing challenges in the process. They argue that, while we can use Wikidata as a valuable source, caution is advised against relying solely on Wikidata for comprehensive filmographic data. This chapter contributes to ongoing conversations about gender representation in film and evolving methodologies in film historiography.

The next chapter, **The Digitization of Silent Films and the Teaching of Film Historiography: Entanglements and Opportunities**, written by film historian Casper Tybjerg and his students Jonatan Bruun Borring and Luan Nhu Vu, discusses the opportunities provided by digitized film collections and related archival documents. Focusing on the website stumfilm.dk, which presents the national collection of Danish silent cinema, the authors reflect on the relation between the research material and the types of research questions it can address. Drawing on a teaching experiment, they elaborate on how historical research is informed by specific archival setups as well as by the particular scholarly context (e.g., disciplinary traditions, archival access, technological resources, national institutions of learn-

ing). In doing so, they refer and draw upon the "entangled film history" approach. The results suggest that automated quantitative style analysis is feasible at scale, allowing a much larger number of films to be investigated. The authors also show that integrating research with teaching can lead to useful results.

The chapter **Collecting Data and Connecting Traces: Researching and Modeling Sources on Doña Francisquita (S 1934)**, written by film scholars Imme Klages and Fabian Kling, critically reflects on the idea of collections as data. The authors argue that the emergence of online archives and public databases allows us to research German film migration between 1933 and 1945 in different ways. Presenting a case study on Doña Francisquita, a film made in 1934 during the Spanish Civil War, they discuss the challenges curator-archivists and film historians face when dealing with the filmographic information of various digital sources. Instead of analyzing large datasets, the authors focus on gathering information from disparate and fragmented sources to confirm and verify historical information. The chapter lays down a critical framework for researching and modeling information about a single film historical event, person, or production with online sources.

In their chapter **Teaching Small-Gauge Formats with Digital Methods**, film scholar Nicole Braida and digital humanities scholar Frauke Pirk offer insights into the challenges of teaching digital methods to undergraduates in film history. Drawing on a course on small-gauge formats – especially Super 8 film – in the context of amateur and non-theatrical films, the authors present a tentative overview of what they consider relevant digital humanities practices and concepts. By sharing teaching experiences with data modeling and management, they reflect on data literacy as a key competency for students.

II Data

The section on archival digitization practices, data management, and the construction and curation of filmographic data opens with the chapter **Managing the Past: Research Data and Film History**, in which film scholar Sarah-Mai Dang looks at research data management (RDM) and its impact on digital film history. Drawing on critical approaches to data-driven projects and infrastructures, as well as her own experience with film historical databases, Dang demonstrates how RDM practices reflect the intellectual conventions and institutional frameworks in which specific concepts of film and gender are inscribed. She also shows how, conversely, our understanding of film culture is affected by RDM. Comparing two examples, the Women Film Pioneers Project and the DFF – Deutsches Filminstitut & Filmmuseum, she lays out why we need to critically scrutinize how

data is collected, organized, and preserved in order to conduct meaningful research. In doing so, the author outlines a framework that allows us to systematically examine RDM in the context of film history.

In the next chapter, **(Re)Visioning Women's Film History: The Women Film Pioneers Project and Digital Curatorial-Editorial Labor**, film scholar Kate Saccone reflects on the development of the Women Film Pioneers Project (WFPP) at its ten-year anniversary. Saccone analyzes the editorial labor in relation to both the iterative nature of feminist film historiography and the mutability of digital scholarship and online publishing. Saccone argues that WFPP relies on and makes visible a digital curatorial-editorial practice that she calls "(re)visioning," which draws on the open-ended processes of creating visibility at the heart of feminist film historiography and the practice of versioning at the heart of digital humanities. Although "(re)visioning" emerges from WFPP's particular subject matter, history, and current workflows, it reflects the many conceptual and practical challenges to notions that this work is ever "done" and "complete" in the digital era. As such, it constitutes a critical-feminist perspective on any digital scholarly editorial labor that is open to continued historiographic movement and its ensuing management online.

The chapter **Data Cleaning and Diversity in Digital Film Historiography** by film scholars Alexandra Schneider and Yvonne Zimmermann is a critical intervention on the preparation of data that is made available for doing digital film history. Drawing on Johanna Drucker's insights, the authors address the tension between the initial interpretative aspirations of projects and the eventual need for interoperability, often leading to the sacrifice of elaborate data constructions for expediency. Through a discussion of various projects, including the Straschek Collection and the Importing Asta Nielsen Database, the chapter highlights the challenges of standardizing and cleaning data while preserving its diversity and locality. The authors call for a rethinking of ethical standards and practices in data handling, emphasizing the need for inclusive and context-aware approaches in computational humanities research.

The section closes with the chapter **Critically Curating Data in Cultural Heritage Collections**, in which EU Projects Coordinator Kerstin Herlt and EFG Project Manager Julia Welter argue that the ongoing discussion about decolonizing archives and museums demonstrates that transnational approaches and curatorial efforts are not a sufficient response to the call of minority communities to increase diversity, equity, and polyvocality in data. Presenting the EU-funded project "DE-BIAS – Detecting and Cur(at)ing Harmful Language in Cultural Heritage Collections," the authors make the case that, while important work has been done on developing common standards, metadata schemas, and vocabularies, there has been little awareness of the need to revise catalog entries and content

descriptions, or resources to do so. From today's perspective, the blind spots and missing narratives are obvious: on the level of metadata, it lies in the use of outdated or discriminatory language for search terms, keywords, and content descriptions. Drawing on a variety of existing cataloging guidelines, the chapter reflects on the proper methodology for taking account of those critical issues.

III Analysis

This section on the analysis of historical films and film historical (meta)data opens with the chapter **Timelines of Scholarly Video Annotation: For a Tool Critical History of Digital Film Historical Scholarship**. In his contribution, media scholar Christian Gosvig Olesen reflects on the development and documentation of the interface of the CLARIAH Media Suite research environment from a historical, tool-critical perspective, paying particular attention to the environment's video annotation tool and its use in film historical research and teaching. This entails two main components: a comparative analysis between previous multimedia scholarship and the Media Suite to understand its timeline functionality as a tool for knowledge production and, subsequently, critical reflection on current historicizations and preservation of digital scholarship, to make the case for a historically informed tool criticism.

The chapter **Distant Viewing the Amateur Film Platform**, written by media historian Tim van der Heijden and digital humanities scholars Taylor Arnold and Lauren Tilton, explores the Amateur Film Platform (2014–2023). This online platform, initiated by the Netherlands Institute for Sound & Vision, hosted a unique collection of more than eight thousand films and videos made by Dutch amateur filmmakers. It featured amateur footage from the early 1900s until the 2010s, recorded on different technological carriers, including 35mm, 16mm, 9.5mm, 8mm, Super 8 film, VHS, MiniDV, and other formats. Utilizing the Distant Viewing approach and toolkit for computationally analyzing large collections of audiovisual materials, the chapter aims to investigate any formal, stylistic, and aesthetic patterns or changes over time in the Amateur Film Platform's collection. The authors argue for a "hybrid heuristics" approach for undertaking data-driven film historical analysis, which combines distant and close viewing with critical reflection on the hermeneutic implications of digital methods and technologies for exploring historical amateur media collections at scale.

The chapter **Pursuing Film History with Digital Images: Towards Visual Literacy in the Age of AI and Social Media**, written by film scholar Franziska Heller and film preservationist and researcher Ulrich Ruedel, is based on a transdisciplinary collaborative project on media history, digital film restoration, and

digital culture. The authors discuss the nature of images when employing film clips or archival film excerpts through online platforms such as YouTube. Questions addressed include the origin of the moving images employed, and the sources this (somehow) digitized material may be derived from. As professors in distinctly different study programs with divergent goals (media studies and audiovisual preservation), both authors practice approaches that lead to specific views on how film history is understood, researched, preserved, and shared, and thus further transferred and communicated. Focusing on the challenge of "doing film history" with images that reach us in digital forms and versions, they propose a transdisciplinary methodology and synergistic approach towards a visual literacy of historical moving images in today's media landscape.

The section closes with the chapter **Managing Tools and Expectations: Dos and Don'ts of Teaching Digital Methods for Film Analysis and Film Historiography** by film scholars Josephine Diecke and Malte Hagener. It explores the challenges and opportunities of integrating digital tools into the teaching of traditional aspects of film studies, including film history, theory, and analysis. It emphasizes the importance of considering digital research methods alongside domain-specific objects of study. Drawing on their co-teaching experience in a course combining film and textual analysis using tools such as VIAN, Voyant, and Arclight, Diecke and Hagener discuss selected dos and don'ts of teaching digital methods. Issues of corpus size, storage requirements, and automated support prompt reflection on digital formats and content negotiation in teaching (with) digital tools. They explore the multifaceted nature of digital film historiography, highlighting the need for interdisciplinary skills to critically engage with digital tools and media. The chapter contributes insights to guide instructors in teaching digital film analysis and historiography.

IV Visualization

The book's final section on the use of visualization tools and methods in digital film historical research and teaching opens with the chapter **Visualization In/As Digital Media Studies** from media scholar Marcus Burkhardt and film scholar Skadi Loist. This chapter examines the role of visualization as a method and mode of research within digital media studies, particularly in the context of digital humanities and digital methods. Emphasizing visualization not just as a means of inquiry but as a critical intervention in the presentation and interpretation of data, the authors adopt a reflexive and experimental approach. Drawing on case studies on film festivals and the evolution of Facebook/Meta's API ecosystem, the chapter highlights how visualization can allow the exploration of large datasets, facilitate iterative analysis, and drive innovative research practices. By employing approaches like Vi-

sual Network Analysis, the authors advocate for a responsive and adaptable research approach, emphasizing the continuous refinement of both data and visualization tools.

The chapter **A Scalable Perspective on Historical Cinema Cultures: Studying Movie Going in Amsterdam (1952–1972) with Digital Data and Tools**, written by cinema historian and digital humanities scholar Julia Noordegraaf, investigates the use of digital data and tools in film historical research, evaluating opportunities and challenges. In particular, it analyzes the epistemological and methodological implications of data-driven media historiography by asking what knowledge it brings that could not otherwise be obtained, and how scholars can negotiate the methodological challenges. Presenting a case study on the programming of Dutch fiction films in Amsterdam cinemas in the first decades after WWII, it develops a framework for "scalable film historical research," which outlines how digital data and tools can be integrated in a research workflow that alternates between the macro level of identifying patterns in large datasets, across space, and through time, and the micro level of one particular movie, scene, shot, location, person, or event.

The chapter **Catastrophe or Pointillism of Disaster? Annotating and Visualizing Patterns of Ecological Imagination**, from film scholar Matthias Grotkopp, presents a case study of an ongoing investigation of the spatiotemporal patterns in audiovisual discourse on anthropogenic climate change. The author outlines a film analytical method based on a structured, machine-readable semantic vocabulary: the Ada Filmontology. He argues that this qualitative empirical method is a valuable addition to other tools and infrastructure within the realm of digital film and media studies. The focus is on interpreting the visualizations that can be generated on the basis of fine-grained semantic annotations. Grotkopp makes the case that we can describe shared and circulating patterns of spatio-temporal scenarios of perception among different modes and formats of audiovisual media that together make up our perception and imagination of the ecological crisis.

The section closes with the chapter **"Pure Information, Not the Real Thing": Digital Hermeneutics and Nelson Sullivan's Videographic Legacy (1983–1989)** from media historians Susan Aasman and Tom Slootweg. The authors explore digital hermeneutics as a key approach for media historical research, with a specific focus on sources that are multi-layered, multimodal, and polyvocal. Their object of study is the work of American videographer Nelson Sullivan (1948–1989), who chronicled queer subcultures and the club scene of New York City during the 1980s. Audiovisual sources that pertain to historical autobiographical and amateur media production are notoriously difficult to reduce to neatly categorized units of analysis, because they do not necessarily adhere to formal aesthetic or narrative conven-

tions. Exploring ways to cope with such archival material and its multimodal and deictic complexity, Aasman and Slootweg elaborate on what digital hermeneutics can offer to further our understanding of a unique collection such as this one – a collection, moreover, that has existed in various iterations and guises over the last three decades, during various moments of media technological change.

Epilogue

In the epilogue **Narration, Agency, and the Digital Film Historiography Group Chat**, media scholar and pioneer in digital media scholarship Eric Hoyt compares *Doing Digital Film History* to a multi-faceted group chat due to the dynamic and collaborative nature of the book. For him, the wide range of critical methods presented in the chapters illustrate once again that history is an ongoing process, a continuous interaction between sources of the past and scholars of the present. He also reminds us not to lose sight of the historical research questions when exploring and reflecting digital methods. In this regard, he sees the Women Film Pioneers Project, discussed by several authors in this book, as an exemplary model because it is equally concerned with narrative, agency, big data opportunities, and self-reflection. He proposes a sequel to this book to continue this collective conversation about film historiography in the digital age.

Overall, this volume offers various insights into the changing practices of doing digital film history. By providing a systematic overview of critical case studies that address some of the most pressing issues arising from digital film history methods, tools, and concepts, the chapters highlight recent advances as well as current challenges facing digital film historiography, while proposing a framework for future directions.

References

Acland, Charles R., and Eric Hoyt, eds. *The Arclight Guidebook to Media History and the Digital Humanities*. Falmer: REFRAME Books, 2016. Accessed April 3, 2024. https://projectarclight.org/book/.

Arnold, Taylor, Jasmijn Van Gorp, Stefania Scagliola, and Lauren Tilton. "Introduction: Special Issue on AudioVisual Data in DH." *Digital Humanities Quarterly* 15, no. 1 (2021). Accessed April 3, 2024. https://www.digitalhumanities.org/dhq/vol/15/1/000541/000541.html.

Burghardt, Manuel, Adelheid Heftberger, Johannes Pause, Niels-Oliver Walkowski, and Matthias Zeppelzauer. "Film and Video Analysis in the Digital Humanities – An Interdisciplinary Dialog."

Digital Humanities Quarterly 14, no. 4 (2020). Accessed April 3, 2024. http://digitalhumanities.org/dhq/vol/14/4/000532/000532.html.

Dang, Sarah-Mai. "Digital Tools & Big Data: Zu gegenwärtigen Herausforderungen für die Film-und Medienwissenschaft am Beispiel der feministischen Filmgeschichtsschreibung." *MEDIENwissenschaft: Rezensionen / Reviews* 35, nos. 2–3 (2018): 142–156. http://dx.doi.org/10.25969/mediarep/4393.

Drucker, Johanna. *The Digital Humanities Coursebook: An Introduction to Digital Methods for Research and Scholarship*, 1st ed. Abingdon, Oxon; New York: Routledge, 2021.

Fickers, Andreas, Juliane Tatarinov, and Tim van der Heijden. "Digital History and Hermeneutics – Between Theory and Practice: An Introduction." In *Digital History and Hermeneutics*, edited by Andreas Fickers and Juliane Tatarinov, 1–20. Berlin: De Gruyter, 2022. https://doi.org/10.1515/9783110723991-001.

Hagener, Malte, and Yvonne Zimmermann, eds. *How Film Histories Were Made: Materials, Methods, Discourses*. Amsterdam: Amsterdam University Press, 2023.

Heftberger, Adelheid. *Digital Humanities and Film Studies: Visualising Dziga Vertov's Work*. Quantitative Methods in the Humanities and Social Sciences. Cham: Springer International Publishing, 2018. https://doi.org/10.1007/978-3-030-02864-0.

Olesen, Christian Gosvig. "Digital Film Historiography – A Bibliography." *Film History in the Making* (blog), August 19, 2020. Accessed April 3, 2024. https://filmhistoryinthemaking.com/digital-film-historiography-a-bibliography/.

Olesen, Christian Gosvig. *Visualizing Film History: Film Archives and Digital Scholarship*. Bloomington: Indiana University Press, 2025.

Sayers, Jentery, ed. *The Routledge Companion to Media Studies and Digital Humanities*. New York: Routledge, 2018.

Smither, Roger. *Evaluating Computer Cataloguing Systems: A Guide for Film Archivists*. Brussels: FIAF, 1988.

Smither, Roger. "Formats and Standards: A Film Archive Perspective on Exchanging Computerized Data." *The American Archivist* 50, no. 3 (1987): 324–337.

Tsivian, Yuri. "Cinemetrics: Part of the Humanities' Cyberinfrastructure." In *Digital Tools in Media Studies: Analysis and Review, An Overview*, edited by Michael Ross, Manfred Grauer, and Bernd Freisleben, 93–100. Bielefeld: Transcript, 2009.

Prologue

Jane M. Gaines

The DH Dilemma: Knowing More & Knowing for Sure vs. Never Knowing At All

Introduction

Why Digital Humanities? In 2004, two U.S. Humanities Center heads published a "Manifesto" in *The Chronicle of Higher Education.* Humanities in research universities were being left in the dust, they argued. In their analysis, humanities knowledge, in contrast with the sciences and social science, had not been historically associated with "discovery."[1] A decade later, the opposition between fields continued, as in Alan Liu's reference to "two cultures," although he saw an "artificial divide" between the sciences and the humanities and envisioned Digital Humanities as bringing the humanities and the sciences closer together.[2] I am not so sanguine about Liu's bridge to the sciences, however. Instead, I note with interest his advocacy for more dialogue between digital humanities and new media studies, as well as media archaeology.[3] For this, I have more hope, especially as the critique of what I call the "computational" develops.[4] As for film and media studies, I propose a contrast *within* the promise of datafication – knowing more, as opposed to knowing for sure, that is, knowing with scientific certainty. But in the end, I veer off into philosophical territory and conclude with the quandaries of what it means to find productivity in the void of *never knowing at all*, that is, never knowing despite having searched.

The 2004 DH "Manifesto" sets up this field dichotomy with its reference to, on the one hand, "critical reflections" on the creation of knowledge and, on the other, the "effects" of knowledge and the attendant implications.[5] It would seem that "knowledge-effects" implies measurability and consequently accountability, whereas

1 Cathy Davidson and David Theo Goldberg, "A Manifesto for the Humanities in a Technological Age," *The Chronicle of Higher Education* 50, no. 3 (February 13, 2004): B7–B9, accessed February 28, 2023, https://www.chronicle.com/article/a-manifesto-for-the-humanities-in-a-technological-age/.

2 Alan Liu, "Where Is Cultural Criticism in the Digital Humanities?" in *Debates in the Digital Humanities*, ed. Matthew K. Gold (Minneapolis: University of Minnesota Press, 2012), 496.

3 Liu, "Where is Cultural Criticism in the Digital Humanities?" 501.

4 I have long preferred "computational" to "digital," but this stems from a conversation I had with Kate Hayles around 2007 in which I recall that she said computer scientists preferred "computational"; while at the time I thought that the term might emerge as more pronounced in the humanities, its functionality remains to be seen.

5 Davidson and Goldberg, "Manifesto."

https://doi.org/10.1515/9783111082486-002

"critical reflections" suggests quite the opposite of the quantifiable and the verifiable. For a moment, however, think of the enormity of the area of inquiry that "critical reflections" can cover. Critical approaches might not only interrogate but might completely abandon the computational project of data collection, storage, and output. Or, critique can even become inquiry into the almost incomprehensible or even unfathomable, either because of the enormity of datasets or the notorious opacity of computational processes themselves.[6]

What is our job? Thinking about how sound and image-making machines were invented and what such machines can do has been basic to film and media theory and historiographic research across the twentieth century and into the twenty-first. In 2009, Tara McPherson charged the field with theorizing new computational research methodologies, arguing that a film theory legacy should translate into a special facility, a ready-made critical orientation, and asking "Who better to reimagine the relationship of scholarly form to content than those who have devoted their careers to studying narrative structure, representation and meaning, or the aesthetics of visuality?"[7] In the decade after her invitation to such reimagining, many scholars trained in the field innovated amazing projects that transformed historical research.[8] But let's add to McPherson's point about film scholars having a background in narrative and visual aesthetics, their knowledge of the history of technology as well as the broadcast of mass culture – preparation for some of the best work on the history and theory of networked communication.[9] Still, as Miriam Posner and Lauren F. Klein remark, the field as a whole has yet to "fully grapple" with data as a medium.[10]

Film and media studies may already be aligned with the broad "critical humanities," as it continues the politics of "critical theory," and remembers post-

6 Christine L. Borgman, *Big Data, Little Data, No Data: Scholarship in the Networked World* (Cambridge, MA: MIT Press, 2015); Mark B. N. Hansen, *Feed-Forward: On the Future of Twenty-First Century Media* (Chicago: University of Chicago Press, 2015), 70, refers to technical processes that are "cognitively inscrutable" to humans.

7 Tara McPherson, "In Focus: Digital Scholarship and Pedagogy," *Cinema Journal* 48, no. 2 (2009): 120.

8 For one list of links, see: https://transformationsconference.net/dh-cinema-projects, accessed February 28, 2023.

9 See, for example, Tung-Hui Hu, *A Prehistory of the Cloud* (Cambridge, MA: MIT Press, 2015).

10 Miriam Posner and Lauren F. Klein, "Editor's Introduction: Data as Media," *Feminist Media Histories* 3, no. 3 (2017): 2. See also Marsha Kinder, "Medium Specificity and Productive Precursors: An Introduction," in *Transmedia Frictions: The Digital, the Arts, and the Humanities*, ed. Marsha Kinder and Tara McPherson (Berkeley: University of California Press, 2014), 3.

structuralism and semiotics.[11] Also emerging here is "critical code studies," which considers a wider range of technologies of datafication, including self-tracking devices like the wearable Fitbit.[12] And, in the following, this critical theory legacy informs my focus on graphic form data visualization or data modeling as a point of departure to discuss the implications of using applications designed for the sciences as one way to configure humanities projects around the question of disciplinary difference. But before we go too far down this road, since we are considering disciplines, the "form of knowledge" issue will inevitably lead back to what, after all, constitutes knowledge, in addition to how we assume it is acquired and, most importantly, stored, as datasets grow to enormous proportions and Google Books scans millions of book titles that it refers to as a "body of knowledge."

And yet introducing the term "knowledge" is always risky, given the centuries-old philosophical approach to thinking about the nature of knowledge. "What is knowledge?" as a question dates from Socrates' dialogue with young Theaetetus to whom he explains that knowledge is not the same as belief but requires "justification." The elaboration of justification as irrefutable evidence then fell to Descartes, who introduced "certainty" as a concept. Descartes is followed by John Locke, who developed empiricism as the philosophy that beliefs are justified by experience. Locke's further elaboration finds knowledge stored in the mind as the collection of ideas.[13] What surprises me in this textbook account is that philosophy, also associated with the imponderable and the unfathomable, was so early associated with evidentiary certainty, historically the terrain of science. And so it is that common sense has come to see knowledge as synonymous with certainty along with its attendant empirical vantage which has its tautological form in *to know is to know*.

We must ask about knowledge in the light of a standard position used to defend Digital Humanities computational methods. These methods are not only defended by the research advantage of "more" knowledge, and "knowledge labor" made efficient. One also encounters the claim that datafication yields "better knowledge." This, however, is where we need to ask: does "better" mean more "correct" or "more data"? If "more data," this is where computation has made an undeniable impact in silent film

11 One often finds reference to Roland Barthes, "From Work to Text," in *Image/Music/Text*, trans. Stephen Heath (New York: Hill and Wang, 1977), 155–164. In the U.S., the journal that keeps the connection vital is *Critical Inquiry*. See, most recently, Orit Halpern, Patrick Jagoda, Jeffrey West Kirkwood, and Leif Weatherby, "Surplus Data: An Introduction," *Critical Inquiry* 48, no. 2 (2021): 197–210, accessed June 30, 2023, https://www.journals.uchicago.edu/doi/full/10.1086/717320.

12 Mark Marino, *Critical Code Studies* (Cambridge, MA: MIT Press, 2020); Sun-ha Hong, *Technologies of Speculation: The Limits of Knowledge in a Data-Driven Society* (New York: NYU Press, 2020).

13 Anthony Appiah, *Thinking it Through: An Introduction to Contemporary Philosophy* (New York: Oxford University Press, 2003), 41–47, 53–54.

history research in the kinds of breakthroughs I call historical *course correction*. One recent example of *course correction* would be the work of the American Film Institute (AFI) project "Women They Talk About: Exploring Female Filmmakers in Early American Cinema," for which see Figure 1, research on topics most used by silent era women writers in the form of a data visualization. With a National Endowment for the Humanities grant, AFI researchers undertook a statistical analysis of silent era film credits. One of their goals was to correct the widely circulated estimate that 50% of silent era screenwriters were women. The AFI study of the U.S. case shows that, on the contrary, between 1910 and 1930, 27.5% of U.S. feature films were credited to women.[14]

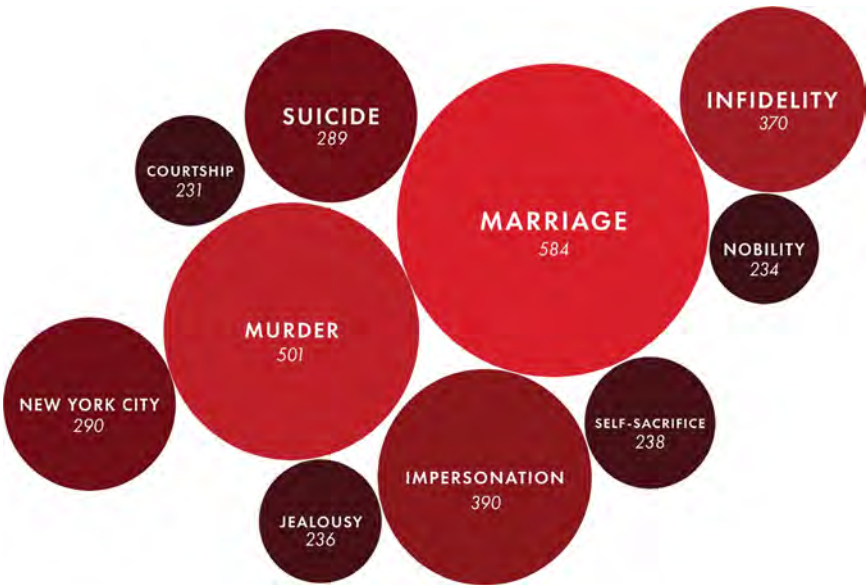


Figure 1: "Women They Talk About" Project Data Visualization. Top Ten Subjects of Films Written About by Women in Silent Era. Courtesy American Film Institute.

When it comes to historical *course correction*, computational methods have been a boon across humanities disciplines, regardless of the cultural status of the object of study. Does this mean that digital humanities has in any way levelled the "cultural

14 American Film Institute: "Women They Talk About" Project (January 2023), accessed June 30, 2023, https://aficatalog.afi.com/wtta/; Jane M. Gaines, "How Wrong We Were: Women in the Silent Era American Film Industry," accessed June 30, 2023, https://devaficalmjediwestussa.blob.core.windows.net/images/sites/3/2023/01/AFI-Women_Gaines.pdf.

playing field," so to speak? Maybe, maybe not. Film and media studies is oddly implicated in the question of knowledge since, as John Hartley has pointed out, our original object of study did not originally "qualify" as knowledge in the academic sense. Unlike word-based print – the "preferred medium of knowledge" – moving pictures were thought to require no specialized knowledge at all.[15] Proof of this "no knowledge needed" was that no education was required to access film and television entertainment. What was there to know or to learn to appreciate about "the popular" anyway? After all, moving images were so effortlessly "known" – that is, recognizable for what they represented (the visible world as seen), and furthermore were widely available as mass culture. Such mass culture was not as scarce and therefore not as valuable as high culture. Against this, semiotics and cultural studies taught that images needed to be "read" by academics. Still, institutions of higher learning have been keepers of the secret key that unlocks access to the print-based culture, that culture on which academic humanities knowledge has historically relied. And one more thing about our originally unacademic object of study – because motion pictures were machine-made, not man-made, or worked automatically without humans, the field compensated for this by privileging old humanities high art approaches. As we all know, the best example of this old humanities hold over is *auteur* studies, dedicated as it has been to elevating the motion picture film to the degree that it could be found to be the artistic "expression" of the director. However, as a consequence of decades of attempts to elevate moving pictures by the analogy to literature, we may have been too slow to embrace the history of technology, relative to which we are now called upon to think how machines produce "art" if "art" is considered the exclusive domain of the artist-human.[16]

In addition, film and media studies, in comparison with literary studies, has had a historically different relation to computational developments in research and teaching. The contemporary moment might appear to witness a trajectory *away* from 1970s close analysis of the single canonical film text toward distribution and exhibition, but at the same time moves *back* to the canonical text with computation aiding ever more sophisticated variations on shot counting.[17] Reach-

15 John Hartley, "Digital Scholarship and Pedagogy, the Next Step: Cultural Science," *Cinema Journal* 48, no. 2 (2009): 141–143.

16 Kashmir Hill, "With A.I. Appropriating Their Style, Some Artists Join the Resistance," *New York Times* (February 19, 2023), 6.

17 For one example, see the study of city films exemplified by BERLIN, SYMPHONY OF A GREAT CITY (1927) and MAN WITH A MOVIE CAMERA (1929) in Eva Hielscher, "The Phenomenon of Interwar City Symphonies: A Combined Methodology of Digital Tools and Traditional Film Analysis Methods to Study Visual Motifs and Structural Patterns of Experimental-Documentary City Films," *Digital Humanities Quarterly* 14, no. 4 (2020), accessed June 30, 2023, http://www.digitalhumanities.org/dhq/vol/14/4/000495/000495.html.

ing back further, however, the academic foundation of film studies, dedicated as it has been to close text analysis and theories of the image, did not commit to countable research outcomes but to the exact opposite. It was instead avowedly anti-empiricist.[18] In addition, the post-structuralist rejection of objectivity charged the documentary with having too close a tie to the camera as objective instrument. In the 1970s *Screen* theory moment, historian Barry Salt became a pariah for counting average shot length (ASL) and correlating film titles based on ASL by decades; his method exemplified empirical tabulation and measurement, then an anathema to film theory, dedicated as that theory was to anti-historicism and the critique of positivism. Yet after Yuri Tsivian's *Cinemetrics* ASL collection project website launched in 2005, Barry Salt was rehabilitated along with his Starword statistical style analysis.[19] In a way, the *Cinemetrics* model, for which see Figure 2, anticipated the revolution in the analysis of film form, an approach which comes to fruition in Barbara Flückiger's Analysis of Film Colors, a massive project testifying to the complexity of the cinematic text, but also reminding us of the multiple numerical aspects of our particular object – edge numbers, shot length in seconds, aspect ratio, film stock gauge measured in millimeters, responsiveness to light represented by numbers, and of course the measurement of film stock color temperature.[20]

Just considering the material specificity of our moving image object of study, as well as the statistical significance of its mass circulation, it might seem that the use of computational tools yielding quantified research outcomes is inevitable and returns us to empiricism with a vengeance. We can point this out, but we cannot leave it at that. Why? Two reasons:

1. Because our assessment may be too soon. I say this, considering the especially creative analytical possibilities. For example, see ALIEN VISIONS (Pablo Nunez Palma & Bram Loogman, 2020), a work based on machine sorting of newsreel footage that takes the problem of machine indecipherability and reverses it to ask: what if humans were the machine and the human was alien?[21]

18 See Jane M. Gaines, "What Happened to the Philosophy of Film History?" *Film History* 25, nos. 1–2 (2013): 70–80.

19 Barry Salt, *Moving Into Pictures: More on Film History, Style, and Analysis* (London: Starword, 2006), accessed June 30, 2023, www.starword.com/index.html.

20 Barbara Flückiger and Gaudenz Halter, "Methods and Advanced Tools for the Analysis of Film Colors in Digital Humanities," *Digital Humanities Quarterly* 14, no. 4 (2020), accessed June 30, 2023, http://digitalhumanities.org/dhq/vol/14/4/000500/000500.html.

21 See https://sensorymovingimagearchive.humanities.uva.nl/index.php/2020/05/17/semia-artist-projects-and-alien-visions-pablo-n-palma-bram-loogman-2020/, accessed June 30, 2023. For those of us who have regretted the paucity of Digital Humanities approaches to sound, listen to *Rendered Environments* (Adam Jurasezek, 2016), processed using SuperCollider software from radio broad-

Figure 2: *Cinemetrics* frame-by-frame count and ASL graph. Man With a Movie Camera (Dziga Vertov, Soviet Union, 1929). Courtesy Cinemetrics.

2. Because we know too much. By this I mean that there are familiar theoretical issues at stake and here I raise only a few. For one, with the computational, the old stand-off between word culture and image culture comes to the fore again. Mark Williams, situating the Media Ecology Project, foregrounds the historic "tension" between visual culture and word culture.[22] For "keywording," "searching," "tagging," and word "prompting" co-exist with the outcome – graphic visualization of data – at the operational level. Note the function of "prompt" in the verb form: words "prompt" images. There is a disciplinary dichotomy between word culture and visual culture, as well as a hierarchy evidenced in the degree

cast snippets: https://sovnrecords.bandcamp.com/album/rendered-environment, accessed June 30, 2023. I would still argue that we need more work that goes further to blur the distinction between experimental video art and digitally enhanced research projects.

22 Mark Williams, "The Media Ecology Project: Collaborative DH Synergies to Produce New Research in Visual Culture History," *Digital Humanities Quarterly* 15, no. 1 (2021), accessed June 30, 2023, http://www.digitalhumanities.org/dhq/vol/15/1/000524/000524.html.

to which word culture research projects far outnumber visual culture projects. As Lev Manovich puts it, Digital Humanities as a field remains dominated by word culture.[23] He also objects to the very term "digital humanities" which he finds both too narrow and too broad.[24] His analysis of broad *and* narrow, however, lends support to my argument that film and media studies projects need to engage with the larger field *and* to carve out a space of expertise within it that might be *for* as well as *against* "digital humanities." But an important caveat: such expertise *must include the film and media studies legacy of critique*. Recall the critique of traditional empirical methodologies that responded to the so-called "historical turn" to New Film History, a critique on behalf of theory: the issue was a concern that the turn to empiricism abandoned the anti-empiricist foundations of the field, as I have just noted.[25] But now the question has been raised as to whether we need to call out a second "historical turn" relative to the "new new cinema history."[26] One might say, however, that this "new new empiricism" is not a "turn" either backward or forward but a "leap" to updated empirical research methods, and ever-more-amazing display design, as well as platforms for testing future technologies. But something is a little strange here and our challenge decidedly different from what it was at the New Film History advent. Critique requires us to historicize and to theorize the software-hardware interface of the computational apparatus *at the same time* that we are relying on that very apparatus for historical research on earlier technologies and their social moments.

23 Lev Manovich, *Cultural Analytics* (Cambridge, MA: MIT Press, 2020), 9. His example is one of the smaller percentage of visual culture projects presented in the 2019 Utrecht Conference on Digital Humanities.

24 Manovich, *Cultural Analytics*, 7. Marsha Kinder and Tara McPherson, "Preface: Origins, Agents, and Alternative Archaeologies," in *Transmedia Frictions: The Digital, the Arts, and the Humanities*, ed. Marsha Kinder and Tara McPherson (Berkeley: University of California Press, 2014).

25 Gaines, "What Happened to the Philosophy of Film History?"

26 Daniel Biltereyst, Richard Maltby, and Philippe Meers, "Introduction: The Scope of New Cinema History," in *The Routledge Companion to New Cinema History*, ed. Daniel Biltereyst, Richard Maltby, and Philippe Meers (London: Routledge, 2019), 9. Chris Yogerst, review of *Explorations in New Cinema History*, posits another "historical turn," *Journal of Cinema and Media Studies* 60, no. 3 (2021): 210.

Arclight Software, *Lantern* Search Tool, and Early Film Datasets

As we know from "critical code studies," whatever the promise of statistical certainty, data is not pristine, given the unavoidability of biases in dataset selection and software design.[27] Cautioning us, the editors of *The Arclight Guidebook to Media History and the Digital Humanities* quote Deb Verhoeven on the kinds of assumptions programmed into databases, leading her to warn that "interrogating software and databases is as important as technical mastery."[28] It might appear, however, that more often than not, interrogation is limited to a cautionary aside within the literature. Where and how do we undertake a robust interrogation, from science-based software development, to server farm energy gluttony, to biases in datasets, to deceptive data visualization? Can we do this all at once?

Although in the following I begin by testing approaches to the critique of data visualization exemplified by the Timeline and the Dendrogram, I first want to highlight a key difference between the literature on film and media studies computational projects and literary studies scholarship, which even insightful analyses of what could be called an "ideology of datafication" may gloss over or postpone.[29] For example, one literary scholar states that, just as in other disciplines, humanities scholars are increasingly dealing with "data." But then he asks whether "data" replaces "books, paintings, movies."[30] The problem here is that this observation only deals with the *word* "data" and not at all with the processes of scanning or digitizing books as opposed to still and moving images. Datafication is, of course, the product of documents having been scanned or digitized, and my impression is that the terms *digitization* and *scanning* appear more in film and media studies than in literary digital humanities project studies. One expects to see in all DH literature the standard thinking about image "resolution" based on scan rate, the difference between 600 and 1200 dpi (dots per inch). Or the pro-

27 Lisa Gitelman and Virginia Jackson, "Introduction," in *"Raw Data" is an Oxymoron*, ed. Lisa Gitelman (Cambridge, MA: MIT Press, 2013), 1–4.

28 As quoted in Eric Hoyt, Kit Hughes, and Charles R. Acland, "A Guide to the Arclight Guidebook," in *The Arclight Guidebook to Media History and the Digital Humanities*, ed. Charles R. Acland and Eric Hoyt (Falmer: REFRAME Books, 2016), 23, accessed June 30, 2023, https://projectarclight.org/book/.

29 However, see Hong, *Technologies of Speculation*, 4–5, for an analysis of how "technologies of datafication" work through a new objectivity to effectively deliver "better knowledge" in the image of data visualization itself.

30 Christof Schöch, "Big? Smart? Clean? Messy? Data in the Humanities," *Journal of Digital Humanities* 2, no. 3 (2013): 1.

cess of datafication in which aesthetic objects that are unlike each other become information and are then transformed into computable data to be operationalized, all of which requires software/computer interface. One wonders as well why in the majority of DH literature one finds so few references to Optical Character Recognition (OCR) software and the process of translating numerical into graphic values. Given that the wider DH field remains so dominated by word culture, I wonder how much longer such seeming obliviousness to the technological can be sustained.[31]

One model of how to foreground the computational is Eric Hoyt's article on how he discovered a hierarchy of influence among early motion picture exhibitors, where he offers detail about the functioning of the Media History Digital Library (MHDL) interface that he helped to design.[32] Even with Hoyt's level of explanation, however, MHDL processes may remain impenetrable to anyone who is not as yet digitally fluent. What Hoyt explains is that the scaled entity search (SES) uses *Arclight* software and the search platform *Lantern* to "draw on" the dataset of the Media History Digital Library. How it works is that "users" input a word into the "query box" located on the MHDL home page. What Hoyt refers to as the "algorithmic backbone" uses Apache Solr, an open-source search technology. From search terms supplied by the "user," the *Archlight* software "generates a line graph." In the next step, the "user" may click on dots that appear in the graph, with each click opening pages of *Photoplay* or *Moving Picture World*, for instance, in *Lantern*.[33] That online search processes and the attendant terminology have already become second nature in the field is all the more reason to foreground steps, name software as well as hardware, and, ideally, interrogate every term. Why, for instance, have we settled so complacently on the term "search"? How did the "user" become constituted as a "user"?[34]

Yet in the same article, Hoyt acknowledges one criticism of digital humanities computational approaches with his metaphor of the way technology *may determine* the kind of research we undertake. As he puts it, if research questions are determined by computational capacity and software design, it may be that, metaphorically, this "allows the tail to wag the dog."[35] To put it another way, it may be that *we look for what we know we can find.* Or, we undertake research to fit available

31 Williams, "The Media Ecology Project."

32 See https://mediahistoryproject.org/, accessed June 30, 2023.

33 Eric Hoyt, "Arclights and Zoom Lenses: Searching for Influential Exhibitors in Film History's Big Data," in *The Routledge Companion to New Cinema History*, ed. Daniel Biltereyst, Richard Maltby, and Philippe Meers (London, New York: Routledge, 2019), 84, 86.

34 Joanne McNeil, *Lurking: How a Person Became a User* (New York: Farrar, Straus and Giroux, 2020).

35 Hoyt, "Arclights and Zoom Lenses," 85.

software tools. The keyword search via Lantern Search engine of publications, such as *Motion Picture World*, invites early cinema "media industries" projects, and *Photoplay* holdings encourage more star studies. Hoyt would counter this with the increased infinity of research possibilities, especially as the MHDL currently contains 2,845,814 pages, 1,944 of which pertain to global cinema sources from China, Iran, and India, now added to the original U.S. and European sources.[36]

Data Visualization

We should still wonder why film and media studies has been relatively peripheral to digital humanities scholarship, given that visual culture analysis is so integral to the broader field. And where more than in the standard use of data display – graphs and charts, maps and timelines? Johanna Drucker in her extensive work on visualization, however, sees a conceptual inconsistency that comes to a head in the use of graphic design in computational data display. Here she finds a paradox: the legacy of *distrusting* images and the *trust* placed in quantitative statistical approaches rendered as data visualization. Why? Charts and graphs appear explicit and unambiguous, even straightforward, in their presentation of information, she answers.[37] But think again. A graph, especially without context, may appear relatively information poor in its non-representational abstraction. Wouldn't such a design signify not explicitness but ambiguity and therefore openness to a range of interpretations? Here, I think, is a second paradox: information certainty is information poverty. Despite this apparent inconsistency (to state the obvious about the computer-generated data chart), trust is crucial, given the representational imperatives built into visualization software: to confirm the assumption of visual equivalency of the mathematical computational; to stand for what it is that we say that we know for sure – to prove beyond doubt, to prove for certain, especially in the context of business and science. To take a field specific example, however, consider Figure 3. The graph represents the use of Project Arclight and Lantern search tools to chart the industry story coverage of Chinese-American motion picture actress Anna May Wong, in contrast with that of German Marlene Dietrich between 1920 and 1945, mapping career highs and lows. What would be our critique of this visualization? After all, the timeline seems so straightforward. But such very apparent straightforwardness is what Drucker wants to call to our attention.

36 See https://mediahistoryproject.org/collections/global-cinema/, accessed June 30, 2023.

37 Johanna Drucker, *Visualization and Interpretation: Humanistic Approaches to Display* (Cambridge, MA: MIT Press, 2020), 6.

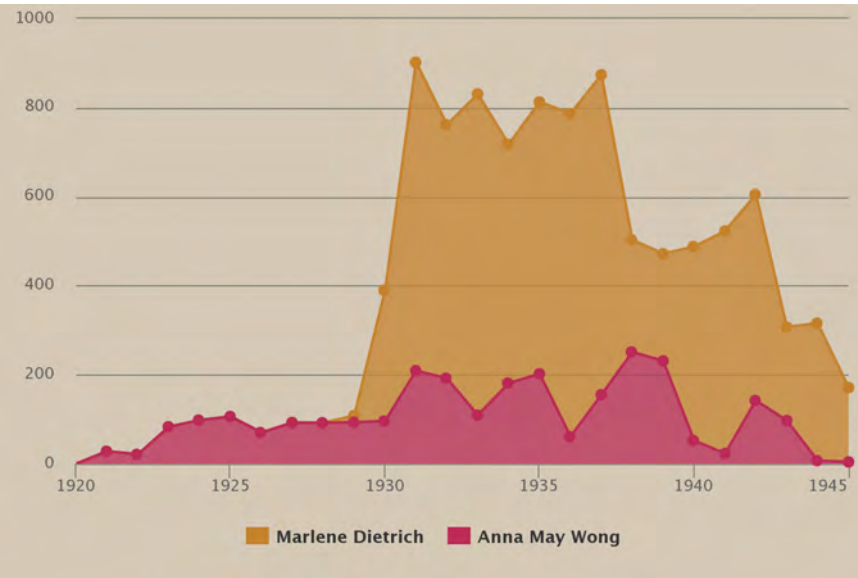


Figure 3: Arclight timeline with graph representing the industry story coverage of Marlene Dietrich and Anna May Wong. Courtesy Media History Digital Library.

Drucker's position is an eerie carryover from the ideological critique of representational images, and a recap of structuralism and semiotics. So when she takes a critical approach to the visualization of datasets and the interpretation of such graphic modeling, it is difficult for media theorists not to experience a sense of *déjà vu* on encountering the phrase "enunciative apparatus of information systems."[38] Because we are adept at the ideological critique of the visual image, as well as narrative as carrier of meaning, one assumes that we would be poised and ready to turn the same critique on forms of *data* or *information visualization*. And what would these visualizations be? Answer: pie chart, dot strip plot, doughnut, ordered bar graph, and timeline. What are these forms expected to represent? Answer: deviation, correlation, distribution, ranking order, spatial relations, magnitude, change over time, part-to-whole relations, and flow, to name a few variable relations of correspondence and non-correspondence.

To start with *data visualization* as one feature that film and media studies computational projects have in common with computational literary studies – the graphic that Drucker treats as a "knowledge form." Here, numbers are translated into graphic values in such a way that their very presentation may be read as in

38 Drucker, *Visualization and Interpretation*, 7.

some way equivalent; that is, the data and the visualization of it are considered, as she says, "one and the same."[39] Then, she asks what "knowledge claims" underly the use of *data visualization*, which as a graph or timeline appears on a screen as something like a "statement of fact" in image form, that is, an abstract non-representational image.[40] Basic to Drucker's critical project is the insistence that the humanities have inherited conventions of data presentation from mathematics, natural sciences, and social sciences, where the very software design assumes that the graph is an "expression" of underlying data. Here, she thinks, is exemplification of the perspectives of empiricism and positivism where the meaning of texts is treated as fixed.[41]

Let's stop a moment here to remind ourselves of how positivism may serve to advance an idea of history, the discipline, as scientific. Then to acknowledge the antidote to this perspective, the school of thought following Nietzsche: "History, so far as it serves life, serves an unhistorical power, and thus will never become a pure science like mathematics."[42] This is also the position of the contemporary philosophy of history, somewhat of an outlier within the discipline of history.[43]

Drucker wants us to foreground the invisibility of the acts of interpretation that disappear in the face of the visualization of statistical data, and to perform something like an ideological critique of the graph, machine/software interface, and search engine navigation. Ideology, as "cultural value," she continues, can be found "in every graphic, layout, format, bit of iconography (as well as interface & navigational features . . . even as it disappears . . . [by means of conventionalization]."[44] Yet we must ask if this doesn't seem elementary. Yes and no. Calling for an ideological critique of *data visualization* might slow the leap onto the digital humanities bandwagon in which functional application comes first, and theory second, if at all. Think back now to the excitement around the possibilities of database reconfiguration of narrative structure based on Marsha Kinder's Labyrinth Project in 1997.[45] Because the Labyrinth projects were received as alternatives to narrative linearity, the first questions they raised did not necessarily have to do with machinic functions.

39 Drucker, *Visualization and Interpretation*, 2.

40 Drucker, *Visualization and Interpretation*, 14.

41 Drucker, *Visualization and Interpretation*, 5–6.

42 Friedrich Nietzsche, *The Use and Abuse of History* [1873] (New York: Cosimo Classics, 2010), 12.

43 For background see Frank Ankersmit, "Bibliographical Essay," in *A New Philosophy of History*, ed. Frank Ankersmit and Hans Kellner (Chicago: University of Chicago Press, 1995), 278–283.

44 Drucker, *Visualization and Interpretation*, 15.

45 See Marsha Kinder, "Designing a Database Cinema," in *Future Cinema: The Cinematic Imaginary After Film*, ed. Jeffrey Shaw and Peter Weibel (Karlsruhe, Germany: ZKM, 2003), 346–353.

In Steve F. Anderson's theorization of recombinant "database histories," they are unlike traditional descriptions of the historical past in that they are instead collections of fragments "infinitely retrievable," categorizable, reconfigurable, and thus open to "continual revision and reinterpretation."[46] Today, thinking about "database history," we might recall Lev Manovich's earlier hypothesis in which the database succeeds the novel and cinema as the dominant form of "cultural expression."[47] Yet in this transition moment, as the database form is no longer compared with narrative form, a different question arises. While we might *hope* that the very invitation to reconfigure that is set up by the "interactive" database might produce critical distance on the production of historical narratives (plural), *the invitation alone is no guarantee*. Engagement is no guarantee that users (as a consequence of use), will come to think of narrative history as an ideological construction. Sarah-Mai Dang, describing the "Aesthetics of Access" project and the interactive Women Film Pioneers Explorer, refers to the tradition of historical research that the Explorer's searches and data reconfigurations are designed to challenge. As she puts it, traditional historiography still carries an "implicit promise" that if enough information is made available, the "whole" historical story (singular) will emerge.[48] We would ask whether resources like the Digital Media History Library offer a "knowing more" that encourages searches for the "whole" historical story. But the critique of "totality" that theorist Siegfried Kracauer associated with photography and film as well as history, for instance, may now be seen as belonging to another tradition, one not only too early but now relegated exclusively to the subfield of film-philosophy.[49] So how do we go about building political analysis in the tradition of film theory's critique of ideology into exercises in computational historiography? This, in a field where we expect researchers to be hyperaware of the constructedness of everything, especially digital materiality and machinic functions.

46 Steve F. Anderson, *Technologies of History: Visual Media and the Eccentricities of the Past* (Hanover, NH: Dartmouth College Press, 2011), 122.

47 Lev Manovich, "Database as Symbolic Form," *Convergence* 5, no. 2 (June 1999): 80.

48 Sarah-Mai Dang, "Unknowable Facts and Digital Databases: Reflections on the Women Film Pioneers Project and Women in Film History," *Digital Humanities Quarterly* 14, no. 4 (2020), accessed February 27, 2023, http://www.digitalhumanities.org/dhq/vol/14/4/000528/000528.html, 4. See also Henri Dickel, Matija Miskovic, Karazm Noori, Christian Schmidt, Atefeh Soltanifard, Sarah-Mai Dang, and Thorsten Thormählen, "Women Film Pioneers Explorer, 2021," accessed June 30, 2023, https://www.online.uni-marburg.de/women-film-pioneers-explorer/index.html.

49 Siegfried Kracauer, "Photography," in *The Mass Ornament: Weimar Essays*, ed. and trans. Thomas Y. Levin (Cambridge, MA: Harvard University Press, 1995), 47–66. See also Siegfried Kracauer, *History: The Last Things before the Last* (New York: Oxford University Press, 1969).

Database to Data Visualization: The Women Film Pioneers Explorer Dendrogram

We can start by taking a field historical perspective on the knowledge form and function of the Dendrogram visualization enabled by the software DISPLAYR on the Women Film Pioneers Explorer's website. First, as an exercise in historical contrast, imagine the 1990s when scholars began collecting documents for a set of Women Film Pioneers source books. That original project conceived in 1993 as a several-volume set never materialized in book form but morphed into an online database launched in New York under the auspices of Columbia University Libraries in 2013.[50] The Dendrogram as a visualized dataset confirms the limitations of the book form to dramatize scope and comparison within categories. Thus constrained, the original goal was to only prove that there were *more* women working as directors and writers in the U.S. in the first two decades of motion pictures than at any time since. Of the three categories made searchable – Name, Geography, and Occupation – the third yielded the most surprising information, indicating a wider range of jobs held by women than we originally projected. This was especially unexpected as we had been more interested in world geography flows, as exemplified by the case of American Fern Andra (Figure 4), a circus performer who left the small town of Watseka, Illinois, to travel to Berlin, Germany, where she founded the Fern-Andra Company in 1917.[51] In 2009, when Women Film Pioneers database design began, Dendrogram-style data display had not been developed for use outside the sciences, and it was only after the partner "Aesthetics of Access" Project used DISPLAYR software to produce data visualization that the research possibilities became evident. It is almost impossible to visualize what the computer can pattern until one has seen the data visualization version of information collected, often the work of decades.

50 See Kate Saccone's article "(Re)Visioning Women's Film History: The Women Film Pioneers Project and Digital Curatorial-Editorial Labor" in this volume for background on how the Columbia University Libraries Digital Research and Scholarship staff worked with scholars to turn Word-files and glossy images prepared for book publication into an interactive website as a pilot project on the future of publishing. Women Film Pioneers Project: https//wfpp.columbia.edu, accessed June 30, 2023.

51 For Fern Andra, see f_films: female film workers in europe, accessed March 27, 2024, https://f-films.deutsches-filminstitut.de/biographien/f_andra_b.htm. Thanks to Paulina Junginger for her continued work on early German filmmakers.

Figure 4: Fern Andra, actress/producer, founder of Fern-Andra Company. Courtesy Deutsche Kinemathek, Berlin, Germany.

In the context of a cultural studies approach to data mining, Adrian Mackenzie asks: "What does machine learning want?" The Dendrogram, as she describes it, is a diagram based on the model of a tree to represent hierarchical clusters, whose arrangement is thought to produce analyses that correspond with its clustering, as in Figure 5. If its basic use is computational biology, and it is designed for representing gene clusters and fine differentiation between those clusters, we can assume an amazing capacity to order and to differentiate. The Dendrogram graph makes it now possible to analyze classifications within occupations, given that the machine can "recognize and render patterns" that people are unable to manage.[52] In the Women Film Pioneers Explorer Project, the machine undertook the labor of sorting to reveal, for instance, variations on Script work, breaking down that category of work as: Script Assistant, Script Consultant, Script Editor, Script Girl, Script Reader, and Script Supervisor. Then, regarding our query as to whether computational machine enhancement is changing the kinds of questions we ask, consider the original question: "What happened to women in the silent

52 Adrian Mackenzie, "The Production of Prediction: What Does Machine Learning Want?" *European Journal of Cultural Studies* 18, nos. 4–5 (2015): 437.

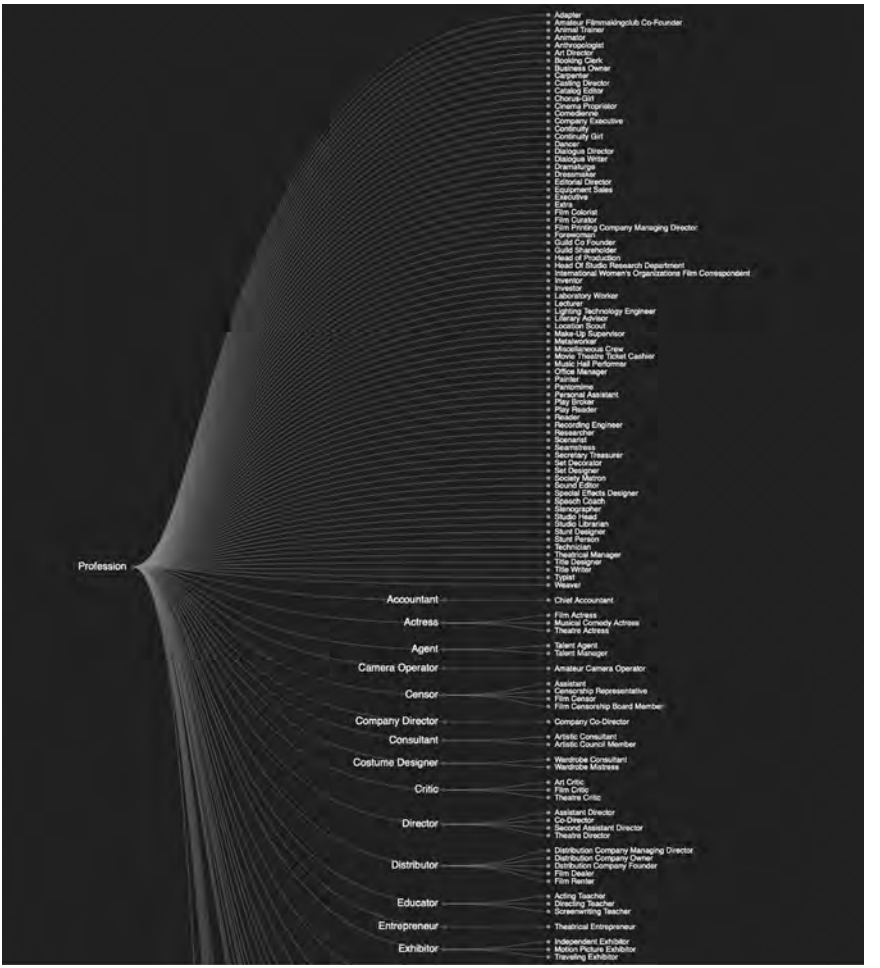


Figure 5: Women Film Pioneers Explorer Dendrogram. Courtesy Aesthetics of Access Project.

film industries who were there in numbers around 1917 and then were phased out?"[53] In effect, "What happened to them?" was an unanswerable question from an empirical standpoint and that is exactly why I asked it. The question that *did* become answerable, however, was "In what occupations were they engaged?" As a consequence of both the European "Aesthetics of Access" with its Explorer inter-

53 See Jane M. Gaines, *Pink-Slipped: What Happened to Women in the Silent Film Industries?* (Urbana: University of Illinois Press, 2018), 16–32 (chapter 1).

active database and the U.S. American Film Institute "Women We Talk About" projects, it is now possible to shift the emphasis from the original pressure to find more women film directors worldwide and to consider instead categories of work that a new industry created around production, exhibition, and distribution, as well as censorship and the promotion of motion pictures, at the turn of the last century when young women were first entering the labor force.[54]

What can computational tools not be made to do? Most difficult to chart would be the shift in epistemological positions – from the 1970s "no women" position to projects like the American Film Institute "Women They Talk About," dedicated to enumerating and classifying, and the Women Film Pioneers Explorer interactive database, which innovates alternative research approaches featuring computational reconfigurations like the Dendrogram, the Cluster, and the Timeline. In a nutshell, here is the dilemma: historically, the feminist theory of the "male gaze" in classical film narrative functioned to disallow any attempt to count, beginning in the 1970s. How, beginning in the 1990s, could we make the case *with* numbers that women were excluded *in numbers*, that is, to *use* numbers to illustrate that, in the first decades, they were *not* excluded? Indeed, researchers were finding more (in numbers) than scholars first anticipated would be found. To put it another way, the original motivation was to challenge 1970s feminism's pessimistic "no women" film theory – *no women* behind the camera, *no women* in the audience, and women on screen *only* for the pleasure of men. There is apparent *contradiction* in the goal to advance the discovery that there were women at all levels, while maintaining that theoretically there were "no women," the feminist film analysis. The challenge was to keep alive the theory of their "absence" and at the same time prove with a preponderance of evidence that women were there in numbers, as well as in influence, despite their having had little to no influence on the classical Hollywood narrative fiction film, the form exported worldwide. In the end, while the question as to what exactly "happened" to them may not be empirically answerable, to say that the question is unanswerable is not to say that there is no more to be discovered about the events in the career trajectories of the figures who are still coming to the attention of scholars. Relative to this paradox of *no women but women*, Sarah-Mai Dang asks our question: how is it possible for a database to take into account the "contradictions and contingencies of history"?[55]

54 See "Women They Talk About: Discovering America's Female Film Pioneers," accessed June 30, 2023, https://aficatalog.afi.com/wtta/; "Aesthetics of Access: Visualizing Research Data on Women in Film History," accessed June 30, 2023, https://wfpp.columbia.edu/the-aesthetics-of-access-visualizing-research-on-women-in-film-history/.

55 Dang, "Unknowable Facts and Digital Databases," 7.

Theories of History: The Timeline

A computer-generated timeline, amazing as it is, especially supplemented with a graph dotted with points, is still a timeline. Let's not forget that 1970s film theory challenged traditional historiography as the "linear" narrative that the historian imposed on events, or that, in the "linguistic turn," theorists of history challenged as the narrativization of historical discourse.[56] But as a graphic display of a data-set, a "translation" of numbers into lines and shapes, the timeline raises issues germane to the philosophy of history as well as film and media theory. Let's take these one at a time, first representationality. There is of course the data-set-to-visualization relation – the issue as to what it is that a timeline graph represents, especially if it is taken to be data that is unproblematically "the same as" its graphic form, to recap Drucker.[57] We might object that in media theory a timeline visualization might be treated not as a representation but as a "remediation" of underlying data. For instance, one can ask how many mediations the Project Arclight timeline (Figure 3) is away from the original historical events – first the pre-1920 early lives, followed by the careers of Anna Mae Wong and Marlene Dietrich. Film historians might supplement the graphic abstraction with what they know about the Chinese-American actress as well as the German actress before, during, and at the end of World War II.[58] Then, one can ask about processes of transformation from events to published fan magazine stories over the period 1920 to 1945, and from magazine pages scanned and converted in bulk to data and as numerical values transformed into screen display. Even with this kind of critical media engagement, however, we might take the timeline for granted and fail to ask why the chronological ordering of events in time in the first place – as though it is the only possible ordering. And let's not fail to notice here that the term "order" has come to be synonymous with "chronology," as though there could be no other kind of ordering.

Thus, second, insofar as the graphic timeline assumes chronology, a critique of graphic conventions cannot avoid the historical backstory of the measurement

56 On the more recent "postnarrativist" shift and reference to the earlier "historical turn," see Frank Ankersmit, "Forum Debate on Jouni-Matti Kuukkanen's Postnarrativist Philosophy of Historiography," *Journal of the Philosophy of History* 11 (2017): 1.

57 Drucker, *Visualization and Interpretation*, 2.

58 See Patrice Petro, "In the Wings," in *Idols of Modernity: Movie Stars of the 1920s*, ed. Patrice Petro (New Brunswick, NJ: Rutgers University Press, 2010), 270–283; Yiman Wang, "'Speaking in a Forked Tongue': Anna May Wong's Linguistic Cosmopolitanism," in *Revisiting Star Studies: Cultures, Themes, Methods*, ed. Sabrina Qiog Yu and Guy Austin (Edinburgh: Edinburgh University Press, 2017), 65–82.

of units tied to particular cultures, which, while leading back to the relativity of such measurement, also finds chronology everywhere, especially as the historian's basic tool. Theorist of history Reinhart Koselleck has weighed in on the dependence of standard historiography on chronology understood as "unalterable succession" of one event following another, establishing a necessary "before" and an "after" for past events. The problem, he explains, is that a chronology cannot possibly contain events and yet it must be "made to conform."[59] Finally, not so surprisingly, the standard chronological timeline also corresponds with common sense notions about temporality as a straight line, as in the metaphor of "time's arrow," and the idea that time moves from past to present.[60]

Here is the opportunity that we don't want to pass up – the chance to critique lines of time, such as the chronology as the too easily calculable ordering principle of historical events. For in Koselleck, historical grasp of past events requires a theory not of singular trajectory but of the relation between temporalities plural. In his theory of multiple times, moments that may be enacted and subjectively felt by humans overlap in such a way as to defy objective time measurement, which brings us to his idea of "subjective historical times," or times as registered, experienced, or felt.[61]

AI Affect and the Race between Humans and Machines

Lest we get stuck at the current stage of computer-aided historical research, fascinated with Project Arclight's interactive timeline, let's consider the current impetus to develop alternatives to the timeline. In Drucker's analysis, the network visualization standard based on a visualization timeline represents dynamic historical conditions as static. Most salient, but currently the most difficult, would be how to represent *change over time*.[62] Even more difficult would be how to represent Kosel-

59 Reinhart Koselleck, *The Practice of Conceptual History: Timing History, Spacing Concepts*, trans. Todd Samuel Presner and others (Stanford, CA: Stanford University Press, 2002), 108–109.

60 Koselleck, *The Practice of Conceptual History*, 115.

61 Koselleck, *The Practice of Conceptual History*, 110.

62 Drucker, *Visualization and Interpretation*, 123–124. For an example of a project, Drucker gives the 3D visualization interface *Grand Canyon* developed to represent all the events of the year 1969 drawn from online image libraries; see John David Miller and John Maeda, "A Stitch in Time: Visualizing History Through Unit Forms and Repetition Structures," 2015, accessed February 27, 2023, https://www.researchgate.net/publication/277250414_A_Stitch_in_Time_Visualizing_History.

leck's theorization of multiple historical times and "subjective historical times."[63] Drucker laments that, translated into the work of computer programmers, there is as yet no technical capacity to generate nonstandard metrics, given the need for "warping transformations" capable of indicating "affective forces."[64] She introduces the concept of "affective metrics" that would be "generated from subjective experience," as opposed to those designed to merely "register" that experience.[65] But, if we stand back from such a hypothetical, we must ask if its goal is still the quantification of the unquantifiable, the measurability of the immeasurable.

Let's consider the question of the future goal of computational modeling as adaptation to the kind of complexity valued in fields associated with the humanities as opposed to the sciences. If the issue is how to program computers to register conceptual and affective complexity, and to present such affect as graphic display, we inch towards the challenges of training computers to operate in those intellectual and affective realms in which human beings excel – the domain of current artificial intelligence research. What are tech companies leading us to think about their progress on that old question as to whether computers can be taught to think like humans, but also to feel "just like" human beings?[66] We may be familiar with the big tech argument that Virtual Reality can teach empathy, and even that VR will be the new "digital novel." Then consider the AI Now Institute's "AI Lexicon Project" blog, which features media theorist Hannah Zeavin, who has tracked "empathy" as a goal of robotics, hiring algorithms, and facial emotion recognition. But why empathy, she wants to know. For "empathy" is a strange programming goal, she thinks, given that "knowing" the other may incite violence as much as encourage understanding between people. And, she goes on, as recognition, empathy is still thought to be "impossible to code."[67]

---

63 For Koselleck's theory of history as relevant to film history, see Jane M. Gaines, "What Next? The *Historical Time* Theory of Film History," in *How Film Histories Were Made: Materials, Methods, Discourses*, ed. Malte Hagener and Yvonne Zimmermann (Amsterdam: Amsterdam University Press, 2023), 59–84. On the complexity of Koselleck's multi-layered theory of history as well as multiple temporalities, see Helge Jordheim, "Natural Histories for the Anthropocene: Koselleck's Theories and the Possibility of a History of Lifetimes," *History and Theory* 61, no. 3 (September 2022): 391–425.

64 Drucker, *Visualization and Interpretation*, 117.

65 Drucker, *Visualization and Interpretation*, 123.

66 Erin Griffith and Cade Metz, "Tech Slump Doesn't Slow New Boom in A.I. Field," *New York Times* (January 7, 2023), B1, B4.

67 Hannah Zeavin, "A New AI Lexicon: EMPATHY," AI Now Institute (September 16, 2021), accessed February 27, 2023, https://medium.com/a-n32-ai-lexicon-empathy-4da12b82e280.

Conclusion: On Not Knowing At All

To return to where I began with the 2004 assessment that science methodologies were more attuned to "discovery" than the humanities. One implication is that the use of metrics in new research initiatives would lead humanities disciplines to be taken more seriously after having reformulated more realizable computer-assisted goals. But, while we are examining disciplines, we also need to acknowledge approaches to the sciences that think about the elusiveness of scientific knowledge proof. Closer consideration reveals the magnitude of unresolved scientific research questions, as dramatized by the end of Siddartha Mukherjee's *The Song of the Cell* with the author's summary on the state of research in cellular biology: "These are mysteries beyond mysteries [. . .] We don't know what we don't know."[68] Or, following neuroscientist Stuart Firestein, who in *Ignorance: How it Drives Science* sees "not knowing" as a "condition of science," that is, the absence of fact as well as understanding. With Firestein's concept of "knowledgeable ignorance" as a way of thought leading to even better questions, we seem to find ourselves on the other side of verifiable knowledge.[69] What, then, if there is another justification for testing models of scientific inquiry on cultural history that is quite the opposite of the measurable outcomes of "discovery"? The obverse would be in "never" discovering, that is, in demonstrating phenomenological loss, forgetting, and non-existence, or the search that yields no data. What is required are methodologies based on the failures of the enumerable and measurable, which call reflexive attention to the immeasurable. What, however, is the reward for the search in vain that hits the limits of the "knowable"? We are asking how we face the impossibility of finding what we expected, of finding at all, when we thought that such a search would at least lead to "knowing more," even if not knowing for sure.

Much of the literature on computation and the humanities seems far from the philosophical literature on epistemology, which leads back to the question of existence, reaching beyond the immeasurable to the completely "unknowable." Perhaps we're called upon to think in two modalities at once – discoverable, categorizable data *and* the realm of the totally undiscoverable. This is reflected in Sarah-Mai Dang's question "Are databases also able to account for what may not be known

68 Siddartha Mukherjee, *The Song of the Cell: An Exploration of Medicine and The New Human* (New York: Simon & Schuster, 2022), 361.

69 Stuart Firestein, *Ignorance: How It Drives Science* (New York: Oxford University Press, 2012), 6–7.

and why it remains unknown?"[70] Taken one way, this question may be an invitation to continue the search for yet-to-be-discovered women. Taken another way, the state of being "unhistoricized" refers not only to women who may have existed, but also those whose lives have either yet to be deemed important enough to search for, or who may never be "discovered." For "unhistoricized" also marks a futile search for those female workers for whom no names were recorded, given the anonymity of early film industry employees.[71] We may be closer to this than one might think. Indeed, developments within the field of feminist media studies point to innovation challenging traditional historiographic scholarship.[72]

A promising model is the kind of counterfactual speculation and "thought experiment" that follows Catherine Gallagher's elucidation of the long legacy of counterfactual history.[73] Although there is as yet relatively little work on how to study the object that never existed, the event that never took place has been seriously taken up by the "counterfactual" approach that answers "What if?" with a fiction that freely departs from factuality. I have argued elsewhere that the counterfactual alternative set of events organized as a timeline is especially effective when fictionalization departs from events verified and consequently established in a field to *fill in* where we do not yet or may never know.[74] One can imagine future projects in which search results are strategically combined with historical speculation. Here we might draw a sharp distinction between the machine's computational timeline and the imaginative alternate account that, defying quantification, cannot be made to "stay on the line."

Yet something may really be afoot. What has emerged coincident with Digital Humanities film and media projects is the serious study of the "lost," the "unwatched," the "incomplete," or "unfinished" artistic project, with an empha-

70 Dang, "Unknowable Facts and Digital Databases," 4. Hoyt, in "Arclights and Zoom Lenses," 87, asks "What historical materials, processes, and experiences do not easily lend themselves to digitization and what effect does their omission have on results"?

71 Dang, "Unknowable Facts and Digital Databases," 11; see Jane M. Gaines, "Anonymities: Uncredited and 'Unknown' Contributors in Early Cinema," in *A Companion to Early Cinema*, ed. André Gaudreault, Nicholas Dulac, and Santiago Hidalgo (London: Wiley-Blackwell, 2012), 443–459.

72 Allyson Nadia Field, "Editor's Introduction: Acts of Speculation," *Feminist Media Histories* 8, no. 3 (2022): 1–7, a double issue on "Speculative History."

73 Catherine Gallagher, *Telling It Like It Wasn't: The Counterfactual Imagination in History and Fiction* (Chicago: University of Chicago Press, 2018).

74 Jane M. Gaines, "Counterfactual Speculation: What if Antonia Dickson Had Invented the Kinetoscope?" *Feminist Media Histories* 8, no. 3 (2022): 8–34, accessed June 30, 2023, https://online.ucpress.edu/fmh/article/8/3/8/190671/Counterfactual-SpeculationWhat-if-Antonia-Dickson.

sis on film and video works produced by women.[75] These projects may not have been conceived in defiance of the measurable or in opposition to the so-called "digital turn." And yet. Taken together, the "unwatched," "incomplete," and "never made," along with the "unhistoricized," constitute a field of inquiry to parallel the computational promise of measurable data.[76] We could then balance knowledge as certainty with "unknowable" phenomena within a speculative historiography that takes seriously the likelihood of never knowing at all.

References

Acland, Charles R., and Eric Hoyt, eds. *The Arclight Guidebook to Media History and the Digital Humanities*. Falmer: REFRAME Books, 2016. Accessed June 30, 2023. https://projectarclight.org/book/.

Anderson, Steve F. *Technologies of History: Visual Media and the Eccentricities of the Past*. Hanover, NH: Dartmouth College Press, 2011.

Ankersmit, Frank. "Forum Debate on Jouni-Matti Kuukkanen's Postnarrativist Philosophy of Historiography." *Journal of the Philosophy of History* 11 (2017): 1–10.

Ankersmit, Frank. "Bibliographical Essay." In *A New Philosophy of History*, edited by Frank Ankersmit and Hans Kellner, 278–283. Chicago: University of Chicago Press, 1995.

Appiah, Anthony. *Thinking It Through: An Introduction to Contemporary Philosophy*. New York: Oxford University Press, 2003.

Barthes, Roland. "From Work to Text." In *Image/Music/Text*, 155–164. Translated by Stephen Heath. New York: Hill and Wang, 1977.

Beeston, Alix, and Stefan Solomon. "Pathways to the Feminist Incomplete: An Introduction, a Theory, a Manifesto." In *Incomplete: The Feminist Possibilities of the Unfinished Film*, edited by Alex Beeston and Stefan Solomon, 1–38. Berkeley: University of California Press, 2023.

Berliner, Lauren S. "Towards a Methodology of Unwatched Digital Media." *Feminist Media Histories* 8, no. 2 (2022): 219–230.

75 Allyson Nadia Field, "The Archive of Absence: A Manifesto for Looking at Lost Film," in *Uplift Cinema: The Emergence of African American Film and the Possibility of Black Modernity* (Durham, NC: Duke University Press, 2015), 23–28; Lauren S. Berliner, "Towards a Methodology of Unwatched Digital Media," *Feminist Media Histories* 8, no. 2 (2022): 219–230; Jane M. Gaines, "Never," in *Incomplete: Feminist Possibilities of the Unfinished Film,* ed. Alex Beeston and Stefan Solomon (Berkeley: University of California Press, 2023), 39–61; Alix Beeston and Stefan Solomon, "Pathways to the Feminist Incomplete: An Introduction, a Theory, a Manifesto," in *Incomplete: Feminist Possibilities of the Unfinished Film*, ed. Alex Beeston and Stefan Solomon (Berkeley: University of California Press, 2023), 1–38.

76 See Alex R. Galloway, "Golden Age of Analog," *Critical Inquiry* 48 (2021): 215, where he sees a paradox: the very age of computation is accompanied by the most analog of approaches – the serious study of affect, sensation, and contingency.

Biltereyst, Daniel, Richard Maltby, and Philippe Meers. "Introduction: The Scope of New Cinema History." In *The Routledge Companion to New Cinema History*, edited by Daniel Biltereyst, Richard Maltby, and Philippe Meers, 1–12. London: Routledge, 2019.

Borgman, Christine L. *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, MA: MIT Press, 2015.

Dang, Sarah-Mai. "Unknowable Facts and Digital Databases: Reflections on the Women Film Pioneers Project and Women in Film History." *Digital Humanities Quarterly* 14, no. 4 (2020). Accessed February 27, 2023. http://digitalhumanities.org/dhq/vol/14/4/000528/000528.html.

Davidson, Cathy, and David Theo Goldberg. "A Manifesto for the Humanities in a Technological Age." *The Chronicle of Higher Education* 50, no. 23 (February 13, 2004): B7–B9. Accessed February 28, 2023. https://www.chronicle.com/article/a-manifesto-for-the-humanities-in-a-technological-age/.

Dickel, Henri, Matija Miskovic, Karazm Noori, Christian Schmidt, Atefeh Soltanifard, Sarah-Mai Dang, Thorsten Thormählen. "Women Film Pioneers Explorer, 2021." Accessed June 30, 2023. https://www.online.uni-marburg.de/women-film-pioneers-explorer/index.html.

Drucker, Johanna. *Visualization and Interpretation: Humanistic Approaches to Display*. Cambridge, MA: MIT Press, 2020.

Field, Allyson Nadia. "Editor's Introduction: Acts of Speculation." *Feminist Media Histories* 8, no. 3 (2022): 1–7.

Field, Allyson Nadia. "The Archive of Absence: A Manifesto for Looking at Lost Film." In *Uplift Cinema: The Emergence of African American Film and the Possibility of Black Modernity*, 23–28. Durham, NC: Duke University Press, 2015.

Firestein, Stuart. *Ignorance: How It Drives Science*. New York: Oxford University Press, 2012.

Flückiger, Barbara, and Gaudenz Halter. "Methods and Advanced Tools for the Analysis of Film Colors in Digital Humanities." *Digital Humanities Quarterly* 14, no. 4 (2020). Accessed June 30, 2023. http://digitalhumanities.org/dhq/vol/14/4/000500/000500.html.

Gaines, Jane M. "Never." In *Incomplete: Feminist Possibilities of the Unfinished Film*, edited by Alix Beeston and Stefan Solomon, 39–61 (Berkeley: University of California Press, 2023).

Gaines, Jane M. "How Wrong We Were: Women in the Silent Era American Film Industry." American Film Institute. Accessed June 30, 2023. https://devaficalmjediwestussa.blob.core.windows.net/images/sites/3/2023/01/AFI-Women_Gaines.pdf.

Gaines, Jane M. "What Next? The *Historical Time* Theory of Film History." In *How Film Histories Were Made: Materials, Methods, Discourses*, edited by Malte Hagener and Yvonne Zimmermann, 59–84. Amsterdam: Amsterdam University Press, 2023.

Gaines, Jane M. "Counterfactual Speculation: What If Antonia Dickson Had Invented the Kinetoscope?" *Feminist Media Histories* 8, no. 3 (2022): 8–34. Accessed June 30, 2023. https://online.ucpress.edu/fmh/article/8/3/8/190671/Counterfactual-SpeculationWhat-if-Antonia-Dickson.

Gaines, Jane M. *Pink-Slipped: What Happened to Women in the Silent Film Industries?* Urbana: University of Illinois Press, 2018.

Gaines, Jane M. "What Happened to the Philosophy of Film History?" *Film History* 25, nos. 1–2 (2013): 70–80.

Gaines, Jane M. "Anonymities: Uncredited and 'Unknown' Contributors in Early Cinema." In *A Companion to Early Cinema*, edited by André Gaudreault, Nicholas Dulac, and Santiago Hidalgo, 443–459. London: Wiley-Blackwell, 2012.

Gallagher, Catherine. *Telling It Like It Wasn't: The Counterfactual Imagination in History and Fiction*. Chicago: University of Chicago Press, 2018.

Galloway, Alex. “Golden Age of Analog.” *Critical Inquiry* 48, no. 2 (2022): 211–232.
Gitelman, Lisa, and Virginia Jackson. “Introduction.” In *“Raw Data” is an Oxymoron*, edited by Lisa Gitelman, 1–4. Cambridge, MA: MIT Press, 2013.
Griffith, Erin and Cade Metz. “Tech Slump Doesn’t Slow New Boom in A.I. Field.” *New York Times* (January 7, 2023), B1, B4.
Halpern, Orit, Patrick Jagoda, Jeffrey West Kirkwood, and Leif Weatherby. “Surplus Data: An Introduction.” *Critical Inquiry* 48, no. 2 (2021): 197–210. Accessed June 30, 2023. https://www.journals.uchicago.edu/doi/full/10.1086/717320.
Hansen, Mark B. N. *Feed-Forward: On the Future of Twenty-First Century Media*. Chicago: University of Chicago Press, 2015.
Hartley, John. “Digital Scholarship and Pedagogy, the Next Step: Cultural Science.” *Cinema Journal* 48, no. 2 (2009): 138–145.
Hielscher, Eva. “The Phenomenon of Interwar City Symphonies: A Combined Methodology of Digital Tools and Traditional Film Analysis Methods to Study Visual Motifs and Structural Patterns of Experimental-Documentary City Films.” *Digital Humanities Quarterly* 14, no. 4 (2020). Accessed June 30, 2023. http://www.digitalhumanities.org/dhq/vol/14/4/000495/000495.html.
Hill, Kashmir. “With A.I. Appropriating Their Style, Some Artists Join the Resistance,” *New York Times* (February 19, 2023), 6.
Hong, Sun-ha. *Technologies of Speculation: The Limits of Knowledge in a Data-Driven Society*. New York: NYU Press, 2020.
Hoyt, Eric. “Arclights and Zoom Lenses: Searching for Influential Exhibitors in Film History’s Big Data.” In *The Routledge Companion to New Cinema History*, edited by Daniel Biltereyst, Richard Maltby, and Philippe Meers, 83–95. New York, London: Routledge, 2019.
Hoyt, Eric, Kit Hughes, and Charles R. Acland. “A Guide to the Arclight Guidebook.” In *The Arclight Guidebook to Media History and the Digital Humanities*, edited by Charles R. Acland and Eric Hoyt, 11–39. Falmer: REFRAME Books, 2016. Accessed June 30, 2023. http://projectarclight.org/book/.
Hu, Tung-Hui. *A Prehistory of the Cloud*. Cambridge, MA: MIT Press, 2015.
Jordheim, Helge. “Natural Histories for the Anthropocene: Koselleck’s Theories and the Possibility of a History of Lifetimes.” *History and Theory* 61, no. 3 (September 2022): 391–425.
Kinder, Marsha. “Medium Specificity and Productive Precursors: An Introduction.” In *Transmedia Frictions: The Digital, the Arts, and the Humanities*, edited by Marsha Kinder and Tara McPherson, 3–19. Berkeley: University of California Press, 2014.
Kinder, Marsha. “Designing a Database Cinema.” In *Future Cinema: The Cinematic Imaginary After Film*, edited by Jeffrey Shaw and Peter Weibel, 346–353. Karlsruhe, Germany: ZKM, 2003.
Kinder, Marsha, and Tara McPherson. “Preface: Origins, Agents, and Alternative Archaeologies.” In *Transmedia Frictions: The Digital, the Arts, and the Humanities*, edited by Marsha Kinder and Tara McPherson, xiii–xvii. Berkeley: University of California Press, 2014.
Koselleck, Reinhart. *The Practice of Conceptual History: Timing History, Spacing Concepts*. Translated by Todd Samuel Presner and others. Stanford, CA: Stanford University Press, 2002.
Kracauer, Siegfried. “Photography.” In *The Mass Ornament: Weimar Essays*, edited and translated by Thomas Y. Levin, 47–66. Cambridge, MA: Harvard University Press, 1995.
Kracauer, Siegfried. *History: The Last Things before the Last*. New York: Oxford University Press, 1969.
Liu, Alan. “Where Is Cultural Criticism in the Digital Humanities?” In *Debates in the Digital Humanities*, edited by Matthew K. Gold, 490–509. Minneapolis: University of Minnesota Press, 2012.
Mackenzie, Adrian. “The Production of Prediction: What Does Machine Learning Want?” *European Journal of Cultural Studies* 18, nos. 4–5 (2015): 429–445.
Manovich, Lev. *Cultural Analytics*. Cambridge, MA: MIT Press, 2020.

Manovich, Lev. “Database as Symbolic Form.” *Convergence* 5, no. 2 (June 1999): 80–99.
Marino, Mark. *Critical Code Studies*. Cambridge, MA: MIT Press, 2020.
McNeil, Joanne. *Lurking: How a Person Became a User*. New York: Farrar, Straus and Giroux, 2020.
McPherson, Tara. “In Focus: Digital Scholarship and Pedagogy.” *Cinema Journal* 48, no. 2 (2009): 119–123.
Miller, John David, and John Maeda. “A Stitch in Time: Visualizing History Through Unit Forms and Repetition Structures.” 2015. Accessed February 27, 2023. https://www.researchgate.net/publication/277250414_A_Stitch_in_Time_Visualizing_History.
Mukherjee, Siddhartha. *The Song of the Cell: An Exploration of Medicine and the New Human*. New York: Scribner, 2022.
Nietzsche, Friedrich. *The Use and Abuse of History* [1873]. New York: Cosimo Classics, 2010.
Petro, Patrice. “In the Wings.” In *Idols of Modernity: Movie Stars of the 1920s*, edited by Patrice Petro, 270–283. New Brunswick, NJ: Rutgers University Press, 2010.
Posner, Miriam, and Lauren F. Klein. “Editor’s Introduction.” *Feminist Media Histories* 3, no. 3 (2017): 1–8.
Salt, Barry. *Moving into Pictures: More on Film History, Style, and Analysis*. London: Starword, 2006. Accessed June 30, 2023. https://www.starword.com/index.html.
Schöch, Christof. “Big? Smart? Clean? Messy? Data in the Humanities.” *Journal of Digital Humanities* 2, no. 3 (2013): 1–13. Accessed February 28, 2023. http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/.
Wang, Yiman. “‘Speaking in a Forked Tongue’: Anna May Wong’s Linguistic Cosmopolitanism.” In *Revisiting Star Studies: Cultures, Themes, Methods*, edited by Sabrina Qiog Yu and Guy Austin, 65–82. Edinburgh: Edinburgh University Press, 2017.
Williams, Mark. “The Media Ecology Project: Collaborative DH Synergies to Produce New Research in Visual Culture History.” *Digital Humanities Quarterly* 15, no. 1 (2021). Accessed June 30, 2023, http://www.digitalhumanities.org/dhq/vol/15/1/000524/000524.html.
Yogerst, Chris. “Review of *Explorations in New Cinema History*.” *Journal of Cinema and Media Studies* 60, no. 3 (2021): 210.
Zeavin, Hannah. “A New AI Lexicon: EMPATHY.” AI Now Institute (September 16, 2021). Accessed February 27, 2023. https://medium.com/a-n32-ai-lexicon-empathy-4da12b82e280.

I Search

Adelheid Heftberger and Paul Duchesne

Finding Female Film Editors in Wikidata: How to Query and Visualize Filmographic Records

Finding Data about Female Contributors to Film Productions

> A common assumption about the German *Filmkleberinnen* [gluers] was that they had a low income and they were regarded as low ranking within the film producing industry. It may well be that "film splicing could make only a very limited creative contribution to storytelling in the earliest days of film, when the job was more or less to join the shots end to end. However, that does not excuse the glaring oversight or possibly even motivated blindness of the historiography of film: that [. . .] it involves creative decision-making."[1]

> Why are editors often invisible? A possible explanation may include "the complexities buried in the multiple names given to the job as it evolved, and the tendency in the discussion of the continuity style to evaluate good editing as 'invisible'."[2]

Let's look at our initial example of film gluers. Judging by oral testimonies, the assumption that this job was regarded as a lowly duty might be a myth, probably based on a contemporary perspective that was influenced by the *auteur cinema* in which the director is the only creative (thus influential) person. In DIE KLEINEN KLEBERINNEN (Eva Maria Hammel & Heide Breitel, 1980) the elderly women talk animatedly about their careers, their craft, and about the pride they still feel as valued and well-paid film workers. Even though they had signed contracts (e.g., for the Decla, a German production studio which existed before the UFA), they worked for other studios as well, and sometimes even from their own homes, thus increasing their salaries even further. Because they were needed as experienced and skillful negative cutters, nobody really challenged this practice. In the same interviews, the women stress that glueing film was far from the only job they knew; they also printed and developed film for their own needs if necessary.

Enriching our traditional filmographies and enhancing our knowledge about historical (and contemporary) film production has certainly been put on the

1 Karen Pearlman and Adelheid Heftberger, "Editorial: Recognising Women's Work as Creative Work," *Apparatus. Film, Media and Digital Cultures of Central and Eastern Europe* 6 (2018), https://doi.org/10.17892/app.2018.0006.124.

2 Pearlman and Heftberger, "Editorial."

https://doi.org/10.1515/9783111082486-003

agenda of film scholars and film historians worldwide. In the last couple of years, pivotal books have been published about female employees in Hollywood by J. E. Smyth[3] and women in British film production by Melanie Bell,[4] as well as research projects like *STUDIOTEC: Film Studios: Infrastructure, Culture, Innovation in Britain, France, Germany and Italy, 1930–60* and research generally that challenges the notion of a predominantly male and national film history by taking emigrés and female film workers into account. These publications have used archival and private sources as well as oral history, thus developing a different methodology for their research needs and addressing questions unanswered by known records. They show that "the myths surrounding women's work as unskilled or biologically determined have been compounded by established film historiographies" and need to be challenged by unearthing more data.[5] The "cherished notion in Western culture of the romantic artist as individual genius"[6] is being discussed, which may have huge implications for how we credit authorship in our databases. Some go as far as challenging the whole structure: "Perhaps asking who should get credit for one or another part of this integrated creative activity is asking the wrong question. What we need to be asking is: what is creative thought in a distributed cognitive system?"[7] Studies like these stress that looking at female careers needs a different methodology as well, because the norm of "continuous work histories" with an ongoing filmmaking record in one studio can't be applied here to assess success in the traditional sense.[8] As Bell underlines, many women's working lives "were shaped by child care and domestic responsibilities," resulting in "episodic waves of working, where career and family alternatively" took center stage.[9] These experiences

3 See J. E. Smyth, *Nobody's Girl Friday: The Women Who Ran Hollywood* (New York: Oxford University Press, 2018).

4 Melanie Bell, *Movie Workers: The Women Who Made British Cinema* (Urbana, Chicago and Springfield: University of Illinois Press, 2021), accessed April 4, 2024, https://www.jstor.org/stable/10.5406/j.ctv1rnpjzz.

5 Bell, *Movie Workers*, 6.

6 Bell, *Movie Workers*, 6.

7 Karen Pearlman, John MacKay, and John Sutton, "Creative Editing: Svilova and Vertov's Distributed Cognition," *Apparatus. Film, Media and Digital Cultures of Central and Eastern Europe* 6 (2018), https://doi.org/10.17892/app.2018.0006.122; Karen Pearlman and John Sutton, "Reframing the Director: Distributed Creativity in Filmmaking Practice," in *A Companion to Motion Pictures and Public Value*, ed. Mette Hjort and Ted Nannicelli (Hoboken, US: Wiley-Blackwell, 2022), 86–105, https://doi.org/10.1002/9781119677154.ch4.

8 Bell, *Movie Workers*, 7.

9 Bell, *Movie Workers*, 7.

seem to be similar across cultural and geographic distances, as suggested by Nadita Dutta's account of Indian women filmmakers,[10] compared with those in Britain discussed by Bell, and in the U.S. by Smyth.

For Bell, the absence of women in film history needs to be addressed with a fitting methodology where, first of all, the concept of the episodic (interrupted) career must serve as a lens "through which to recover women's occupational labor." Secondly, women's achievements need to be contextualized within a climate of discrimination; lastly, a more flexible, inclusive model of creativity is needed "to accommodate the many and varied tasks women undertook in the performance of their professional duties in below-the-line roles."[11] Of course we also need to take national and studio traditions into account and a great deal more work needs to be done there in order to be able to write a more comprehensive history of women in film production. Sometimes the careers in different countries and studios were exactly the opposite: while women only slowly acquired more significant roles in the British film and television industry in the 1950s and 1960s, for example, they started to disappear from important positions in Hollywood studios at the end of the 1950s due to the decline of the studio system.[12]

One persistent myth can be refuted: that women were just not present in early film production and that is why we don't know their names. Thanks to scholars like Smyth we now know that Hollywood women, for example, actually played a major role, with many of them having influential positions and able to bring their vision and creativity to the screen. Other scholars have looked into Russian film history and unearthed early female film producers and cinema owners particularly.[13] Smyth argues that it is not some sort of law of nature that these names disappear from film history: "However much Hollywood acted as an advocate for working women during the heyday of the studio system, male historians tended to edit out women's roles, leaving them on the cutting room floor of conventional film histories."[14] She cites authors like Terry Ramsaye, Benjamin Hampton, Lewis Jacobs, Leo Rosten, and Arthur Knight, who "all focused on a progressive historical model founded on technological,

10 Nandita Dutta, *F-Rated: Being a Woman Filmmaker in India* (Uttar Pradesh: Harper Collins Publishers India, 2019).

11 Bell, *Movie Workers*, 8.

12 Smyth, *Nobody's Girl Friday*, 11.

13 Natascha Drubek, "Hidden Figures: Rewriting the History of Cinema in the Empire of All the Russias," *Apparatus. Film, Media and Digital Cultures of Central and Eastern Europe* 13 (2021): 109–144, accessed January 7, 2023, https://doi.org/10.17892/app.2021.00013.284; Peter Bagrov and Anna Kovalova, "Elizaveta Thiemann," *Academic Commons* (2021), https://doi.org/10.7916/d8-hnnd-rk78.

14 Smyth, *Nobody's Girl Friday*, 13.

artistic, and commercial innovation dominated by a male cast of producers and directors."[15] Natascha Drubek argues in similar ways about how film workers who were not Russian, but had an emigrant background, were actively left out.[16] More and more thorough research is needed to get an increasingly comprehensive picture on how film historiography was written for different countries and for marginalized groups. It needs to be recognized that there is possibly a parallel with how history in general is written by the dominant group and based on patriarchal and capitalist notions of success and creativity, as Bell points out.

How then to find data on female film workers? Were they actively forgotten? Were they ever recorded? Where should we start looking? As Sarah-Mai Dang points out, the awareness of missing sources has become a key issue of feminist debate:

> How can we identify and include blind spots when trying to reconstruct the past? In what way can we narrate ellipsis and absences while avoiding the pitfall of implicitly promising to grasp the "whole" story once enough information will be gathered? How is it possible to explore uncharted territory when faced with the lack of historical objects?[17]

On the other hand, more sources than ever have been digitized and put online. There is an imminent risk, however, that, by focusing on online sources, we neglect the plentitude of diverse and most likely unrecorded offline sources[18] – information that stems from viewing the moving images themselves, for instance, which so often have not been preserved and made accessible. In short, only by archivists and researchers joining forces, to create more sources and analyze the information (for example, deciphering pseudonyms which were frequently used by women), can we gather more data for our investigations and draw conclusions about the involvement of female film workers.

Women in film history are not easy to locate in film archives' databases and available filmographies for several reasons. Not many databases offer the option to distinguish their cast and credits information into at least a binary distinction of male/female. And even if there was the option, it arguably didn't feel mandatory until recently. What would a film archive do with such information? The project *BFI Filmography* provided a novel view on filmographic data by considering their metadata as something that could and should be mined and analyzed as a valuable corpus of data. Probably not many outside film archives can under-

15 Smyth, *Nobody's Girl Friday*, 13.

16 Drubek, "Hidden Figures."

17 Sarah-Mai Dang, "Unknowable Facts and Digital Databases: Reflections on the Women Film Pioneers Project and Women in Film History," *Digital Humanities Quarterly* 14, no. 4 (2020), accessed April 4, 2024, http://digitalhumanities.org/dhq/vol/14/4/000528/000528.html.

18 Dang, "Unknowable Facts and Digital Databases."

stand how pioneering (and costly) this project was. It was a political and strategic decision to dedicate resources (both human and technical) to enriching their data with gender information by combining the film archives' data (in that case forenames) with "available gendered" forenames on the internet. Only then was the BFI able to show the gender distribution in British film production. A process like this will not be without errors, but it supports intellectual, manual work by catalogers.

There is no space to elaborate further on this, but it is worth stressing, that "data is trimmed or transfigured to match the expectations of the machine," what can be called "schematic bias":[19]

> For those involved in building data systems, a schema is a kind of blueprint, a map of which types of information will be stored, in what form, and which types of information will be rejected. In cognitive science, a schema is a pattern of thought, a framework of preconceived ideas that directs how a person sees the world: if you observe something that fits neatly into your schema, it gets filed easily and efficiently into your memory. On the other hand, schema-foreign things will often not be noticed or remembered, or they will be modified to fit into what you expect based on the frameworks you constructed.[20]

We need to keep these thoughts in mind when talking about how we organize information, what we leave out for various reasons, and where we can find as yet unavailable information.

Collecting and Working with Data

Databases of film archives are – alas – not always the most available or accessible sources. Another caveat is that often cataloguers either take their cast and credits information from secondary sources (which should not, however, be sneered at, because they can also complete fragmented physical elements), or have to fit the data visible on the print/file itself into a rigid information framework. Complementary action would involve viewing the films themselves, especially those at the fringes of industrial film production, like documentaries, the German genre of *Kulturfilme*, amateur films, experimental films, etc. Another option would be to carry out formal analysis to find out more about stylistic conventions, studio style, or even detecting gendered styles to enrich existing data. These highly spec-

19 Jer Thorp, *Living in Data: A Citizen's Guide to a Better Information Future* (New York: MCD, Farrar, Straus and Giroux, 2021), 131.
20 Thorp, *Living in Data*, 132.

ulative, but nevertheless fascinating investigations might back up longstanding film theories on female involvement. Or they might shatter our understanding of how to describe a "female" style. By pursuing the quest for female names in film production, we can test our hypotheses about what we want to uncover/find. Do we think that women had a unique voice? Is their contribution distinguishable from the contribution of their male co-workers? Or do we rather just want to close knowledge gaps, no matter whether male or female? As film archivists our preliminary answer might be – again – a pragmatic one at least initially. Collecting existing data would serve as a first step to connect already digitized sources and, by comparing and mapping them, at least complete some records.

The pioneering publications mentioned earlier relied on data sources that were mostly mined manually. It is worth asking how the unearthed names and pieces of information are stored again for others to analyze and use for their own research questions. These infrastructures rarely exist within academia. The "Women Film Pioneers Project" offers a wealth of information, albeit in the form of essays. If we, however, need to query, compare, and visualize the existing data, we need something more like a database. Of course, as mentioned above and summarized elsewhere,[21] databases themselves are not neutral collections of knowledge.

Dang points out that there is an increasing risk "of letting offline sources be consigned to oblivion" because of the easily and immediately available sources on the internet.[22] There are many data collections which are not yet digitized and data which, especially for our topic, can only be created by scholars to be later (if ever) provided online. Some valuable data can be found in, for example, archives' databases and cataloguing notes not yet transferred, filmographic sources and other aggregated online databases like Wikidata, historical film journals, memoirs and interviews, studio archival collections, guild reports, trade papers, local reportage, nationally syndicated journalism, press publication, and trade union records.

For this chapter we will focus on Wikidata and evaluate what information is available via the platform, the provenance of these statements, and the data gaps. We have chosen Wikidata as it is currently the most prominent and largest knowledge graph openly available to cultural data researchers, and features extremely permissive data licensing. We will concentrate on female film workers and on data regarding female editors particularly. Before we do this, let's look at how gender is assigned in Wikidata and how reliable this information is.

21 Dang, "Unknowable Facts and Digital Databases."
22 Dang, "Unknowable Facts and Digital Databases."

Classification of Male/Female and How Wikidata Deals with This

Traditionally, in historical data sources and probably many times still today, we can find a binary structure dividing people/agents into male or female. This view is at best simplistic, as feminist theorists, writers, and activists like Simone de Beauvoir, Julia Kristeva, Luce Irigaray, Michael Foucault, and Judith Butler have pointed out and deconstructed. Early on, one guiding thought was that it is necessary to distinguish sex from gender: while the first is biologically determined, the second is socially constructed. However, we as a (Western) society have come to learn (arguably slowly), that the bi-polar man-woman model does not do justice to the diversity of human bodies:

> In biology and medicine, sex determination takes place in four stages [. . .]: 1. genetic sex (xx-xy chromosome model), 2. gonadal sex (gonadal sex, ovaries versus testes), 3. body sex (internal and external sexual organs such as vagina, clitoris, penis, etc.), 4. hormonal sex (female versus male hormones).[23]

Heike Wiesner states clearly that the fact that our society generally only wants to see two sexes is a social construction as well.[24]

There is no doubt that the subject is complex. However, if we want to take a pragmatic approach in order to at least start the process of analysis, the binary classification can be helpful. We need to be pragmatic because in many cases, we won't find more differentiated data, and probably, as a film archive, never will be able to produce complete and accurate datasets retrospectively. If the archives' databases or other data sources have any gender classification at all, the legacy data is in most cases binary. Even if one wants to enrich data afterwards, e.g., by assigning first names to provide a gender via external sources, the result remains an approximation. The best answer to overcoming the binary bias would be to have film workers classify themselves, as is already technologically possible on platforms such as Wikipedia and Wikidata. It should be flagged, however, that in practice this self-declaration could be less straightforward than expected, as when Philip Roth found himself unable to modify a statement written without providing "secondary sources" even though it was about the inspiration of a book he himself had writ-

23 Heike Wiesner, *Die Inszenierung der Geschlechter in den Naturwissenschaften: Wissenschafts- und Genderforschung im Dialog* (Frankfurt, New York: Campus-Verlag, 2002), 220.
24 Wiesner, *Die Inszenierung der Geschlechter in den Naturwissenschaften*, 221.

ten.[25] A promising initiative in this direction is the SOLID project,[26] led by Sir Tim Berners-Lee, where a user controlled datastore allows individuals to declare data about themselves and selectively allow access from third-parties, as opposed to the current paradigm where an individual has personal (and possibly conflicting) data spread across many different databases and platforms.

We argue that being pragmatic about working within existing limitations can be an option, but it needs to be tightly linked to being transparent about our methods and our underlying ontologies and classifications. Data collection and creation ought to be critically discussed, as well as our vocabularies, search options, training data for AI projects, etc. – even if more differentiated data proves difficult to find, for example, because an individual involved in film production might find certain information too personal to share. But we can allow for more options in our databases and ontologies because we don't want to make invisible crucial data that could be relevant for cultural/political decisions. "'What gets counted counts', feminist geographer Joni Seager has asserted, and [Maria] Munir is one person who understands that. What is counted – such as being a man or a woman – often becomes the basis for policymaking and resource allocation."[27] The BFI, for example, stated clearly that their analysis of credits and their distribution over male/female members of staff in film production "can lead to an increase in data-driven policy development, including outreach and engagement with film training and education providers, to identify departments and roles where women are under-represented, and advocate and encourage for greater diversity."[28]

Data relevant to this conversation can be found in significant quantities on Wikidata, under the property "sex or gender" (P21). This concatenation of two attributes which could be ontologically distinguished has been criticized from within the community, although currently upheld as being consistent with language used by various governments (for example, the United Kingdom).[29] When applied to "humans" (Q5) there are seven recommended "sex or genders" on Wikidata: "male" (Q6581097), "female" (Q6581072), "non-binary" (Q48270), "intersex"

25 Philip Roth, "An Open Letter to Wikipedia," *The New Yorker* (September 6, 2012), accessed April 4, 2024, https://www.newyorker.com/books/page-turner/an-open-letter-to-wikipedia.

26 See https://solidproject.org/, accessed April 4, 2024.

27 Catherine D'Ignazio and Lauren F. Klein, *Data Feminism* (Cambridge, MA: MIT Press, 2020), 97.

28 See https://www.bfi.org.uk/bfi-national-archive/search-bfi-archive/bfi-filmography/bfi-filmography-project-overview, accessed April 4, 2024.

29 See https://www.wikidata.org/wiki/Property_talk:P21#UK_government_combines_sex_and_gender_in_forms, accessed April 4, 2024.

(Q1097630), “trans woman” (Q1052281), “trans man” (Q2449503), and “agender” (Q505371). Note that this is not strictly a controlled vocabulary, as Wikidata allows for the entry of any data statement, which are only policed retroactively, and there are many more terms which have been applied in this context.

Queries pertaining to all “humans” on Wikidata are no longer possible, as the query service can no longer process a single request on this scale. Instead, we will look at a specific subset of “humans”: “film editors” (P1040), to understand where “sex or gender” labels are applied, whether they are referenced, and what the sources of these references are. Our definition of “film editors” are those who are “human” (Q5) and have a statement of “film editor” (P1040) against an entity that is an “instance of” (P31), or a “subclass of” (P279), a “film” (Q11424).

It should be noted that of the 360,654 films matching the above definition currently recorded in Wikidata, only 52,862 (or 15%) contain one or more editor credits. This is a relatively low representation compared to credits for director (281,371 or 78%), indicative of the prioritization of perceived primary credits roles, but also due to the director credits being used as a common method for entity disambiguation (e.g., Psycho “1960 film by Alfred Hitchcock”[Q163038] against Psycho “1998 American film by Gus Van Sant” [Q979196]). This means that any further conclusions regarding editors are drawn from a small slice of actual film contributions, and we should ask what factors have led to this specific dataset being available to us.

The 52,862 extant editor credits are the work of 10,296 distinct individual editors, of which 10,108 (or 98%) contain one or more “sex or gender” data statements (see Figure 1). Six editors have been attributed multiple “sex or gender,” but it is not clear whether this is an intentional attempt at attributing non-binary status, or data error.

Figure 1: Example of “sex or gender” (P21) data statement for “Agnès Varda” (Q229990).

Statements in Wikidata can also be supported by reference links, to provide provenance as to the source of the claim beyond being simply an assertion. This is an important attribute as Wikidata is primarily a crowd-sourced data platform, and, while references may be subject to link-rot or seldom verified, it is a good start towards incorporating data authentication into the platform (see Figure 2).

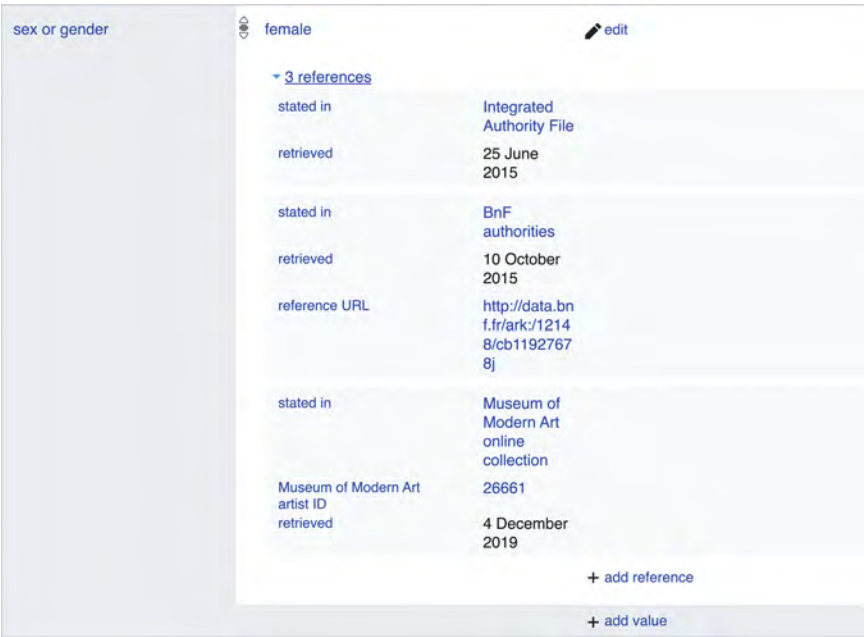


Figure 2: Example of references for "sex or gender" (P21) data statement for "Agnès Varda" (Q229990).

Of the 10,115 "sex or gender" statements pertaining to film editors, only 4,098 (or 41%) contain one or more provenance references. Interestingly, there is a noticeably higher rate of "male" (Q6581097) film editors with references to back up their "sex or gender" statements than "female" (Q6581072) (see Figure 3). All statements related to the "sex or gender" of non-binary film editors are referenced.

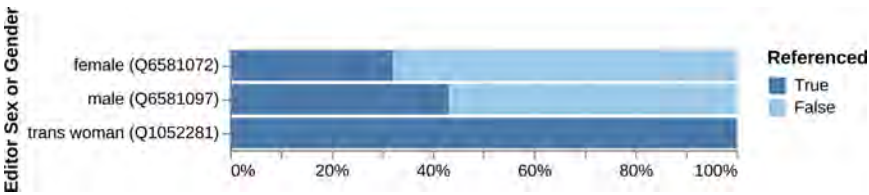


Figure 3: Percentages of "sex or gender" (P21) values referenced to one or more sources.

Let us look more closely at the sources themselves for these references, and how they are derived. The vast majority of these statements are made as either "stated in" (P248) or "imported from Wikimedia Project" (P143). Taking the exact source of these derivation references produces the chart included here in Figure 4.

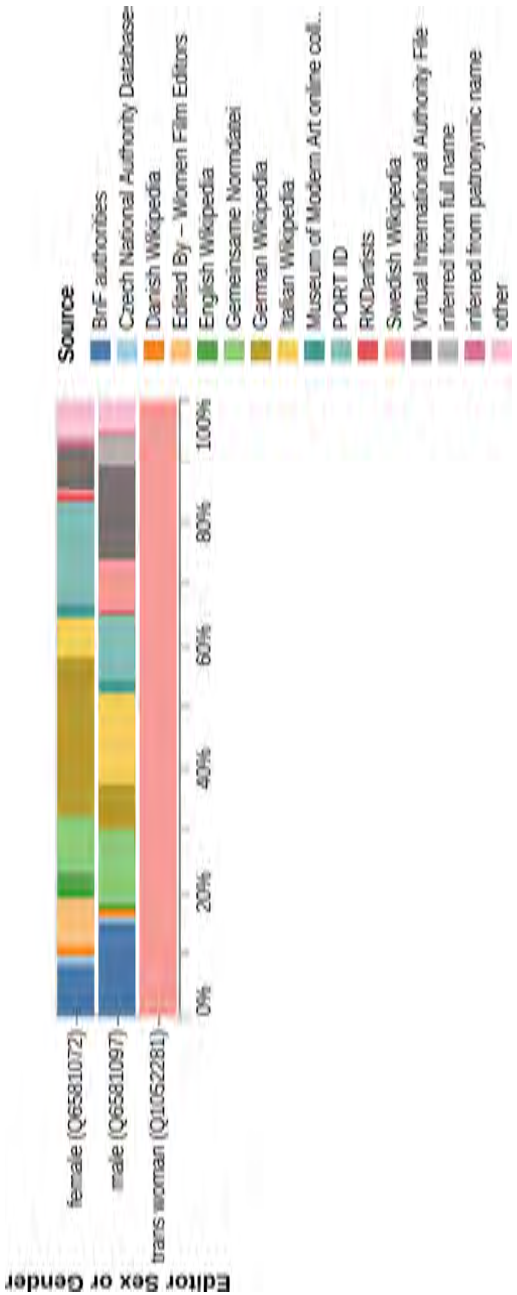


Figure 4: Reference sources for "sex or gender" (P21) values.

Some references for "sex or gender" statements come from sources we deem fairly reliable, such as national authority files, for example the *Gemeinsame Normdatei* (GND) and the Bibliothèque nationale de France (BnF). Wikipedia is also a common source for information, although the exact transformation path of free-text into data statements is unclear. Figure 5 shows the same graph, but with the sources broadly categorized.

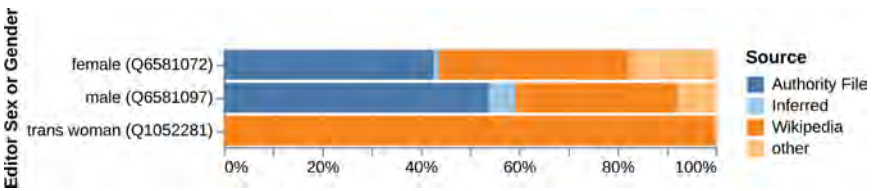


Figure 5: Reference sources for "sex or gender" (P21) grouped into general categories.

A minority of statements are generated by "inferred" data, which is the process of determining "sex or gender" information based only on name analysis, the same approach used for the BFI Filmography project. This involves making calculated assumptions about the correlation between certain names and "sex or gender" associations, and cannot provide any insight into identifying non-binary individuals.

Queries and Data

Having looked at the language and provenance of "sex or gender" data found through Wikidata, we will now shift focus to look at leveraging them in relation to other data available on the platform, and critically discuss the results.

One of the most obvious queries to execute in this context is to compare the ratio of female film editors to the "Country of Origin" (P495) of the film. We have filtered out low counts per country to guarantee a representative sample for each, yet we see something that may immediately appear interesting in the greater context of this chapter: the almost complete dominance of female film editors in the German Democratic Republic (Q16957) (see Figure 6).

What to make of this? Would this be confirmed by those who were working in the industry during these years? Or are we seeing a distortion due to only certain prominent GDR female film editors having been added to Wikidata? While a complete history of female editors of the GDR still remains to be written, we can find references in older publications, where editing staff is mentioned as exclusively female:

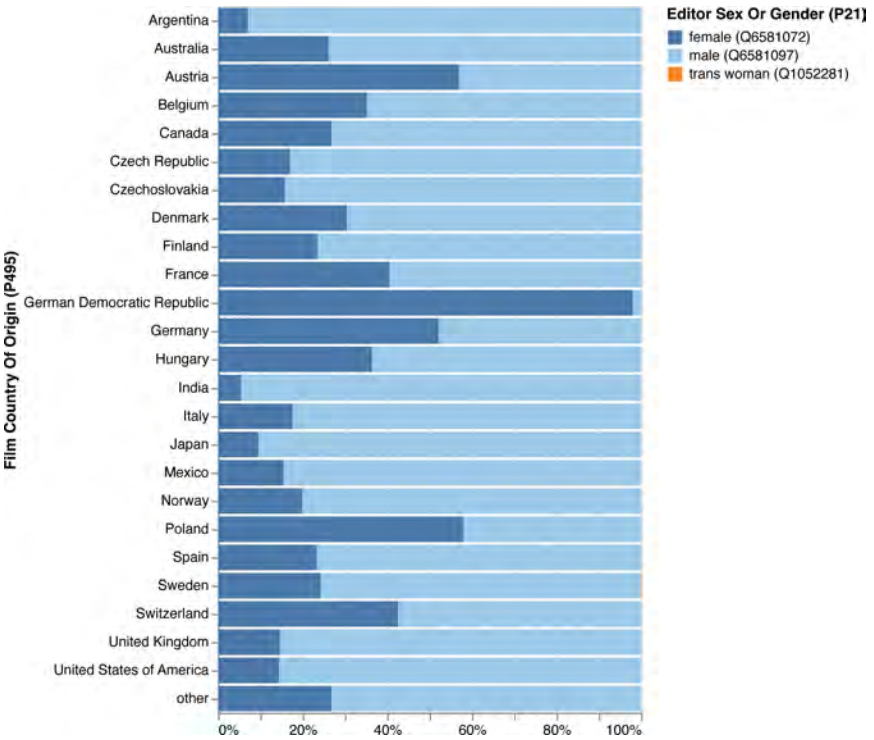


Figure 6: "sex or gender" (P21) of "film editor" (P1040) by the "country of [film] origin" (P495).

> Final production includes editing, dubbing, music and sound recording, and mixing. The editing supervisor [*Schnittmeisterin* in German] is responsible for managing these operations. Here, too, proven teams have emerged at DEFA. Every director knows how much his film depends on good cooperation with the editor [*Schnittmeisterin*]. [. . .] At or shortly after the start of shooting, an editor [*Schnittmeisterin*] and an assistant editor [*Schnittassistentin*] are available for production.[30]

As has been mentioned previously, this raises a core problem when using Wikidata as a sole basis for research and analysis, in that it does not claim to be fully representative (as a national filmography would), but is rather populated with whatever data individuals are interested in populating it with.

30 Dirk Jungnickel, "Produktionsbedingungen bei der Herstellung von Kinospielfilmen und Fernsehfilmen," in *Filmland DDR. Ein Reader zu Geschichte, Funktion und Wirkung der DEFA*, ed. Harry Blunk and Dirk Jungnickel (Cologne: Verlag Wissenschaft und Politik, 1990), 55.

However, we can continue to explore correlations that we might superficially expect to coincide with the presence of a female film editor. It should be noted that none of these metrics relate to any formal analysis of the films themselves (e.g., shot duration, image, or sound characteristics), as such granular data is not currently available to us via Wikidata.

The "Bechdel Test" (Q4165246) is a well-known metric for measuring the agency of female characters in film works, and surprisingly well populated in Wikidata (see Figure 7). Graphing against film editor "sex or gender" does show a weak correlation, although anecdotally we are of course aware that some of the most celebrated female film editors ("Thelma Schoonmaker"[Q166887] or "Anne V. Coates" [Q31294]) made notable contributions to films which feature barely any or no female protagonists.

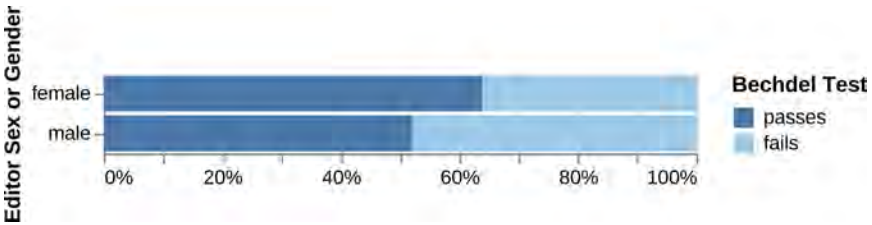


Figure 7: The "sex or gender" (P21) of "film editor" (P1040) against the film "Bechdel test" (Q4165246) score.

There is a greater correlation to be found between female film editors and female film directors, although further exploration would be required to confirm whether these individuals are in fact distinct, and the results are not being distorted by one individual working in both capacities (see Figure 8).

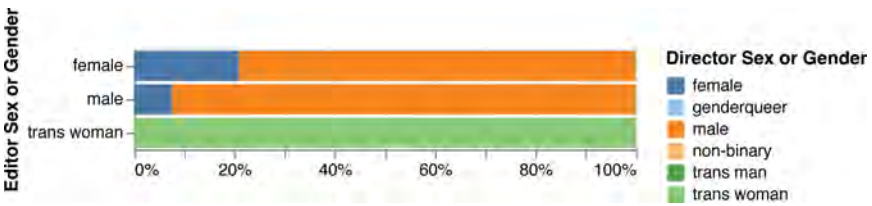


Figure 8: The "sex or gender" (P21) of the "film editor" (P1040) against the "sex or gender" (P21) of the film "director" (P57).

Lastly, we can talk about the information which is held against the film editors themselves. Here we have graphed the most popular data attributes, and the fill rate. As in the previous discussion, there is an almost complete representation of "sex or gender" (P21) data, as well as some other expected bibliographic data points: "occupation" (P106), "given name" (P735), "date of birth" (P569), "country of citizenship" (P27), "family name" (P734), and "place of birth" (P19) (see Figure 9).

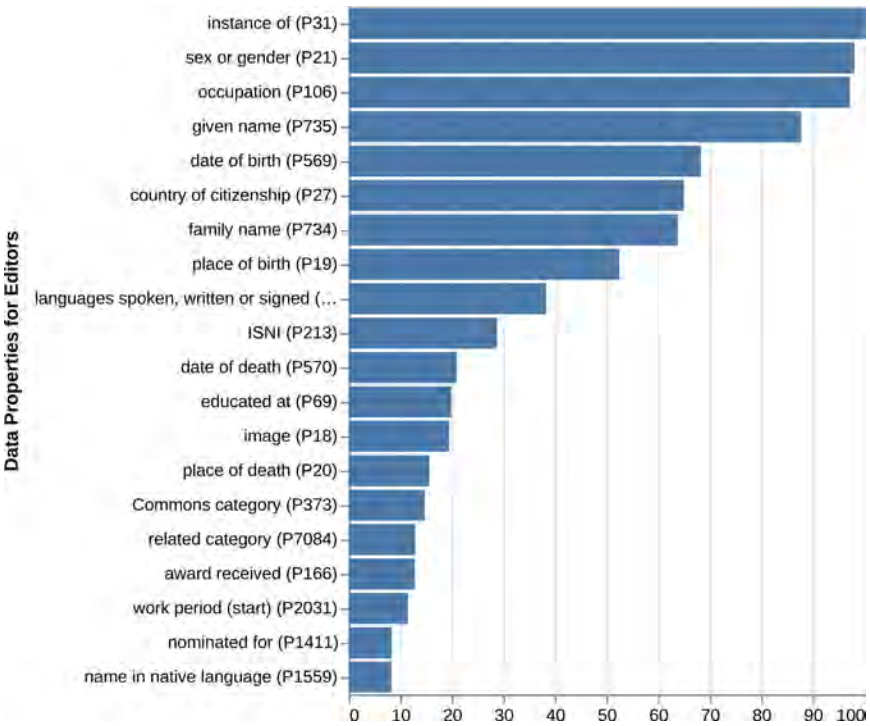


Figure 9: Other available data properties for "film editor[s]" (P1040).

It is worth pointing out that there is a quick decrease in widely available data, with only the top eight data points being provided for more than 50% of film editors. Unusual data that can be found in only few instances include: "dance style" (P10741), "hairstyle/hairlength" (P8839), "had as last meal" (P3902).

Discussion and Outlook

Interpreting our findings is not easy and cannot be based entirely on the data retrieved from Wikidata. First of all, it relies on crowd-sourced data added by individuals and institutions with different motivations and we can safely assume that they are rarely striving for completeness of any given metric. However, we started from the position that it would be worth looking at the data Wikidata provides as an open data source, in order to discover what information we can derive from it for our questions relating to the quantitative distribution of female editors and correlations within international film production. We will summarize our thoughts and preliminary results under the following headers: national filmography, data gaps in film history, bias of focusing on one role only, and bias of overrating female influence on the final work.

Wikidata does not tell us anything about how many film works exist, thus we have no reference point for our queried statistics. It might be a very theoretical question, but how would we be able to achieve that basis – not currently provided by Wikidata – for data analysis? It is tricky, but not impossible. But, alas, the task of writing a national filmography (in the best sense) is in most countries not assigned to a single institution; sometimes it is not assigned to any institution at all. Furthermore, unlike book publications, there are no comparable identifiers like ISBN for film works, which would allow for some statistical data source. Some countries achieve better results than others because of a more centralized structure (France for example), mostly when funding agencies, film production, and the archive are working together to create meaningful numbers. Wikidata remains a valuable resource, because it is growing and there is always the chance that more data will be contributed by institutions. It might therefore be advisable to monitor the statistics and repeat the queries from time to time. Nonetheless, for further research it would be advisable to include other data sources like film archives' data and/or aggregated film-related data (as from Europeana[31] or European Film Gateway[32]). It might also be advisable to analyze national datasets first and then query datasets according to nationality, although obviously the notion of "national film heritage" has its pitfalls as well, and must probably be viewed just as pragmatically as sex/gender data.

We are well aware of the fact that, in many cases, the people (male or female) who actually worked on a film production, and many times in an important creative or decision-making capacity, are not mentioned for various reasons. Thus,

31 See https://www.europeana.eu/, accessed April 4, 2024.

32 See https://www.europeanfilmgateway.eu/, accessed April 4, 2024.

unfortunately, it is not easy to find reliable numbers for film production since the birth of cinema. Some countries were better at compiling (national) filmographies than others. Thus, we should consider the numbers of women on Wikidata, as on every other data source that has been involved in filmmaking, with the necessary caution. As stated at the beginning of our chapter, archivists can first collect information from different sources and compile it for researchers and anyone interested to use. Archives could do more in looking at the primary object and be more transparent about their data sources as well, by being clear, for example, whether the name and role was taken from what appears on the film print, or from a secondary source. Wikidata's ability to include provenance information for each one of its statements in some ways puts it *ahead* of many archival cataloguing systems, where data is declared "as is" with no context as to where it was sourced or the method by which it was derived. In many cases we won't find the creators of our film history, but with qualitatively better datasets and critical discussion, based on knowledge of historical film production, we can still contribute to film historiography, whether it is from a feminist perspective or not. Our observation would be that it is too much to ask that Wikidata provides meaningful data for every period of film history, and it may be more helpful when it comes to contemporary film production.

We have focused on available data around female editors and what could be found on Wikidata, but let's not forget that filmmaking is a collaborative practice. As Karen Pearlman reminds us, filmmaking is the result of "distributed creativity."[33] There is a danger that by only looking at certain roles for female film workers, we lose sight of that fact; we need to consider the creative process heuristically in order to understand more about the influences and decision making. Wikidata will not be helpful when it comes to more differentiated analysis, because it is – after all – a database which reduces complex circumstances like every other database. Network analysis thus needs to be treated with caution and must come with a critical discussion of the data context. Assigned "roles" tell us only so much about the actual work of the individual person, which might vary enormously in different studio production contexts. It is one of the latest achievements of film historiography that we question these job descriptions and see beyond the credit lines. A good example is the story behind Penny Eyles' contribution as a script supervisor, presented by Wendy Russell at a FIAF workshop.[34] As archivists, we still need to dis-

33 Pearlman and Sutton, "Reframing the Director."

34 Wendy Russell, "The Materiality of an Archival Object: A Reading of Penny Eyles' Continuity Script for *Kes*" (presented at So Much More Than Non-Film: Cataloguing Film-Related Materials, Swedish Film Institute, Stockholm, June 2023), accessed April 4, 2024, https://www.fiafnet.org/images/tinyUpload/2023/06/Kes_Continuity_Script_-_WR.pdf.

cuss our definitions, adjust our vocabularies, and perform close readings of specific projects, accompanied by oral history in order to understand the working environment. Data analysis can give us hints and provide statistics to start the investigation at meaningful points.

Finally, does the fact that I am female (or male or non-binary) really influence the final film in a foreseeable way? And if so, in what way would that be? If we tried to formulate theories on how an *écriture féminine* might manifest itself, would that not run the risk of perpetuating stereotypes? How much does the cultural background, the budget, and the power structures in film production influence the film in the end? Again, not to repeat ourselves, we did not claim that data analysis would necessarily give us more insight on questions that we just cannot find in binary (or at least highly structured) data. However, starting with data can be a significant and useful first step to understand what can be retrieved and where the historical gaps are. Just the banal fact that, in a fairly large open crowd-sourced dataset like Wikidata, there are only x% of a certain kind of data, raises questions about the general availability of data. It might lead to a situation where archives consider contributing more to fill data gaps or provide more comprehensive datasets themselves on other platforms. We can start by discussing where data gaps might originate. We might also be surprised by how varied the statements for "sex and gender" are and, as archivists, compare it to our vocabularies. Nuanced automated textual analysis, beyond simple pronoun detection, could also enable the ability to apply complex data statements. Film archivists will probably read our contribution differently from scholars. While some might be more interested in where the data comes from, others might want to compare the data with other sources collected so far. Generally, trying to work out what can be found on Wikidata and which queries could be useful has proven an interesting process and might influence the way we as archivists could query our own databases and domain data sources into the future.

References

Bagrov, Peter, and Anna Kovalova. "Elizaveta Thiemann." *Academic Commons* (2021). https://doi.org/10.7916/d8-hnnd-rk78.

Bell, Melanie. *Movie Workers: The Women Who Made British Cinema*. Urbana, Chicago and Springfield: University of Illinois Press, 2021. Accessed April 4, 2024. https://www.jstor.org/stable/10.5406/j.ctv1rnpjzz.

Dang, Sarah-Mai. "Unknowable Facts and Digital Databases: Reflections on the Women Film Pioneers Project and Women in Film History." *Digital Humanities Quarterly* 14, no. 4 (2020). Accessed April 4, 2024. http://digitalhumanities.org/dhq/vol/14/4/000528/000528.html.

D'Ignazio, Catherine, and Lauren F. Klein. *Data Feminism*. Cambridge, MA: MIT Press, 2020.

Drubek, Natascha. “Hidden Figures: Rewriting the History of Cinema in the Empire of All the Russias.” *Apparatus. Film, Media and Digital Cultures of Central and Eastern Europe* 13 (2021): 109–144. https://doi.org/10.17892/app.2021.00013.284.

Dutta, Nandita. *F-Rated: Being a Woman Filmmaker in India*. Uttar Pradesh: Harper Collins Publishers India, 2019.

Jungnickel, Dirk. “Produktionsbedingungen bei der Herstellung von Kinospielfilmen und Fernsehfilmen.” In *Filmland DDR. Ein Reader zu Geschichte, Funktion und Wirkung der DEFA*, edited by Harry Blunk and Dirk Jungnickel, 47–57. Cologne: Verlag Wissenschaft und Politik, 1990.

Pearlman, Karen, and Adelheid Heftberger. “Editorial: Recognising Women’s Work as Creative Work.” *Apparatus. Film, Media and Digital Cultures of Central and Eastern Europe* 6 (2018). https://doi.org/10.17892/app.2018.0006.124.

Pearlman, Karen, John MacKay, and John Sutton. “Creative Editing: Svilova and Vertov’s Distributed Cognition.” *Apparatus. Film, Media and Digital Cultures of Central and Eastern Europe* 6 (2018). https://doi.org/10.17892/app.2018.0006.122.

Pearlman, Karen, and John Sutton. “Reframing the Director: Distributed Creativity in Filmmaking Practice.” In *A Companion to Motion Pictures and Public Value*, edited by Mette Hjort and Ted Nannicelli, 86–105. Hoboken, US: Wiley-Blackwell, 2022. https://doi.org/10.1002/9781119677154.ch4.

Roth, Philip. “An Open Letter to Wikipedia.” *The New Yorker* (September 6, 2012). Accessed April 4, 2024. https://www.newyorker.com/books/page-turner/an-open-letter-to-wikipedia.

Russell, Wendy. “The Materiality of an Archival Object: A Reading of Penny Eyles’ Continuity Script for *Kes*.” Presented at the So Much More Than Non-Film: Cataloguing Film-Related Materials, Swedish Film Institute, Stockholm, June 2023. Accessed April 4, 2024. https://www.fiafnet.org/images/tinyUpload/2023/06/Kes_Continuity_Script_-_WR.pdf.

Smyth, J. E. *Nobody’s Girl Friday: The Women Who Ran Hollywood*. New York: Oxford University Press, 2018.

Thorp, Jer. *Living in Data: A Citizen’s Guide to a Better Information Future*. New York: MCD, Farrar, Straus and Giroux, 2021.

Wiesner, Heike. *Die Inszenierung der Geschlechter in den Naturwissenschaften: Wissenschafts- und Genderforschung im Dialog*. Frankfurt, New York: Campus-Verlag, 2002.

Filmography

Die kleinen Kleberinnen, dir. Eva Maria Hammel and Heide Breitel. West Germany, 1980.

Casper Tybjerg, Jonatan Bruun Borring, and Luan Nhu Vu

The Digitization of Silent Films and the Teaching of Film Historiography: Entanglements and Opportunities

Introduction

The present chapter presents research results concerning the cutting rates of Danish silent films, results obtained through the use of automated scene detection. Perhaps more importantly, the article also presents reflections on two methodological issues: first, the importance of integrating quantitative style analysis more closely with historical and archival research and, second, the potential of integrating historical and stylistic research with teaching. The article has grown out of a strategic initiative at the University of Copenhagen to advance the integration of research into teaching. More specifically, it emerges from a methods seminar in the first year of the MA program in Film and Media Studies.[1]

The most common form of qualitative style analysis has been the examination of cutting rates.[2] In our estimation, quantitative style analysis is most valuable if it can be carried out at scale. Establishing how fast a particular film is cut will not be particularly informative unless it can be compared to some sort of broader pattern. For instance, a more qualified discussion of the stylistic *norms* shaping Danish silent cinema would require a fairly large number of movies to be stylistically quantified. Comparisons between, say, different national traditions would require even more data.[3]

1 To underscore the collaborative character of this article, we largely refer to ourselves as "we" in the rest of the article. In practice, the bulk of the research work, particularly the development of the PySceneDetect parameters, was carried out by Borring and Vu (students in the methods seminar), whereas the framework discussion of research-teaching integration and entangled film history has largely been written by Tybjerg (teacher of the methods seminar).

2 The most important resource for this kind of work remains the Cinemetrics website (www.cinemetrics.uchicago.edu), founded by Yuri Tsivian. See Yuri Tsivian, "Cinemetrics: Part of the Humanities' Cyberinfrastructure," in *Digital Tools in Media Studies*, ed. Michael Ross, Manfred Grauer, and Bernd Freisleben (Bielefeld: Transcript, 2009). The research group headed by András Bálint Kovács has also made significant contributions; see András Bálint Kovács, "Shot Scale Distribution: An Authorial Fingerprint or a Cognitive Pattern?" *Projections* 8, no. 2 (2014): 50–70; Mattia Savardi et al., "CineScale: A Dataset of Cinematic Shot Scale in Movies," *Data in Brief* 36 (2021): 107002, https://doi.org/10.1016/j.dib.2021.107002.

3 Radomír D. Kokeš, "Norms, Forms and Roles: Notes on the Concept of Norm (Not Just) in Neoformalist Poetics of Cinema," *Panoptikum* 22 (2019): 52–78.

https://doi.org/10.1515/9783111082486-004

The proposal for the first iteration of the integrated course, taught in the fall of 2021, had foreseen that students would be able to gather the necessary data easily using digital tools. If each two-student team could quantify four or five films, a sizable dataset could be assembled, especially if the exercise was repeated across several annual iterations of the course. Numerous practical issues arose, however, involving both access to the films and the performance of the digital tools; together, they made the proposed approach seem altogether impracticable. Because of this, that kind of research was given less attention in the second iteration of the course. Even so, Borring and Wu were able to solve at least some of the technical issues involved, opening up the possibility of doing further work along these lines in subsequent iterations of the course.

Since this chapter has emerged from work on integrating research with teaching, we decided to start out by briefly reviewing the established models for such research integration, allowing us to reflect on the degree to which our collaboration fits these models. An important result of our work is a heightened awareness of the importance of the way digital archival materials are curated and presented, what kinds of metadata are made available, and so on. To approach these issues, we have found the *Entangled Film History* approach useful, and we discuss it in the second section. While entangled film history is generally focused on transnational topics, we have drawn on its self-reflexive component and attention to the context of the research. The actual research conducted was a quantitative style analysis of a limited number of Danish silent films. The third section will provide a brief introduction to this kind of research, while the fourth examines some of the practical pitfalls encountered while conducting the research, to which the freeware program PySceneDetect has provided a good practical solution. The results of the research are presented in the fifth section, and our conclusion sums up the methodological lessons we believe we can take from our work.

Integrating Research in Teaching

In this section, we will briefly look at the conceptualization of research integration in teaching underlying the plan for the course and discuss where our research collaboration can fit within that model.

In Denmark, “research-based teaching” is written into the University Law, the legislation governing the operations of the Danish universities, as a foundational principle.

> §2. It is the task of the university to conduct research and offer research-based education up to the highest international level within its disciplinary subjects.[4]

There is broad agreement among stakeholders that "research-based teaching" is foundational to the role of the universities as educational institutions; accordingly, there has been a good deal of discussion about what it actually entails. In practice, it has generally been assumed that the requirement was satisfied if a substantial proportion of courses were taught by active researchers. However, because of its definitional importance, the University of Copenhagen has made it a strategic goal to "Further develop models for student involvement in research activities and make it a credit-bearing element of their programme."[5] To advance this goal, a certain amount of funding was made available to faculty who wanted to experiment with a higher degree of research integration in the classes they taught.

The fact that the university's leadership has made research integration a strategic goal underscores that it is not a neutral concept. As was pointed out some years ago in a working paper surveying the literature on research integration, it is assumed to be a good thing from the outset – which also leads to the assumption that the more of it, the better: "In the main the literature is characterized by a normative perspective which argues that there is strong value in enhancing the teaching-research nexus in terms of improving student learning and in other areas."[6] These assumptions are also present to a certain degree in the models developed by those who research and write on the practice of university teaching. As part of the University of Copenhagen strategic initiative, a group of researchers conducted an ethnological investigation of how research was integrated into teaching in practice, and what students and researchers thought about it.[7] Based on this research, they developed a model of research integration, which was proposed as an alternative to models de-

4 Universitetsloven, LBK no. 778, August 7, 2019, accessed October 23, 2023, https://www.retsinformation.dk/eli/lta/2019/778 (translation by the authors).

5 "Talent and Cooperation," University of Copenhagen 2023 Strategy document, accessed April 11, 2024, https://web.archive.org/web/20231128025913/https://about.ku.dk/strategy2023/education/.

6 Paul Trowler and Terry Wareham, "Tribes, Territories, Research and Teaching: Enhancing the Teaching-Research Nexus" (January 1, 2008), Working Paper, 13, accessed October 23, 2023, https://www.researchgate.net/publication/252423791_Tribes_territories_research_and_teaching_Enhancing_the_teaching-research_nexus.

7 Tine Damsholt and Marie Sandberg, *Af lyst eller nød: En etnologisk undersøgelse af integration mellem forskning og undervisning i praksis* (Copenhagen: University of Copenhagen, 2018), accessed February 20, 2023, https://curis.ku.dk/ws/files/213594738/AF_LYST_ELLER_N_D_ELEKTRONISK_VERSION_2.pdf.

veloped elsewhere, particularly in Great Britain.[8] The description of the Danish model (Figure 1) explicitly stresses that it is intended to convey the message that different types of research integration are *equal*: "The types of research–based teaching and research integration are regarded equally as the selected types must fit the individual degree programme and the students' level."[9] The model replaced an earlier one that looked like a staircase, with the various types of research integration drawn as steps, signaling a hierarchical progression from lower to higher. However, the new model does not entirely avoid this trap either. The colors in the image of the model deepen as one moves upwards and to the right, which still at least implicitly suggests a hierarchy: the deeper the color, the better.

In its first iteration, the course was intended to include different levels of research integration. For the purposes of Tybjerg's plan to investigate the aesthetic norms of Danish silent cinema, the most important component would involve students working on data collection and processing (area b. on the model). The idea was for students working in pairs to work through a handful of films, producing complete shot lists (allowing cutting rates to be computed) and adding information about shot scale, camera movements, and scene boundaries. All the films would be Danish and from the same span of years. The hope was that the creation of a dataset of this sort would enable conclusions to be reached about the norms governing film production during the period in question.

While this would be the students' most important contribution from the point of view of Tybjerg's research interests within the overall plan of the course, it was only intended as a fairly small preparatory assignment – unless individual students specifically chose to do further work on the data. In practice, however, this proved to be unrealistically demanding. Part of the reason for planning this data-collection work as a side exercise was that the ambition for the course had been to enable students to conduct independent research – an ambition moti-

8 The key text is Mick Healey's 2005 article "Linking Research and Teaching"; its central importance is stressed in an extensive report surveying the literature on research integration in teaching: "In our review, we have not been able to identify an alternative typology that would represent a substantial divergence from this model" (Mari Elken and Sabine Wollscheid, *The Relationship between Research and Education: Typologies and Indicators. A Literature Review*, Nordic Institute for Studies in Innovation, Research and Education (NIFU) (Oslo, 2016), 16, accessed February 6, 2023, https://www.nifu.no/en/publications/1351162/; citing Mick Healey, "Linking Research and Teaching: Exploring Disciplinary Spaces and the Role of Inquiry-Based Learning," in *Reshaping the University: New Relationships Between Research, Scholarship and Teaching*, ed. Ronald Barnett (Maidenhead: McGraw-Hill Education, 2005), 67–78).

9 "Research integration in teaching," project home page, accessed February 8, 2023, https://kunet.ku.dk/work-areas/teaching/teaching_development/funds-for-experiments-with-research-integration/Pages/default.aspx.

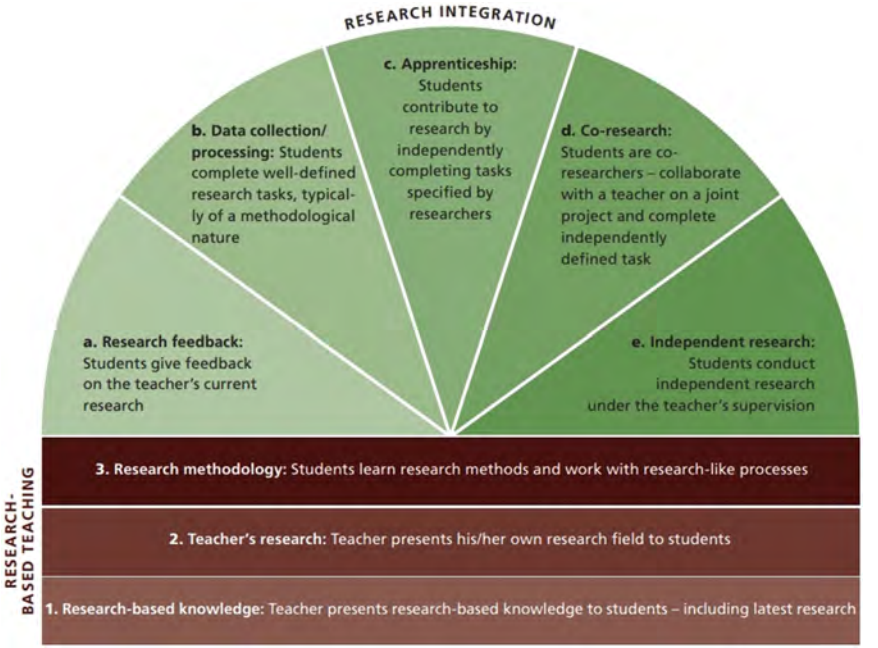


Figure 1: Research-teaching integration model. The darker the colors, the greater the degree of research independence and student autonomy. https://kunet.ku.dk/work-areas/teaching/teaching_development/funds-for-experiments-with-research-integration/Pages/default.aspx (accessed April 11, 2024).

vated to some extent by the way the models gave pride of place to independent research conducted by students. Another consideration was the hope that students could come up with innovative ways of using the material available on the website www.stumfilm.dk.

At stumfilm.dk, the Danish Film Institute (DFI) has made and continues to make available a rich trove of material on Danish silent cinema: all extant Danish silent fiction films are being digitized and presented, along with many related archival documents. The great majority of these films were produced by the dominant company Nordisk Film, which still exists and gave permission to make their films available, greatly reducing copyright concerns.

Students were given suggestions for possible ways of approaching the material but were also urged to come up with their own ideas. They were also given the opportunity to tour the DFI's archive facility, to give them an understanding of the physical character of the archival objects, the extensive infrastructure required to ensure their continued preservation, and the skilled labor involved in making digital versions available to users – an understanding, in other words, of

the efforts and support structures required to maintain the apparently seamless and effortless experience of using the stumfilm.dk website.

A complex archive or archive-like resource like stumfilm.dk may open up new avenues of investigation. It prompts us to ask: what can we do with it? What new questions will it enable us to answer? Thinking of the research process in this way may appear to turn the "proper" relation between research material and research questions on its head. Textbooks on empirical methodologies tend to insist on the primacy of research questions: only when researchers have formulated their research questions should they begin to gather data. While social science disciplines generally adhere to this model, scholarly work in the humanities, particularly historical scholarship, commonly begins with the archive. Researchers immerse themselves in documents, and only later formulate a research question or questions based on what they discover. Film historian Eric Schaefer disarmingly refers to this procedure as "critical mess historiography."[10] The historians' immersion-first approach is a well-established research mode, although not always recognized as such in the methodological literature.

Even if our study ended up conforming to the research-question-driven model, the presence of this sort of alternative reveals the need for a framework for thinking about how research may be shaped by institutional and disciplinary contexts; in particular, for thinking more closely about how the archive is set up and what kinds of research procedures it does or does not facilitate – the archive's affordances or *dispositif*, as it were. *Entangled film history* provides such a framework.

Entangled (Digital) Film History

The entangled approach emphasizes how film historians themselves are enmeshed in a particular context (scholarly traditions, archival access, technological resources, national institutions of learning).[11] In Film and Media Studies, the en-

10 Eric Schaefer, "The Problem with Sexploitation Movies," *Iluminace* 24, no. 3 (2012): 151.

11 See Malte Hagener, "Introduction: The Emergence of Film Culture," in *The Emergence of Film Culture: Knowledge Production, Institution Building, and the Fate of the Avant-Garde in Europe, 1919–1945*, ed. Malte Hagener (New York: Berghahn Books, 2014); Daniël Biltereyst and Philippe Meers, "Comparative, Entangled, Parallel and 'Other' Cinema Histories. Another Reflection on the Comparative Mode within New Cinema History," *TMG Journal for Media History* 23, nos. 1–2 (2020); Casper Tybjerg, "The European Principle: Art and Border Crossings in Carl Theodor Dreyer's Career," in *A History of Danish Cinema*, ed. Isak Thorsen, C. Claire Thomson, and Pei-Sze Chow (Edinburgh: Edinburgh University Press, 2021), 41–50; Casper Tybjerg, "Danish-German Cinematic Interconnections and the Prospects of an Entangled Film Historiography," in *Danish and*

tangled approach has mainly been discussed as a framework for investigating the border-crossing and transnational character of many filmmaking careers and media industries. However, in its original formulation as *histoire croisée*,[12] it also encourages (film) historians to reflect on the context within which they do their work, how their work is constrained and facilitated by it, and how it makes some research questions salient while obscuring others.

To take a simple example: to do a quantitative style analysis of a film, or segment and annotate it in some specialized software, you really need to have the film as a file on your computer. The stumfilm.dk website, however, does not allow you to download the films directly. While we had initially hoped to work on all the 235 films available on the site, this was not possible in practice. We would have needed staff from the DFI to download the titles individually onto a hard drive for us, and that was just not feasible given the time available to carry out the investigation. We ended up with a much smaller sample of films; further, given the technical issues involved (of which more below), it would not have been realistic to attempt to analyze the full collection of films, even if it had been accessible.

This example is important to keep in mind when thinking about the way historical research is facilitated and shaped by the archive's affordances. The digital availability of large numbers of silent films enables researchers to view them systematically, but if they cannot be downloaded, certain types of investigation are very difficult to carry out. Another factor to consider is that many archives are organized as national (and nation-bound) institutions. This is apparent in the very names of both the website and the institution that runs it: for both stumfilm.dk and the Danish Film Institute, the nation of Denmark is evidently an important organizational principle. While this may seem unproblematic if you want to work on Danish silent film, the institutional framework makes it easy to think in only national terms and to overlook cross-border entanglements.

In particular, when trying to map out stylistic norms, film historians have long argued that national cinemas are not the most fruitful framework. David Bordwell has shown that the "standard story" of the history of film style has been resolutely internationalist from very early on.[13] From the first, film has been a highly international medium. While the emergence and persistence of certain craft traditions may be best explained at the national level, this cannot be as-

German Silent Cinema, 1910–1930: Towards a Common Film Culture, ed. Lars-Martin Sørensen and Casper Tybjerg (Edinburgh: Edinburgh University Press, 2023), 24–50.

12 Michael Werner and Bénédicte Zimmermann, "Beyond Comparison: Histoire Croisée and the Challenge of Reflexivity," *History and Theory* 45, no. 1 (2006): 30–50.

13 David Bordwell, "The Power of a Research Tradition: Prospects for Progress in the Study of Film Style," *Film History* 6, no. 1 (1994): 59–79.

sumed from the outset. Film workers have often travelled to work in other countries, transferring and absorbing skills and norms. Hollywood films have been shown in most places, making their stylistic and storytelling devices at least theoretically available as creative options for filmmakers elsewhere.

Even so, nationally focused works of film historiography predominate. Like archives, research institutions tend to be organized on national lines, and both films and written sources use a particular language, creating a strong countervailing pressure in the direction of methodological nationalism. It is not realistic to imagine that we can easily escape from the path dependencies created by these institutional facts. Instead, we should strive for compatibility and comparability, allowing scholars from different countries to work with each other's data and results.

The need to ensure compatibility was also present when thinking about how quantitative style analysis could be brought into the classroom in a way that would produce data useful for further research. Students would need to follow a clear, predefined procedure for gathering and organizing the data. As mentioned in the introduction, the original idea was to have students work in pairs or small groups[14] to gather quantitative stylistic data on a handful of films, each group working with a different set for films. With enough students and a repetition of the exercise over several iterations of the course, a large and hopefully coherent dataset could be assembled.

However, the curriculum for the Film and Media Studies MA program of which the course was a part requires that the students work on more independent projects for their exams, allowing them to develop their own methodological reflections based on this work. Since this was a key learning outcome, the data gathering exercise had to be a preliminary one that would be practicable for the students to carry out; as a minimum, they should be able to easily access the films for analysis. The difficulties with access, outlined above, led to students being offered alternative options for their exercises. A few, however, were sufficiently intrigued by quantitative style analysis to focus on that approach.

14 András Kovács had three-person teams do the coding for his project: "Shot scale annotations were provided by three human coders (2 coders + 1 who made decision in case of disagreement)." See Savardi et al., "CineScale," 8.

Quantitative Style Analysis

The quantitative analysis of film style involves counting different types of stylistic features in a film, typically for the purpose of comparison with other films. The process is fairly demanding in terms of the resources it requires, which has been an obstacle to its widespread adoption. The earliest practitioners of the approach Barry Salt, Raymond Bellour, David Bordwell and Kristin Thompson[15]) would work on 35mm prints, but few researchers had the necessary access to extensive film archives and viewing tables. The advent of analog and digital video formats held out considerable promise, and the approach came within reach of anyone with an interest in the field (rather than just a privileged few) with the development of purpose-built and freely accessible digital tools, of which the most important has been Cinemetrics, launched in 2005.[16]

The film historian Barry Salt has been one of the biggest proponents of this quantitative method, from his pioneering articles in 1974 to his *magnum opus*, *Film Style and Technology* (1983, revised and expanded in 1992 and again in 2009), his findings based on years of collecting film style data from across the whole of film history.[17] Salt argues that a central parameter of cinematic style is *average shot length* (ASL), which is the length of a film in seconds divided by the number of shots in it.[18] The advantage of ASL is that it is easy to calculate: you just need to count each cut, which can be done with a clicker or by making marks on a piece of paper. Existing studies outside Denmark have mostly used this measurement. It has therefore also been used for the present study because it allows for easy comparison with these already existing studies, including Salt's.

Some researchers have pointed out that ASL can be misleading as a measurement, particularly if the film contains some very long or very brief shots. These outliers will distort the measurement, giving two films that are fairly similar in tempo quite different ASLs if one contains such outlier shots and the other does

15 Barry Salt, "Statistical Style Analysis of Motion Pictures," *Film Quarterly* 28, no. 1 (1974): 13–22; Raymond Bellour, "The Unattainable Text," *Screen* 16, no. 3 (1975): 19–27; David Bordwell, Janet Staiger, and Kristin Thompson, *The Classical Hollywood Cinema: Film Style and Mode of Production to 1960* (London: Routledge, 1985).

16 For more historical detail, see Casper Tybjerg, "En introduktion til kvantitativ æstetisk filmanalyse i praksis," *Kosmorama* (2021).

17 Barry Salt, *Film Style and Technology: History and Analysis* (London: Starword, 1983), 2nd ed. (London: Starword, 1992), and 3rd ed. (London: Starword, 2009). See also Salt's compendium of his articles (with useful retrospective commentary), *Moving into Pictures: More on Film History, Style, and Analysis* (London: Starword, 2006).

18 Salt, *Film Style and Technology*, 3rd ed., 160.

not.[19] To compensate for this distortion, these researchers recommend using *median shot length* (MSL) instead: recording the length of each shot, ordering them from the shortest to the longest, and then taking the one in the middle. MSL, according to these researchers, is a more reliable parameter when it comes to comparing the pace of individual films. However, it is somewhat more difficult to calculate since you need to know the length of each shot in the film.

Since the present study was initially conceived as one that would examine a very large corpus of films, the decision was made to focus on tempo. Tempo is an important stylistic characteristic of the film medium; it makes the medium's characteristic temporal dimension meaningful. Tempo has changed importantly over time and constitutes an essential part of the historically changing and developing stylistic conventions of cinema. For the purposes of this study, it was decided to treat cutting rates as indicative of tempo, using ASL and MSL as parameters. The quantification of stylistic parameters allows you to observe the development of different film techniques over time and to compare different groups of films (national cinemas, films by the same director, etc.) with each other. Here, Salt's data and research have laid the groundwork for subsequent studies.

In his contribution to the volume *Finnish Cinema: A Transnational Enterprise*, Jaako Seppälä proceeds from Salt's work in an attempt to explain the peculiarity of Finnish silent fiction films and how foreign films influenced their stylistic development.[20] Like the present study, Seppälä looks at the development of tempo; Finnish silent films tend to be slow-paced. Seppälä has analyzed a substantial number of films from the period from 1920 to 1931, not only calculating their average shot length and median shot length, but also analyzing shot scale. As Seppälä points out, however, there is some disagreement among scholars working in this field on how best to code for shot scale, particularly regarding more distant shots.[21]

---

19 Nick Redfern, "The Average Shot Length as a Statistic of Film Style," *Cinemetrics: Film Statistics: Give and Take* (n.d.), accessed April 11, 2024, https://web.archive.org/web/20230328210123/http://cinemetrics.lv/fsgt_q1b.php.

20 Jaakko Seppälä, "Finnish Film Style in the Silent Era," in *Finnish Cinema: A Transnational Enterprise*, ed. Henry Bacon (London: Palgrave Macmillan UK, 2016), 51.

21 Seppälä, "Finnish Film Style," 60. Although there is a slight difference in nomenclature, both Barry Salt and András Bálint Kovács use seven shot scales, from Big or Extreme Close Up to Extreme or Very Long Shot; see Barry Salt, "[Data Method] Statistical Style Analysis" (n.d.), accessed October 23, 2023, http://www.starword.com/Data_Method/data_method.html; Kovács, "Shot Scale Distribution," 50. However, Kovács observes that "Salt's definition of shot scales is slightly different from what I have used toward the long end of the scale" (54). Moreover, Kovács adds the category of "foreground shot," which combines two different shot scales, with one significant pictorial element close to the camera and another fairly distant from it (51). Taken together, these differences were significant

It would not have been impossible to make an informed decision on how to code shot scale distributions in our course, but it would have required a fair amount of exploratory work and discussion of the advantages and disadvantages of different coding schemes. This would have meant that fewer movies could be analyzed within the time available, which is why we decided to concentrate on collecting data on tempo.

Practical Challenges and Solutions

One key challenge for carrying out the proposed research involved finding the right tools for the job. In order to gather the necessary data at scale, it seemed clear that it should be easy to gather the data (preferably through an automated process) and they should be output in a format that would be straightforward to share and would allow further data to be added fairly easily. For instance, if one already had a dataset containing a list of shots for a number of films with the length of each shot, it would be ideal if information about shot scale, the number and identity of characters, entrances and exits, camera movement, etc., could then be added. Cinemetrics is a very elegant piece of software, but it is not automated, and it does not really allow you to extract data. Regrettably, because of the untimely death of its developer Gunars Civians, it was not maintained for a long time, and since the web interface was a Flash program, it became unworkable for most users, although a new version (a Google Chrome extension) has just been released.[22] The problem of software developed for research becoming obsolete because programs are not updated has also affected other specialized film studies tools, as they have often relied on funding for specific, time-limited research projects.

Tybjerg had hoped that the problem could be solved with software providing *automated scene* (or *shot-boundary*) *detection*, where the computer looks for changes in the pixel composition in the image that exceed a certain tolerance.[23] With the right software and settings, this method is surprisingly accurate, al-

enough that, when Kovács compared his analyses of the same films to Salt's data, "in all cases the [shot scale distribution] patterns of Salt's measurement and of mine were very different" (55).

22 Cinemetrics Measurement Tool, chrome web store, accessed April 11, 2024, https://chromewebstore.google.com/detail/cinemetrics-measurement-t/bekhkeilpophjchjdbhnfhecjkhjlgme?hl=en.

23 "Fully automatic shot-boundary detection (SBD) has become the holy grail of video indexing," wrote Jeremy Butler in his article "Statistical Analysis of Television Style: What Can Numbers Tell Us about TV Editing?" *Cinema Journal* 54, no. 1 (2014): 30–31.

though some false positives and negatives do occur.[24] DaVinci Resolve, a powerful professional video editing program, seemed a promising candidate: it had the requisite automated scene detection capability; further, after running scene detection on a video file, the results can be viewed as an *editing decision list* that the program can export, allowing a shot list to be generated (almost) automatically; and, finally, it is available in a free version that does not restrict any features needed for the data-gathering process.

However, the DaVinci Resolve program is quite demanding in terms of the computer resources it needs to function optimally; it requires a lot of processing power to run scene detection on an entire film. Even powerful laptops struggle. Moreover, the results proved to be frequently unsatisfactory, with a lot of false positives or negatives (depending on the sensitivity chosen for the scene detection settings). Silent films look quite different from modern films with respect to the features the scene detection algorithms use to determine whether a frame is so different from its predecessor that a cut must have occurred. Silent films mostly lack color (many are of course tinted, but tints tend to be uniform within scenes, meaning that only the black-and-white patterns change from shot to shot), and prints often contain flaws (scratches, white blotches, uneven exposures) that are registered as shot changes by the scene detection algorithms. It is certainly possible to adjust the parameters of the scene detection feature of DaVinci Resolve to produce satisfactory scene detection results, but the processing power requirements meant that the necessary experimentation would have been very time-consuming and, worse, that it would be unlikely that students would be able to get the program to run satisfactorily on their laptops.

These difficulties had discouraged Tybjerg from making quantitative style analysis a central part of the course, but Borring and Vu were able to find a practicable alternative: PySceneDetect, a free program built with the Python programming language, designed specifically to detect shot boundaries in video files.[25]

PySceneDetect requires users to have Python installed on their computers (along with a few supplementary pieces of software), but no skill in Python programming is needed to use it. The scene detect program runs off the command line. Once you have the program installed, you simply write `scenedetect` followed by the name of the video file you want to analyze at the command line prompt. You add further commands to control the parameters used by the scene detection

24 A similar method is described in Brett Adams, Chitra Dorai, and Svetha Venkatesh, "Toward Automatic Extraction of Expressive Elements from Motion Pictures: Tempo," *IEEE Transactions on Multimedia* 4, no. 4 (2002): 472–481.

25 Available at https://www.scenedetect.com/, accessed April 11, 2024.

algorithm and to modify the output format (the scenedetect.com web page offers detailed instructions for using the program).

There are three algorithms for automatic scene detection in PySceneDetect: adaptive, content aware, and threshold.[26] After extensive experimentation between the different detection algorithms and thresholds on various parameters, we found a formula that works well on silent films. The command of the formula reads:

```
detect-adaptive --luma-only --threshold 4.5 --min-content-val 10 --frame-window 3
```

The adaptive algorithm is used; for each individual frame, it compares an average luma value (*luma-only* as the films are mostly grayscale) of the three previous and subsequent frames (*frame-window* value) to determine if a cut has happened. If the difference is above 4.5 (*threshold* value) and the difference between the frame in question and the previous frame is higher than 10 (*min-content-val* value) it is considered a cut. These values deviate significantly from the default values in the program. This is likely because of the way the visuals of silent films differ considerably from the modern movies from which the default values were created. Using the adaptive method (i.e., including previous and subsequent frames in the comparison) is especially useful when working with archival film material, as it avoids false positives if there are a few damaged frames. Precisely because the program includes the six surrounding frames in the assessment, it is not considered a cut if there is a sudden flash of light over a frame or two.

This formula was tested by holding the program's results up against a manual count. Three different films were tested. First, shots were counted by hand using a simple smartphone counting app. Next, the same film was run through the PySceneDetect program. When the shot counts were compared, the results were as shown in Table 1.

With deviations of 3 percent, 1 percent and 0.3 percent, the PySceneDetect algorithm has proven to be reliable.

PySceneDetect makes it easy to export the results to a spreadsheet or database. If the command list-scenes is added, the program outputs a shot list as a CSV file that can be imported into Microsoft Excel or other programs. The command save-images exports frame grabs of the first, middle, and last frames of each shot, allowing easy visual verification of the algorithm's segmentation. The frame grabs can also be used to code for other stylistic parameters.

26 See https://www.scenedetect.com/docs/latest/api/detectors.html, accessed April 11, 2024.

Table 1: Comparison of PySceneDetect with manual counting.

Den hvide slavehandel (1910)			
Manual count: 65 shots	PySceneDetect: 67 shots	Variance: 2	Deviation: +3%
Der var engang (1922)			
Manual count: 674 shots	PySceneDetect: 681 shots	Variance: 7	Deviation: +1%
Nedbrudte Nerver (1923)			
Manual count: 670 shots	PySceneDetect: 668 shots	Variance: 2	Deviation: −0.3%

Results

Armed with this effective tool, we proceeded to analyze a sample of 22 Danish silent feature films. We initially decided to focus on feature-length films (excluding short films and incomplete films from consideration) available on stumfilm.dk. The earliest available feature film was from 1910 and the latest from 1932. This fits with what we know about the history of Danish film production: the first feature films were made in 1910, and the 1932 film was the last silent film to be released (the other six features released that year were talkies).

We decided that, for the purposes of the study, it would make most sense to analyze a single random film for each year, ensuring that the whole period was covered. Since the number of feature films produced fluctuated a great deal over the period in question – from dozens every year in the mid-1910s to a handful or less at the end of the 1920s – and survival rates also differ a great deal, it is not possible to say whether the films chosen can be said to be a representative indicator of the tempo of Danish films in a given year.

The sampling of one film for each year in the period was possible until the years 1930 and 1931, where there were no surviving films available (two silent films were released in 1930 and just one in 1931, along with a total of three talkies). Each film from each year was chosen completely randomly from among those available. For all the years up to 1929, there was a minimum of three films to choose from. For the year 1932, there was only a single film. The sample selected in this way included 22 silent feature films from 1910–1932 (excluding 1930 and 1931), which was also a manageable quantity of films and data to work with. We were given access to file copies of the films in the sample by the Danish Film Institute.

An overview of the results of the analysis can be seen in Table 2.

Table 2: Cutting rates for the films included in the sample.

TITLE	YEAR	LENGTH	ASL (sec.)	MSL (sec.)
Den hvide Slavehandel	1910	33 min.	29.6	19.9
Ved Fængslets Port	1911	41 min.	37.6	26.1
Dødens Brud	1912	41 min.	32.4	18.9
Ballettens Datter	1913	56 min.	25.8	17.2
Det hemmelighedsfulde X	1914	90 min.	16.3	12.3
Ned med Vaabnene	1915	65 min.	18.3	12.4
Verdens Undergang	1916	73 min.	15.5	11.5
Klovnen	1917	62 min.	18.1	13.2
Himmelskibet	1918	81 min.	7.3	6
Præsidenten	1919	71 min.	7.9	5.5
En Skuespillers Kærlighed	1920	47 min.	14.5	8.6
Vor fælles Ven	1921	139 min.	8.8	6.7
Store Forventninger	1922	113 min.	8.6	6.4
Nedbrudte Nerver	1923	75 min.	6.7	5
Hadda Padda	1924	83 min.	8	5.1
Fra Piazza del Popolo	1925	107 min.	8.4	6.6
Klovnen	1926	139 min.	9.7	7.3
Vester-Vov–Vov	1927	101 min.	6.2	4.3
Jokeren	1928	102 min.	5.7	4
Højt paa en Kvist	1929	97 min.	6.2	4.5
I Kantonnement	1932	99 min.	6.6	4.4

The table includes information about length, since silent films did not have a set projection speed.[27] Individual projectionists could (and did) show a given film at very different speeds. Video versions or digitizations of the same film may also have been transferred at different speeds. Therefore, discrepancies may arise between different analyses of the same film. Barry Salt's dataset includes ASLs for 11 Danish silent films, of which two are also part of the sample examined here (see Table 3).

The ASL of Klovnen (1917) is given as 18.0 by Salt, while we have calculated it to be 18.1 – a negligible difference. However, Salt gives the ASL of Det hemmelighedsfulde X as 12.0, much faster than the ASL of 16.3 recorded here. The most likely explanation is that Salt worked from a print run or transferred at a much faster speed, probably 24 frames per second, whereas the DFI's transfer was made at 16

27 Kevin Brownlow, "Silent Films – What Was the Right Speed?" in *Early Cinema: Space – Frame – Narrative*, ed. Thomas Elsaesser and Adam Barker (London: BFI, 1990), 282–290; Casper Tybjerg, "About the Frame Rates," Visual essay, in Carl Th. Dreyer, La Passion de Jeanne d'Arc, Blu-ray ed. (New York: Criterion Collection, 2017).

Table 3: From Barry Salt's dataset: cutting rates for Danish silent films.[28]

TITLE	YEAR	ASL
Ekspeditricen	1911	43.0
De fire Djævle	1911	21.0
Dødspring til Hest fra Cirkuskuplen	1912	17.0
Ekspressens Mysterium	1914	21.0
Den fremmende	1914	16.0
Det hemmelighedsfulde X	1914	12.0
Den mystiske Fremmende	1914	17.0
Verdens Undergang	1916	13.0
Klovnen	1917	18.0
Blade af Satans Bog	1919	7.0
Der var Engang	1922	5.0

frames per second.[29] The transfer rates are unfortunately rarely given in open metadata, which is really regrettable. We would strongly recommend that archives digitizing films include transfer speeds in frames per second as part of the immediately available metadata for each title, preferably along with the length in meters of the source print. Until that happens, researchers doing this kind of research should provide the length in minutes of the video file used (as well as full information about its source), making it possible to see and compensate for differences in the speeds at which the film may have been viewed, transferred, or digitized.

As can be seen, the pace of Danish silent films increased significantly from 1910 to 1932. Both ASL and MSL decrease consistently over time. The slowest film is Ved Fængslets Port (Temptations of the Big City, 1911) with an ASL of 37.6 seconds and an MSL of 26.1 seconds. The film with the highest tempo in our sample is Jokeren (The Joker, 1928) with an ASL of 5.7 seconds and an MSL of 4 seconds. A visualization of this data can be seen in Figure 2. As the graph makes clear, the pace increases sharply over the course of the 1910s, then levels off in the 1920s. This is clearly seen in the exponential trend line for both ASL and MSL.

28 Source: Barry Salt's Average Shot Length data table, accessed October 23, 2023, http://starword.com/Data_Method/Data_Tables/data_tables.html. For Der var engang, our manual test count produced an ASL of 6.7 rather than Salt's 5.0, but this is likely due to Salt working from an older print without the many rather long intertitles of the 2002 restoration.

29 Claus Greffel, digital restoration and mastering technician, DFI, personal communication, June 15, 2023.

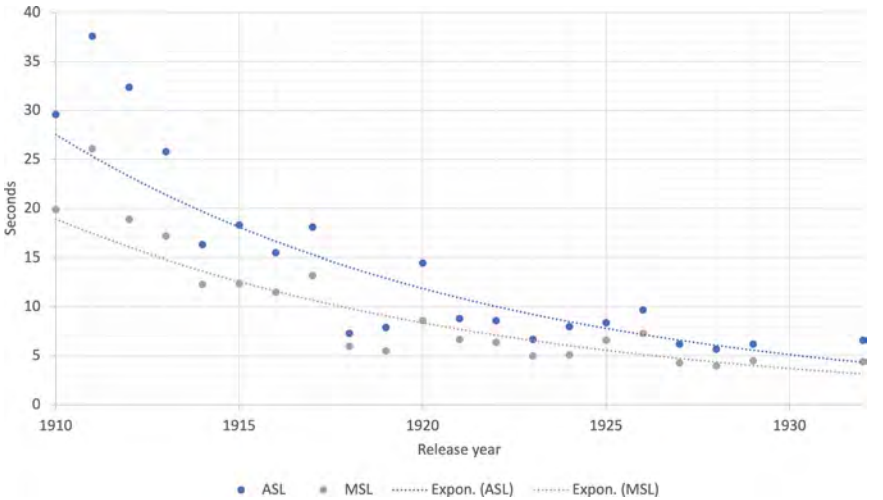


Figure 2: The pace of Danish silent films.

The overall development is clear and relatively homogeneous. There are not many deviations in the overall trend. A good example of this is A. W. Sandberg's KLOVNEN from 1917, which was remade in 1926 by the same director for the same production company. The 1917 version has an ASL of 18.1 while the 1926 version's ASL is 9.7. This strongly suggests that the drop in ASL between the two versions is due to the overall development towards a faster tempo in Danish film during the period. We also looked at whether any significant difference from the overall development could be observed if films made by the same directors or production companies were examined separately. No significant correlation was found, though the sample is really too small to draw conclusions. Still, the pattern seems to be an overall development, a change in stylistic norms, rather than something specific to particular filmmakers or organizations. One possible explanation could be the international adoption of scene dissection and the abandonment of the tableau style in the European cinema of the early 1920s.

We also sought to compare the data to international patterns. The lack of information about projection speeds makes it problematic to use Salt's data for silent films, but the patterns seem clear. Salt divides the silent feature period into three: 1911–1917, 1918–1923, and 1924–1929. A similar division of the sample yields the following results: from 1911 to 1917, the Danish films examined have an average ASL of 15.9 seconds (if we include the 1910 film, the average becomes 16.4). From 1918 to 1923, the ASL drops sharply to an average of 6.4. In the last period, from 1924 to 1932, the average ASL is 5.3. Table 4 compares the averages from the

sample with the average ASLs Salt calculates from his data on American and European films.[30]

Table 4: Average ASLs in the United States, Europe, and Denmark.

	1911–1917	1918–1923	1924–1929
U.S.	9.6	6.5	5.8
Europe	15.0	8.5	6.5
Denmark	15.9	6.4	5.3

In the first period (1911–1917), the Danish films resemble the rest of Europe when it comes to tempo, with an ASL of 15.9 compared to the European average of 15.0. American films are significantly faster with an average ASL of 9.6 seconds. During the next period, the Danish films leave the European one behind and reach the American pace; while the rest of Europe has an average ASL of 8.5, Denmark and the U.S. are at 6.4 and 6.5. In the last period (1924–1929), the European average has fallen to 6.5, but the American average has fallen to 5.8 and the Danish even further to 5.3. On the face of it, it seems unlikely that Danish films of the late 1920s were actually faster paced than Hollywood pictures, and this is almost certainly an artifact of the small number of films in our sample. Still, the result is intriguing and suggests the need for further research.

Conclusion

The present study indicates that the cutting rates and thus the tempo of Danish films increased over the course of the silent period, rapidly at first, then more slowly. It suggests that Danish films caught up with the cutting rates of American ones in the late 1920s, a surprising result that calls for further research.

Perhaps more importantly, the present study can be seen as a proof of concept with respect to the automation of quantitative film style analysis. The reliability of the PySceneDetect algorithm when using our command parameters seems strong enough for generalizable conclusions. An obvious next step is to conduct the same

30 Data for U.S. and Europe from Salt, *Film Style and Technology*, 3rd ed., 161, 192. The Danish titles individually listed by Salt (see n. 27) produce averages of 19.7 for 1911–1917 (9 films) and 6.0 for 1918–1923 (2 films). Salt lists no Danish films from 1924–1929.

study with a much larger sample. When DFI has finished digitizing all preserved Danish silent films, the study could be repeated with all available feature films. A large part of the work time lies in writing commands to the program and then processing the data for each individual movie, but this could easily be automated. It would not be overly difficult for someone with programming skills to write a program that automatically fed PySceneDetect with commands for each movie and then processed the data and compiled it into a spreadsheet. You could let the program run for a few days and then have all the data ready.

Of course, this would likely be a national database at first, and the entangled film history approach reminds us of the way methodological nationalism can limit our perspectives. Ideally, the database could gradually expand to include international data, perhaps allowing international patterns to emerge. Ahead of doing this kind of study, however, it would be important to secure the availability of metadata regarding the provenance of prints, the speeds at which the digital files have been transferred, and so on. To conduct a proper analysis of archival objects like the films presented on the stumfilm.dk site, researchers need to have a solid understanding of what these objects are and how they came to be. If our study shows anything, it shows the importance of understanding archival objects properly.

When planning courses that integrate digital methods, we think that it is important to recognize that it is very difficult to foresee all the technical and practical challenges that may arise, and that one should be ready to improvise and modify one's plans. Nevertheless, one outcome of this study has been to show that it would be feasible to do a quantitative style analysis exercise as part of the next iteration of the methods course. Video files for all the films from a particular year could be assembled and analyzed by two- or three-person teams.

Turning back to the research-integration model (Figure 1), we find that its categories do not wholly fit the work described in this article; they all make the process appear more structured and pre-planned than it tends to be in practice. Borring and Vu's work on cutting rates can fairly be described as category e.: "conduct[ing] independent research under the teacher's supervision." But arguably they made a more significant contribution by developing a set of commands for the PySceneDetect program that actually allows such research to be carried out at scale. In a sense, this is closer to c., as they "contribute[d] to research by independently completing tasks specified by researchers," even if "specified" in this case amounted to no more than Tybjerg saying, "it's cool that you want to do quantitative style analysis, but you'll have to figure out how to make it work in practice yourselves." While the model may suggest that e. is more "independent" (and therefore "better") than c., our experience indicates that the opposite was the case here.

Even as a retrospective categorization tool, the research-integration model (Figure 1) seems somewhat restrictive. It is useful when thinking about how to

integrate research with teaching, as long as the model and its categorization scheme is not applied too rigidly. From the point of view of students, the work certainly seems more rewarding if they are afforded at least some flexibility and have the opportunity to figure things out for themselves. From the instructor's point of view, having students carry out routine but time-consuming tasks like coding large datasets is certainly convenient and predictably useful. But even if giving students more loosely defined tasks and more freedom to grapple with how to resolve them may often fail to produce results that contribute to the instructor's own research projects, good students may come up with unforeseen solutions that advance projects in ways that the instructors could not have achieved on their own.

References

Adams, Brett, Chitra Dorai, and Svetha Venkatesh. "Toward Automatic Extraction of Expressive Elements from Motion Pictures: Tempo." *IEEE Transactions on Multimedia* 4, no. 4 (2002): 472–481.

Bellour, Raymond. "The Unattainable Text." *Screen* 16, no. 3 (1975): 19–27.

Biltereyst, Daniël, and Philippe Meers. "Comparative, Entangled, Parallel and 'Other' Cinema Histories. Another Reflection on the Comparative Mode within New Cinema History." *TMG Journal for Media History* 23, nos. 1–2 (2020): 1–9.

Bordwell, David. "The Power of a Research Tradition: Prospects for Progress in the Study of Film Style." *Film History* 6, no. 1 (1994): 59–79.

Bordwell, David, Janet Staiger, and Kristin Thompson. *The Classical Hollywood Cinema: Film Style and Mode of Production to 1960*. London: Routledge, 1985.

Brownlow, Kevin. "Silent Films – What Was the Right Speed?" In *Early Cinema: Space – Frame – Narrative*, edited by Thomas Elsaesser and Adam Barker, 282–290. London: BFI, 1990.

Butler, Jeremy. "Statistical Analysis of Television Style: What Can Numbers Tell Us about TV Editing?" *Cinema Journal* 54, no. 1 (2014): 25–44.

Damsholt, Tine, and Marie Sandberg. *Af lyst eller nød: En etnologisk undersøgelse af integration mellem forskning og undervisning i praksis*. Copenhagen: University of Copenhagen, 2018. Accessed February 20, 2023. https://curis.ku.dk/ws/files/213594738/AF_LYST_ELLER_N_D_ELEKTRONISK_VERSION_2.pdf.

Elken, Mari, and Sabine Wollscheid. *The Relationship between Research and Education: Typologies and Indicators. A Literature Review*. Oslo: Nordic Institute for Studies in Innovation, Research and Education (NIFU), 2016. Accessed February 6, 2023. https://www.nifu.no/en/publications/1351162/.

Hagener, Malte. "Introduction: The Emergence of Film Culture." In *The Emergence of Film Culture: Knowledge Production, Institution Building, and the Fate of the Avant-Garde in Europe, 1919–1945*, edited by Malte Hagener, 1–18. New York: Berghahn Books, 2014.

Healey, Mick. "Linking Research and Teaching: Exploring Disciplinary Spaces and the Role of Inquiry-Based Learning." In *Reshaping the University: New Relationships Between Research, Scholarship and Teaching*, edited by Ronald Barnett, 67–78. Maidenhead: McGraw-Hill Education, 2005.

Kokeš, Radomír D. "Norms, Forms and Roles: Notes on the Concept of Norm (Not Just) in Neoformalist Poetics of Cinema." *Panoptikum*, no. 22 (2019): 52–78.

Kovács, András Bálint. "Shot Scale Distribution: An Authorial Fingerprint or a Cognitive Pattern?" *Projections* 8, no. 2 (2014): 50–70.

Redfern, Nick. "The Average Shot Length as a Statistic of Film Style." *Cinemetrics: Film Statistics: Give and Take*. n.d. Accessed April 11, 2024. https://web.archive.org/web/20230328210123/http://cinemetrics.lv/fsgt_q1b.php.

Salt, Barry. *Film Style and Technology: History and Analysis*. 3rd ed. London: Starword, 2009.

Salt, Barry. *Moving into Pictures: More on Film History, Style, and Analysis*. London: Starword, 2006.

Salt, Barry. *Film Style and Technology: History and Analysis*. 2nd ed. London: Starword, 1992.

Salt, Barry. *Film Style and Technology: History and Analysis*. London: Starword, 1983.

Salt, Barry. "Statistical Style Analysis of Motion Pictures." *Film Quarterly* 28, no. 1 (1974): 13–22.

Salt, Barry. "[Data Method] Statistical Style Analysis." n.d. Accessed October 23, 2023. http://www.starword.com/Data_Method/data_method.html.

Savardi, Mattia, András Bálint Kovács, Alberto Signoroni, and Sergio Benini. "CineScale: A Dataset of Cinematic Shot Scale in Movies." *Data in Brief* 36 (2021): 107002. https://doi.org/10.1016/j.dib.2021.107002.

Schaefer, Eric. "The Problem with Sexploitation Movies." *Iluminace* 24, no. 3 (2012): 148–152.

Seppälä, Jaakko. "Finnish Film Style in the Silent Era." In *Finnish Cinema: A Transnational Enterprise*, edited by Henry Bacon, 51–80. London: Palgrave Macmillan UK, 2016.

Trowler, Paul, and Terry Wareham. "Tribes, Territories, Research and Teaching: Enhancing the Teaching-Research Nexus." January 1, 2008. Working Paper. Accessed October 23, 2023. https://www.researchgate.net/publication/252423791_Tribes_territories_research_and_teaching_Enhancing_the_teaching-research_nexus.

Tsivian, Yuri. "Cinemetrics: Part of the Humanities' Cyberinfrastructure." In *Digital Tools in Media Studies*, edited by Michael Ross, Manfred Grauer, and Bernd Freisleben, 93–100. Bielefeld: Transcript, 2009.

Tybjerg, Casper. "Danish-German Cinematic Interconnections and the Prospects of an Entangled Film Historiography." In *Danish and German Silent Cinema, 1910–1930: Towards a Common Film Culture*, edited by Lars-Martin Sørensen and Casper Tybjerg, 24–50. Edinburgh: Edinburgh University Press, 2023.

Tybjerg, Casper. "The European Principle: Art and Border Crossings in Carl Theodor Dreyer's Career." In *A History of Danish Cinema*, edited by Isak Thorsen, C. Claire Thomson, and Pei-Sze Chow, 41–50. Edinburgh: Edinburgh University Press, 2021.

Tybjerg, Casper. "En introduktion til kvantitativ æstetisk filmanalyse i praksis." *Kosmorama* (2021).

Tybjerg, Casper. "About the Frame Rates." Visual essay. In Carl Th. Dreyer, La Passion de Jeanne d'Arc, Blu-ray ed. New York: Criterion Collection, 2017.

Werner, Michael, and Bénédicte Zimmermann. "Beyond Comparison: Histoire Croisée and the Challenge of Reflexivity." *History and Theory* 45, no. 1 (2006): 30–50.

Imme Klages and Fabian Kling

Collecting Data and Connecting Traces: Researching and Modeling Sources on Doña Francisquita (S 1934)

Introduction

> Bringing the details together is the most important thing. In my recent work, I see very clearly, as an insight, that the most minor experiences, the smallest events, sometimes in later times, are the most important. (Raul Hilberg 2001)[1]

In October 2019, the Filmoteca Española in Madrid presented a digital edition of DOÑA FRANCISQUITA after its costly restoration and digitization. In her new digitized form, the film not only invites us to reflect on its historical context, but also on the contribution of digital sources to a more comprehensive understanding of its history. The film was shot in 1934 in Spain before the Spanish Civil War. The crew of DOÑA FRANCISQUITA mostly consisted of German exiles who, dismissed from the UFA in 1933 because of their Jewish origins, were banned from working, and had fled Germany via France to Spain, as they did not need a work permit to live there. Regarding the bigger picture, the film – like many films of German film exile in this period – represents an interesting case study of transnational cinema.[2] The analog sources (documents, letters, production files, etc.) on film ex-

1 Prof. Dr. Raul Hilberg, "Zur Integration des Wissens um die Judenvernichtung in die Gedenk- und Erinnerungsarbeit" (lecture in German at the 14th Annual Conference of Alemannia Judaica in Breisach, March 18, 2001), accessed January 13, 2023, https://www.youtube.com/watch?v=kmpRMaN2RH0 (Minute: 15:37; viewed January 13, 2023). Translation by the authors. Original quote: "Das Zusammenbringen der Einzelheiten ist das wichtigste. In meiner letzten Arbeit erkenne ich sehr deutlich, als Einsicht, dass die kleinsten Erfahrungen, die kleinsten Geschehnisse, manchmal in späteren Zeiten, die wichtigsten sind."

2 There has been profound film historical work done on the film DOÑA FRANCISQUITA in Spanish: J. B. Heinink and Alfonso C. Vallejo, *Catálogo del cine español. Volumen F3, Films de ficción 1931–1940* (Madrid: Cátedra/Filmoteca Española, 2009); Marta Munoz Aunion and Fernando Gonzalez Garcia, "El paraiso de los amigos de cine," in *Faros y torres vigía: el cine español durante la II República (1931–1939)*, ed. Julio Pérez Perucha and Agustín Rubio Alcover (Madrid: Asociación Española de Historiadores del Cine, 2016), 149–164; Valeria Camporesi, "The Tuneful 1930s. Spanish Musicals in a Global Context," in *Global Genres, Local Films. The Transnational Dimension of Spanish Cinema*, ed. Elena Oliete-Aldea, Beatriz Oria, and Juan A. Tarancón (New York: Bloomsbury 2017), 19–30; Valeria Camporesi and Fernando Gonzalez Garcia, "Un progreso en el arte na-

https://doi.org/10.1515/9783111082486-005

iles are manifold but scattered worldwide and always fragmented.[3] Using digital and digitized sources that are available online – filmographic information or historical sources on the film's crew, production, and reception – allows one to engage with the film's history without overlooking, but rather emphasizing, its transnational dimension. Conflating digital sources from archives and other cultural heritage institutions across Europe and the world – in this case study, Frankfurt, Bonn, Berlin, Bremen, Vienna, Barcelona, Madrid, and Madison (Wisconsin) – not only allows us to rediscover the lives of exiles but also embodies an argument for a transnational and data-driven approach to researching the topic.

Through zooming in on this research process, this chapter proposes three complementing methodological approaches to researching heterogeneous information from varying sources: a practical approach to source, data, and tool criticism, and the possibility of using data collections to assemble and structure information, as well as an examination of the idea of data as traces. As we reflect on our research on DOÑA FRANCISQUITA as a case study of digital film history, we explore the historical implications of collecting and modeling data.

Encountering *Data* as *Traces*

In order to explore the idea of data as traces, three different perspectives on the term *trace* are taken into account, each implying a methodological mindset that might be productive when dealing with heterogenous and potentially fragmented historical information or data. Data-based film historiography runs the risk of imagining a highly incomprehensible subject to be lucid or quantitatively manageable. As Marcus Burkhardt contends, databases create the fiction of absolute infor-

cional? Ibérica Films en España, 1933–1936," *Boletín del Seminario de Estudios de Arte y Arqueología* 77 (2011): 265–285.

3 A list of archives containing useful sources of film exiles are (but not limited to): Paul Kohner Agency, Deutsche Kinemathek Berlin; Marta Mierendorff papers at the Feuchtwanger Memorial Library, USC, Los Angeles; The USC Max Kade Institute, Los Angeles; Fritz Lang Papers at the American Film Institute, Los Angeles; Files of Warner Bros. Archive; Fritz Lang Collection at Cinematic Arts Library, USC, Los Angeles; The estate in the Billy Rose Theatre Division, the New York Public Library for the Performing Arts; Deutsches Literaturarchiv Marbach; Archiv der Akademie der Künste Berlin; Deutsches Exilarchiv 1933–1945, Deutsche Nationalbibliothek and much more. See Helmut Asper, *"Etwas Besseres als den Tod . . .": Filmexil in Hollywood: Porträts, Filme, Dokumente* (Marburg: Schüren, 2002).

mation, raising hopes for a fully accessible and controllable world.[4] Tyler Reigeluth describes a similar phenomenon when he notes a "tendency towards a naturalization of data, endowed with inherently 'objective' qualities and capable of 'speaking the truth'."[5] In his text on Big Data, Reigeluth argues for using the term *digital traces* rather than data to describe the information users produce or leave behind online, in order to develop a critical perspective on digital interactions:

> In the English language, very little academic research or media discourse refers to digital "traces." Instead the preferred lexical field seemingly revolves around "data." One term or lexical paradigm is not necessarily inherently better than another, especially considering that every language has its relevant and specific meanings for homologous words, but a challenging and critical perspective of our digital interactions can be developed by analyzing information as the traces that are produced, abandoned or captured in digital environments.[6]

Although Reigeluth focuses on a different field and emphasizes the physical traces data leaves on digital storage, there is an argument for avoiding the terms *source* or *data* in the field of historiography and using the less concrete term *trace*. It captures the potentially fragmented, partial, and unstable nature that is inherent in historical data.

In "Morelli, Freud and Sherlock Holmes: Clues and Scientific Method," Carlo Ginzburg traces the idea of the *conjectural paradigm* throughout the work of art historian and physician Giovanni Morelli, more precisely his method of identifying unknown painters on the basis of supposed minor details such as ears or hands;[7] through literary figure Sherlock Holmes;[8] and through Sigmund Freud's "proposal of an interpretative method based on taking marginal and irrelevant details as revealing clues."[9] Even though Anna Davin uses the term *clues* in her English translation of Ginzburg, he uses "tracce"[10] in the Italian original, which might well be translated as *traces*. For Ginzburg, neglecting the concept of systematic knowledge in "a social structure of ever-increasing complexity"[11] – which might be an apt de-

4 Marcus Burkhardt, *Digitale Datenbanken: Eine Medientheorie im Zeitalter von Big Data* (Bielefeld: Transcript, 2015), 333, https://doi.org/10.25969/mediarep/835.

5 Tyler Reigeluth, "Why Data Is Not Enough: Digital Traces as Control of Self and Self-Control," *Surveillance & Society* 12, no. 2 (2014): 243.

6 Reigeluth, "Why data is not enough," 248–249.

7 Carlo Ginzburg, "Morelli, Freud and Sherlock Holmes: Clues and Scientific Method," trans. Anna Davin, *History Workshop* 9 (1980): 9.

8 Ginzburg, "Morelli, Freud and Sherlock Holmes," 8.

9 Ginzburg, "Morelli, Freud and Sherlock Holmes," 11.

10 Carlo Ginzburg, "Spie. Radici di un paradigma indiziario," in *Miti emblemi spie. Morfologia e storia*, ed. Carlo Ginzburg (Torino: Einaudi, 1986), 165.

11 Ginzburg, "Morelli, Freud and Sherlock Holmes," 27.

scription of both the internet and German film migration – doesn't necessarily mean neglecting the concept of totality:

> On the contrary; the existence of a deep connection which explains superficial phenomena can be confirmed when it is acknowledged that direct knowledge of such a connection is impossible. Reality is opaque; but there are certain points – clues, signs – which allow us to decipher it. This idea, which is at the heart of the conjectural or semiotic paradigm, has made itself a place in a wide range of intellectual contexts, most deeply affecting the human sciences.[12]

While not referring specifically to historical sources, Ginzburg, like Hilberg, emphasizes the need to trace, connect, and contextualize information or sources, which seems inherent in the notion of *trace* as opposed to *source* or *data*.

Drawing on Marc Bloch, Paul Ricoeur describes history as a "science of traces."[13] While his book *Memory, History, Forgetting* presents a wide-ranging and complex analysis of immaterial and material memories, it seems productive to engage with his observations on the "persistence of traces,"[14] which he discusses in his last chapter, *Forgetting*:

> As has been said, the notion of trace can be reduced neither to the documentary trace nor to the cortical trace. Both consist of "external" marks but in different senses: that of the social institution for the archive, that of biological organization for the brain. There remains the third sort of inscription, the most problematic but the most significant for what follows in our investigation; it consists in the passive persistence of first impressions: an event has struck us, touched us, affected us, and the affective mark remains in our mind.[15]

Although Ricoeur attributes this persistence to a specific kind of trace and evaluates it rather negatively, there is an argument to be made that there is a certain persistence to traces (that are available to researchers) and that this persistence is a productive quality. Particularly in the field of German film exile, it seems appropriate to emphasize the persistence of a trace and thereby point to the historical reason for a seeming lack of information. Especially as the storage of data has a rather negative connotation, the idea of a trace as a resistant element in the production of knowledge captures the value of the information accumulated on topics like the historical German Film Migration. Envisioning data as traces forces researchers to engage digital or digitized information with a source-critical approach, which should be part of the research, as well as the modeling of historical data.

---

12 Ginzburg, "Morelli, Freud and Sherlock Holmes," 27–-28.

13 Paul Ricoeur, *Memory, History, Forgetting*, trans. Kathleen Blamey and David Pellauer (Chicago: University of Chicago Press, 2004), 13.

14 Ricoeur, *Memory, History, Forgetting*, 427.

15 Ricoeur, *Memory, History, Forgetting*, 427.

Performing Source, Data, and Tool Criticism

While digital research requires a critical approach to the historical sources found, a critical method should also be applied to the way in which they were found, i.e., the online tools, archives, and search engines used. Even setting aside the fact that research tools – in a broader sense – are black boxes anyway, it is a demanding, difficult task for humanistic researchers without advanced technical knowledge to evaluate them. Rather than focusing on a detailed evaluation of the tools used,[16] this article outlines ways to critically engage with tools during research. As Karin van Es, Maranke Wieringa, and Mirko Tobias Schäfer put it in their article "Tool Criticism," tool criticism shouldn't just reflect on the tool itself, but include "its influence on the research process and the results and their presentation, and the way in which the user interact[s] with it".[17]

> In our practice we see tool criticism as a reflexive and critical engagement with tools. In this reflexive and critical practice, the limitations and presuppositions built into the tool and its output need to be put under scrutiny, as well as the user's interaction with the tool.[18]

As we will show later in this article, the different spellings of names and versions of film titles stand out as examples. It is unclear to what extent search engines are supplied with data that assigns different names to the same person or film. It is the responsibility of humanities scholars to discover missing connections and question those provided by the algorithms – or rather the people who create metadata.

In order to critically engage with a platform or tool, it might be productive to question how these sites provide information to their users and where they get their information from. Encyclopedic platforms, for example, should be approached differently from digital archives, because of the different ways in which they accumulate data or information – working with search engines presents different challenges from filmographic databases because of their varying specificity to film.

Given that many platforms are labeled or self-labeled according to different standards, contextualization is essential. The example of *genealogy databases* might illustrate how classifying the origin of information can be productive. Although most of them are labeled similarly, genealogy databases are not all the same. Some of them are digital archives that provide digitized historic documents,

16 Karin van Es, Maranke Wieringa, and Mirko Tobias Schäfer, "Tool Criticism: From Digital Methods to Digital Methodology," in *Proceedings of the 2nd International Conference on Web Studies* (New York: Association for Computing Machinery, 2018), 24–27, https://doi.org/10.1145/3240431.3240436.

17 Es, Wieringa, and Schäfer, "Tool Criticism," 26.

18 Es, Wieringa, and Schäfer, "Tool Criticism," 25–26.

such as marriage certificates, boarding passes, or birth records – often behind paywalls. Prominent examples include *Ancestry* and *FamilySearch*.

Other genealogy databases like *Geni* are encyclopedic platforms, where users can gather and conflate information to reconstruct family trees. Standards for referencing original sources vary across different online communities and, of course, from user to user. As with other online encyclopedias, the documentation of the research in itself might be as or more revealing than the articles themselves. In the case of different spellings of names or missing or faulty OCR, sources that should theoretically be discovered by text search alone may be revealed by the work and documentation of other (amateur) researchers. A third type of genealogy database, such as *genealogy.net* or *GlobalGenSearch*, are essentially meta-catalogues – specialized search engines that scan many different websites under the genealogy label. It's hard to evaluate how thorough they are, but at least they are valuable tools to double check potential search results when working with genealogy databases. To conclude the example of genealogy databases (whose potential and limitations for film historiography could be discussed in a separate article): when researching digital sources from a genealogy database with the intention of making them accessible through one's research, it may be important to reflect on, or contextualize, which type of database provides which information. In this context, a collection of heterogeneous data should capture these references or documentations.

Creating and Providing Data Collections

Data collections can be seen as a methodological approach to the conflation of digital sources. They can be characterized as accumulations of consistently restructured datasets that originate from different databases.[19] A dataset in this case is a digital representation or model of a historical source in a broader sense. Data collections are a valuable tool in humanities research because they focus on a concrete subject rather than simply representing the historical collection of an archive, library, etc.[20] There are fundamental differences between digital research approaches that investigate a particular topic instead of an already existing database. The latter approach deals with large sets of homogeneous data and has to work within the rather static framework of the database. The former, on the other hand, requires researchers to collect and structure data in the first place to evaluate it further. As Johanna

19 Christof Schöch, "Aufbau von Datensammlungen," in *Digital Humanities – Eine Einführung*, ed. Fotis Jannidis, Hubertus Kohle, and Malte Rehbein (Stuttgart: Metzler, 2017), 223.
20 Schöch, "Aufbau von Datensammlungen," 227.

Drucker suggests, contrary to the impression that data simply exist, data must always be made.[21] While Drucker establishes this concept regarding modeling data, I would argue that, in the case of data collections, making data extends the modeling part of the process. In Christof Schöch's contribution to *Digital Humanities – An Introduction*, the process of creating a data collection is structured as follows:

1. Defining the subject of the data collection;
2. Researching Information on the subject from a broad perspective;
3. Selecting datasets according to different strategies;
4. Gathering, modeling, and cleaning the datasets;
5. Adding metadata;
6. Making the collection accessible.[22]

As the author points out, these phases of the process aren't strictly separated, but rather intertwined.[23] Projects like this need to be approached holistically. When accumulating highly heterogeneous sets of data or information, and aiming to conflate them in a consistent form, it is crucial to constantly adjust the structure of your data collection – and perhaps your research goals as well – according to your findings. This gives the source, data, and tool criticism what can be described as a historiographic dimension of modeling.

Researching DOÑA FRANCISQUITA with Digital Methods

In 1933/1934, before the Spanish Civil War, film exiles from Germany found a flourishing film industry in Barcelona, which was just beginning to produce sound films on a large scale. As a sound film operetta, DOÑA FRANCISQUITA suited the idea of reaching a wide audience. Together with Kurt Flatau, producer David Oliver hoped to conquer a worldwide Spanish-speaking market with the film.

A consideration of the opening credits of the film DOÑA FRANCISQUITA gives a first hint of the international scope of the crew: among the Spanish surnames of the actors in the film are names such as Herbert Lippschitz, Hans Behrendt, and Heinrich Gärtner. A closer look at the film reveals aesthetic nuances in the cinematography of Heinrich Gärtner, demonstrating a range of camera experience,

21 Johanna Drucker, *The Digital Humanities Coursebook: An Introduction to Digital Methods for Research and Scholarship*, 1st ed. (Abingdon, Oxon; New York: Routledge, 2021), 19.
22 Schöch, "Aufbau von Datensammlungen," 232.
23 Schöch, "Aufbau von Datensammlungen," 232.

such as a well thought out lighting design and sweeping camera pans, suggesting studio experience. The inspection of the credits and the aesthetics reveals a peculiarity of the film already visible in many details.

Personal Data on Filmographic Databases

The names in the credits allow us to find out who worked on the film. Searching their names via IMDb allows us, in some cases, to find additional information, such as short biographies. In the existing biographies, places of birth and death can be found and might be a first suggestion of an exile background: e.g., Vienna as a place of birth and Los Angeles as a place of death. Using IMDb's filmographic data on the film crew, a first list of German exiles emerges, which now needs to be verified with other sources. The people who are identified with a presumed exile background can sometimes be compared with existing film historical research on them: part of the German film crew had previously shot the film Gado Bravo (PT 1934) in Portugal,[24] before they were hired for Doña Francisquita. They included cameraman Heinrich Gärtner and Herbert Lippschitz, who was responsible for the film's architecture. Also from Germany were director Hans Behrendt, screenwriter Hans Jacoby, and editor Paul Falkenberg. The film's music was adapted by Jean Gilbert, and 23-year-old Peter Paul Weinschenk served as assistant cameraman. Kurt Flatau from Berlin organized the production on location with Edith Oliver, the producer David Oliver's wife.[25] The verification and review of a second source is a prerequisite for continuing to work on the people who were found through the name search, as the IMDb should be treated as a first clue and not a reliable source.

Norm Data of the German National Library

To discover whether a person appears in other sources with the same or similar biographical data,[26] the GND, the norm file of the German National library, is

24 Hagener Malte, "Nationale Filmproduktion und Exil: Zur Produktion und Rezeption des Films Gado Bravo," in *Exil in Portugal*, ed. Stiftung Deutsche Kinemathek (Munich: Text + Kritik, 2002), 52–53.
25 Ufa Man, "The Story of Ibérica Films," accessed January 13, 2023, https://ufa-man.com/category/original-sources/#jp-carousel-306.
26 A difference of one to three years in the dates of life is not at all unusual in the field, since in some cases passports and birth certificates were no longer available, and actors in exile sometimes pretended to be younger for the job market.

used.[27] For the film exiles and their name verification, a relevant source is the Straschek Estate at the Exile Archive in the German National Library in Frankfurt am Main, which was successively transferred to the GND until 2022. Günter Peter Straschek spent over thirty years researching the German film exile.[28] He sent out questionnaires about biographical details and escape routes, corresponded with exiles worldwide in letters over a period of years, conducted telephone interviews, visited exiles, and meticulously searched Austrian, German, English, and Dutch archives for traces of film exiles. His estate is now listed in the National Library's Exile Archive with 3,607 archival records on film exile, including over 3,500 personal files, each of which has a GND ID created or linked to an existing GND ID. For example, a personal file on DOÑA FRANCISQUITA's co-producer Kurt Flatau can be found here.[29] However, the Flatau file may not contain information about DOÑA FRANCISQUITA. The specific contents of the files can only be determined with certainty by visiting the Deutsche Exilarchiv in Frankfurt am Main. But caution is called for with the Straschek collection because a file being in the Straschek Archive does not in itself imply an exile past. Straschek also collected the marginal figures, the uncertain traces of people who left Germany and returned before 1945, who went into hiding, but in some cases worked in the German film industry again.[30]

Another problem for digital research is the different spellings of the names: pseudonyms, birth names, spouses' names, and anglicized names. Sometimes several people can be found under one name. Herbert Lippschitz is listed as "Arquitecto" in the film's opening credits. Via the name search of the National Library (GND), we find Arnold Lipschitz when looking for *Herbert Lippschitz*; however, he is noted as a screenwriter and not as an architect. The screenwriter is listed with the primary name *Arnold Phillips*, but also several pseudonyms: *Arnold Lippschitz*, *Arnold Lipp*, *Erich Philippi*, *Arnold Lippschütz* (real name), and *Arnold Phillip-Lipschitz*.

Was the screenwriter also an architect, or are they two different people? Spain is not listed as a country of exile for Arnold Lipschitz/Phillips in the GND, but the GND does show one family connection to a brother called Herbert O. Phillips. His first name gives us the idea for another investigation, and we find that Herbert

27 See "The data are reliable," accessed January 13, 2023, https://gnd.network/Webs/gnd/DE/UeberGND/GNDEignungskriterien/eignungskriterien_node.html#doc831580bodyText4.

28 Imme Klages and Alexandra Schneider, "Mapping German Film Migration. Digital Film Historiography Using the Example of the Estate of Günter Peter Straschek," in *Archives and Museums of Exile*, ed. Bettina Bannasch et al. (Berlin: De Gruyter, 2019), 222–238.

29 See "Personal file Kurt Flatau in the Straschek estate," accessed January 13, 2023, https://d-nb.info/1167563905.

30 See "Mapping German Film Migration," accessed January 13, 2023, https://filmexil.uni-mainz.de.

O. Phillips is listed with another name, the *Herbert Lippschitz* we seek. Wikipedia (on the name *Herbert Lippschütz*) says Herbert Lippschitz was a celebrated film architect, who often designed the buildings for the films for which Arnold wrote the screenplays.[31]

This example shows that the task of verifying a person across various online platforms can be a challenge for film historians. To verify the filmographic data on IMDb means checking that there is not another ID created under the pseudonym of a person, under which further filmographic data on the same person is hidden. For this very reason cross-checking between different online platforms is essential. Another example of this is the screenwriter Egon Eis with the additional names Egon Eisler (birth name), Egons Eis, Baby van Eyss, Edgar Eis, Etienne Reynard, George Turner, Georg Turner, Thomas B. Foster, Tennyson-Holme, and other variations.[32] And different sources can be found under one of the names alone.

Digitized Sources on Web-Based Platforms

Original sources such as letters, production correspondence, or photographs can be found in personal inheritance papers in archives, source publications, online repositories, or photographic databases. DOÑA FRANCISQUITA's editor Paul Falkenberg represents an interesting case, as his estate is in the Deutsche Kinemathek in Berlin. An online search for his name leads to a project of the Arbeitskreis selbständiger Kultur-Institute e.V., which has created a detailed and well-researched portrait of the artist and, in this context, includes many digitized sources from the estate in the website: letters, audio recordings of an interview with Paul Falkenberg, photos, and materials on the film DOÑA FRANCISQUITA, including a program booklet in Spanish and correspondence with the production company *Ibérica Film*.[33] The project describes itself as a digital memory project. The information on the website is verified through the official institutions that are its co-sponsors, such as the Deutsche Kinemathek. The digitized letters provide essential information about Falkenberg's escape route from Germany, and, in its letterhead, Ibérica Film's correspondence reveals the address in Barcelona and its offices in Madrid, Seville, Valencia, and Bilbao. Digitized personal letters reveal insights from a private perspective, leaving a different kind of trace.

31 See "Herbert Lippschütz," accessed June 10, 2023, https://d-nb.info/gnd/1062032721.

32 See "Eis, Egon," accessed January 13, 2023, https://d-nb.info/gnd/122897617.

33 See "TSURIKRUFN!" Arbeitskreis selbständiger Kulturinstitute e.V., accessed January 13, 2023, https://www.tsurikrufn.de/portraits/falkenberg/.

The personal writing of the time demonstrates the fear and powerlessness of the exile situation and gives the researcher a unique insight that adds another dimension to the research. Falkenberg writes to Rudolf Maté from Paris on June 5, 1935: "And all these émigré films – we made them to the exclusion of Germany – stop. Only 'Aryan' films are made in Vienna, i.e. a Jew and subhuman like me cannot work there, not even in disguise. Et nous voilà à Paris. We await the things to come."[34] The letter illustrates the highly precarious nature of the work, the need for perseverance, and the dependence on other contracts, which always provided work for only a short time.

Genealogy Databases

Genealogy databases offer another way of adding biographical insight and cross-checking references. GlobalGenSearch,[35] a platform that allows searches in a variety of genealogy databases, provides a total of about 200 results for the first five exiles mentioned in this article (Heinrich Gärtner,[36] Herbert Lippschitz,[37] Hans Behrendt,[38] Hans Jacoby,[39] Paul Falkenberg[40]). Although the platform allows us to specify surname, first name, date and place of birth, and date of death, there are no further options or filters to narrow the search. Since search results will be listed that match the name category but lack information on death or birth, each result must be checked individually to confirm that it matches the exile in question. It should also be mentioned that although GlobalGenSearch searches the

34 Collection Kinemathek Berlin Estate Paul Falkenberg. Translation by the authors. Original: "Und diese ganzen Emigrantenfilme -wir haben nämlich unter Ausschluss von Deutschland gedreht- hören auf. Es werden in Wien nur noch 'arische' Filme gemacht, d.h. ein Jude und Untermensch wie ich kann dort, nicht einmal getarnt mitarbeiten. Et nous voilà à Paris. Wir harren der Dinge, die kommen sollen."

35 See "GlobalGenSearch," accessed January 22, 2023.

36 See "Heinrich Gärtner," accessed January 22, 2023, https://ggs.spdns.eu/?CN=Heinrich&SN=G%C3%A4rtner&BD=1895&DD=1962&PLACE=.

37 See "Herbert Lippschitz," accessed January 22, 2023, https://ggs.spdns.eu/index.php?CN=Herbert&SN=Lippschitz&BD=1904&DD=1972&PLACE=.

38 See "Hans Behrendt," accessed January 22, 2023, https://ggs.spdns.eu/index.php?CN=Hans&SN=Behrendt&BD=1889&DD=1942&PLACE=.

39 See "Hans Jacoby," accessed January 22, 2023, https://ggs.spdns.eu/index.php?CN=Hans+&SN=Jacoby&BD=1898&DD=1967&PLACE=.

40 See "Paul Falkenberg," accessed January 22, 2023, https://ggs.spdns.eu/index.php?CN=Paul+&SN=Falkenberg&BD=1903&DD=1986&PLACE=.

website Deutsche Biografie,[41] which contains eleven versions of Heinrich Gärtners' name,[42] this specific information doesn't seem to be implemented in the GlobalGenSearch search engine; thus the checking and verifying of names and persons as described above again becomes necessary. This example may illustrate how much further the search for biographical information can be extended beyond this exemplary case study.

Digital Periodical Archives

The personal data and film data cannot tell us much about the context of the film city of Barcelona in 1934, except that the IMDb mentions the production company Ibérica Films. The Ibérica letter in Paul Falkenberg's estate, in turn, shows Kurt Flatau's signature under the production company document. Kurt Flatau also appears in the digitized documents of producer David Oliver, made available online by his grandson Marc Oliver.[43]

Was Kurt Flatau David Oliver's representative in Spain? Sabine Pamperrien, a freelance journalist and author, writes in an online article about David Oliver, published in the Weser-Kurier, that he was still in Germany in 1934.

> When Jewish artists were forced out of the film business after the Nazis came to power, he [David Oliver] founded Iberica Films in Barcelona in 1933, where numerous Jewish filmmakers and family members found employment. He remained in Berlin with his family, and as late as April 1933 celebrated the marriage of his only daughter Gertrud to a son of the respected Hamburg coffee-roasting family J.W. Darboven. [. . .] An assassination attempt in March 1934 caused a worldwide sensation when David Oliver's limousine was hit by a hand grenade on Unter den Linden in Berlin.[44]

This information is from April 23, 2022, and historical sources are not cited in the article. Only a telephone conversation with the journalist can confirm that the data comes from the City Archives in Bremen and the Federal Archives. The infor-

41 See "Deutsche Biographie," accessed January 22, 2023, https://www.deutsche-biographie.de/home.

42 See "Heinrich Gärtner," accessed January 22, 2023, https://www.deutsche-biographie.de/pnd137417241.html.

43 Ufa Man, "The Story of Ibérica Films," accessed January 22, 2023, https://ufa-man.com/category/original-sources/#jp-carousel-306. In the photographs of the grandson Marc Oliver, he introduces his grandmother Edith Oliver as Production Chief of Ibérica.

44 See "Ein Filmproduzent mit Bremer Vergangenheit," Sabine Pamperrien, accessed January 13, 2023, https://wkgeschichte.weser-kurier.de/ein-filmproduzent-mit-bremer-vergangenheit/?fbclid=IwAR1tJGe66MP0CB4n2Ji2SWxJ29jw89QLzQLfvKI2CVctfCHXJwjtGhwSkKg.

mation from journalistic articles on the film exile must be checked if they do not indicate any sources. This research work is a time-consuming and almost insurmountable task, because the various files and information on exiles are available not only in the above-mentioned archives (see footnote 3), but also in city and state archives, the Federal Archive, national cinema and film libraries, film museums in Europe and the wider world, in the exile press and the Paul Marcus Newsletter (PEM), the archives of various embassies, and in suitcases in attics that we do not yet know about, scattered and invariably only discoverable in excerpts.

In order to learn more about the historical context, production, and reception of DOÑA FRANCISQUITA, we turn to digitized film journals. The ZDB journal database lists 48 digitized film journals,[45] including the Austrian paper *Mein Film* and the Spanish *Popular Film*. Through the ZDB, there is a direct link to the respective libraries that hold the digitized material; for example, *Popular Film* links directly to the Hemeroteca Digital of the Biblioteca Nacional de España and to the repository of the Filmoteca de Catalunya.[46] The articles are in Spanish, and the high-resolution PDFs of the journals are also available to download. It is difficult to search the individual journals by name and the film title, however. On the page of the Filmoteca Catalunya, there is still a search window within the open PDF of the individual magazine, but one has to supply the exact name; no letters may differ, otherwise you will not find the person in question. In the case of the cameraman Heinrich Gärtner, for example, the German name is not to be found, but his Spanish name Enrique Guerner is, if one knew of it beforehand. The individual copies of the journals can be retrieved and then searched by opening each PDF.

The various repositories for film periodicals also differ greatly in their search options. The German periodicals database ZDB redirects a search for the historic context of the film DOÑA FRANCISQUITA to the Austrian National Library Website, except that the ZDB portal takes you directly to the selection of publication years for the magazine *Mein Film*.[47] For the production year of the film in 1934 alone, fifty magazine covers are displayed. Each issue can be downloaded individually as a PDF and all pages can be viewed in thumbnail format. It is like looking through a microfiche copy in an archive.

45 See "Film," accessed January 13, 2023, https://zdb-katalog.de/list.xhtml?t=Film&dig=digitalisiert&asc=false.

46 See "Popular Film," Biblioteca Nacional de España and Generalitat de Catalunya, accessed January 13, 2023, https://hemerotecadigital.bne.es/hd/card?oid=0004189218 and https://repositori.filmoteca.cat/handle/11091/8836.

47 See "Jahresauswahl – Mein Film," Österreichische Nationalbibliothek, accessed January 13, 2023, https://anno.onb.ac.at/cgi-content/anno-plus?aid=mfi.

If one enters the film title DOÑA FRANCISQUITA in the search mask of the Austrian National Library, 13 results appear, which already show the hits in the text and directly link to the digital copy via a "hits in the text"-button. The problem is, however, that none of the references of the "hits in the text" lead to the film title of 1934, but to the music of the Spanish zarzuela, which was performed in plays before 1934, among other things on radio programs, which announced a music performance of the famous zarzuela in Madrid.

The article in the magazine "Mein Film"[48] on the Spanish film industry in 1934 is filtered out and found only after a long search, using historical association combined with archival experience; only the trained eye can discover it more easily[49] and filter it out of the multitude of digitized pages. Examining the article, one discovers that the film is spelled differently: "Doña Francesquitta" with "e" and "tt," so that the film title search did not lead directly to the article. Despite these limitations, the "hits in text" display is a great asset to film historical research, allowing for serendipitous discoveries that might have taken months if the microfiche had had to be searched.

A search in the Media History Digital Library for the film title DOÑA FRANCISQUITA across the entire database for the period 1934 and 1935, returns 174 results,[50] of which only the first eighteen relate to the film DOÑA FRANCISQUITA, as the other hits contain only the word "Doña" in their text. The short notices in *Variety* and *Film Daily* about DOÑA FRANCISQUITA in their "Foreign Films" section contain information about the European distribution context of the film. For example, the *Film Daily* article of November 26, 1935, indicates that the film, along with 52 other foreign features, was picked up for distribution by United Artists Distribution. "Most of the pictures being handled for foreign distribution are of foreign make and are being distributed in most cases in countries or territories where they have particular appeal."[51]

In 1935, United Artists acquired the distribution rights for 19 feature films, including DOÑA FRANCISQUITA, and four short films from Barcelona-based Ibérica. This news further clarifies the financing models, or the distribution channels of

48 See "Mein Film: Filmproduktion in Spanien," Österreichische Nationalbibliothek, accessed January 13, 2023, https://anno.onb.ac.at/cgi-content/anno-plus?aid=mfi&datum=1934&page=716&size=45.

49 We would like to thank Helmut G. Asper for pointing out the article.

50 See "Doña Francisquita," Wisconsin Center for Film and Theater Research, accessed January 13, 2023, https://lantern.mediahist.org/catalog?op=AND&keyword=Doña+Francisquita&second_keyword=&third_keyword=&title=&Author=&subject=&dateString=&publisher=&description=&sort=score+desc%2C+dateStart+desc%2C+title+asc&search_field=advanced&range%5Byear%5D%5Bbegin%5D=1934&range%5Byear%5D%5Bend%5D=1935&commit=Refine.

51 N.N., "U.A. handling foreign films outside U.S.," *Film Daily* 68, no. 125 (1935): 10.

the film. The large Digital Media History Library portal is helpful for this type of research information. Each portal needs to be examined and checked for resources on the topic in question. The website of the Spanish Film Library offers more film reviews and articles on the making of the films, as well as stories about the stars and the popularity of the actors.

Modeling the Data Collection

The information found is transformed into a data collection by supplementing and enriching the biographical information and data. The data collection on DOÑA FRANCISQUITA is realized as a spreadsheet (Figure 1). It can be described as a digital (partial and exemplary) listing of sources on the film's production crew. Reflecting our findings, the data collection includes different versions of names, as well as more basic data, such as job descriptions and biographical information. Each piece of data collected is attributed to its source. Each of the source archives or cultural heritage institutions is briefly contextualized and located to emphasize the transnational perspective.

As a humanities researcher, one tends to find epistemological rather than technical solutions for integrating different sets of information. One must be open-minded about the structuring and selection of information during the process of collecting as well as modeling. Again, engaging with the data as traces reflects and emphasizes the need for constant critical engagement with data and tools during the research, the selection and structuring of information, and the contextualization. Although the critical framing of information and connecting of traces isn't implemented as data in the data collection on DOÑA FRANCISQUITA, the data's selection and structure directly reflect these processes. The research question, the available information, the accumulation and structuring of data, and the concepts or ideas for displaying the findings (visually or in text) constantly influence each other, making the process more fluid and holistic.

Name in the original film credits	Credited as	Verified / identified via IMDb (as part of the crew) as	Credited as (on IMDb)	Verified / identified through other sources as	Credited as (in the source)	Source	Origin of source	Type of platform	Verified through Archiv Günter Peter Straschek on DNB
Fernandez Shaw	inspirado en la zarzuela del mismo nombre de los Senores	Guillermo Fernández Shaw	"Writing Credits" (book)						
Federico Romero	inspirado en la zarzuela del mismo nombre de los Senores	Federico Romero	"Writing Credits" (book)						
Francisco Elias	dialogada por	Francisco Elías	"Writing Credits"						
D. Amadeo Vives	música del insigne maestro	Amadeo Vives	"Music by"						
Jean Gilbert	**adoptada para la película por el maestro**	Jean Gilbert	**"Music by"**						Konvolut von Unterlagen zu Jean Gilbert
Hans Behrendt	**DIRECTOR:**	Hans Behrendt	**"Directed by"**						Konvolut von Unterlagen zu Hans Behrendt
José Vivés Giner	realizada bajo la supervisión de	-	-						
Luis Marquina	Ingeniero de sonido	-	-						
Heinrich Gaertner	**Camera:**	Heinrich Gärtner [as Enrique Guerner]	**"Cinematography by"**						Konvolut von Unterlagen zu Heinrich Gärtner
Herbert Lippschitz	**Arquitecto**	-	-	Herbert Lippschütz	*Filmarchitekt*	Article	*Wikipedia*	*Online Encyclopedia*	Konvolut von Unterlagen zu Arnold Phillips
-	-	Hans Jacoby	**"Writing Credits"**						Konvolut von Unterlagen zu Hans Jacoby
-	-	Paul Falkenberg	**"Editing by"**						Konvolut von Unterlagen zu Paul Falkenberg
-	-	-	-	David Oliver	*Producer*	Digitised Prints and Photos	*UFA MAN*	*Grandson David Oliver's Private Archive*	Konvolut von Unterlagen zu David Oliver
-	-	-	-	Kurt-Louis Flatau	*Co-Producer*	Digitised Prints and Photos	*UFA MAN*	*Grandson David Oliver's Private Archive*	Konvolut von Unterlagen zu Kurt Flaufau
-	-	-	-	Edith Oliver	*Production Chief*	Digitised Prints and Photos	*UFA MAN*	*Grandson David Oliver's Private Archive*	-

Figure 1: Screenshot (detail) of the data collection realized as a spreadsheet.[52]

52 Imme Klages and Fabian Kling, "Data Collection: Crew of DONA FRANCISQUITA (ESP 1934) [Dataset]," *Zenodo*, accessed March 26, 2024, https://doi.org/10.5281/zenodo.10657958.

Conclusion

The sources found on DOÑA FRANCISQUITA's film crew and its production in Spain had to be sorted and put into context, matching the available data and the sought-after information. This process has to be constantly connected with the following questions for film-historical work: Which sources are available at all? Which information will be used? Which ideas will lead the research further? The digital research path is determined by constant reflections on one's process, including knowledge gaps and further questions. It is an iterative process, as new sources open new perspectives on research questions. Existing encyclopedic articles on the film's collaborators only create a starting point as further research on the connections between the film and the film crew must be collected. The research process always includes reflection on one's approach and findings.

At the Vienna conference on the occasion of the founding of the Vienna Wiesenthal Institute for Holocaust Studies on June 7–8, 2006, historian and Holocaust researcher Raul Hilberg described, with a gentle humor that belies his years of research, his work in the archives in Washington in 1951/1952. He reflected about how he and eight colleagues had to process 10,000 meters of Leitz folders of files produced by the National Socialists during their years in power: "There I developed my method, which I don't recommend to anyone, of blindly pulling out the files, it doesn't even matter what they are. And that's where you find what you're not looking for."[53]

In digital film historical research, there's no choice but to proceed blindly at first since a simple review no longer does justice to the ever-increasing number of sources and data. It requires a different mindset to adjust your research and modeling of data to reflect the necessary critical approach towards them. To envision data as traces means reflecting on their incompleteness, emphasizing the necessity to connect and contextualize them, and acknowledging their resistance to being comprehensively connected and categorized. While this chapter refuses to conclude with a concrete methodological framework for certain research designs, by reflecting on a particular case study with its specific challenges, it has outlined how theoretical perspectives can constructively influence practical research.

53 Raul Hilberg and Walter Manoschek, "The Legacy of Simon Wiesenthal for Holocaust Studies." Lecture in German at the IFK Internationales Forschungszentrum Kulturwissenschaften in Vienna, June 8, 2006, accessed January 13, 2023, https://www.youtube.com/watch?v=epUoqWxwjj8&t=908s&ab_channel=WienerWiesenthalInstitutf%C3%BCrHolocaust-Studien (Minute: 13:13).

The DFG research project "Mapping German Film Migration 1930–1950" (2021–2024) produced some of the research findings on German film exiles.[54]

References

Asper, Helmut. *"Etwas Besseres als den Tod . . .": Filmexil in Hollywood: Porträts, Filme, Dokumente*. Marburg: Schüren, 2002.

Biblioteca Nacional de España. "Popular Film." Accessed January 13, 2023. https://hemerotecadigital.bne.es/hd/card?oid=0004189218.

Burkhardt, Marcus. *Digitale Datenbanken: Eine Medientheorie im Zeitalter von Big Data*. Bielefeld: Transcript, 2015. https://doi.org/10.25969/mediarep/835.

Camporesi, Valeria, and Fernando Gonzalez Garcia. "Un progreso en el arte nacional? Ibérica Films en España, 1933–1936." *Boletín del Seminario de Estudios de Arte y Arqueología* 77 (2011): 265–285.

Camporesi, Valeria. "The Tuneful 1930s. Spanish Musicals in a Global Context." In *Global Genres, Local Films. The Transnational Dimension of Spanish Cinema*, edited by Elena Oliete-Aldea, Beatriz Oria, and Juan A. Tarancón, 19–30. New York: Bloomsbury, 2017.

Deutsche Nationalbibliothek. "Eis, Egon." Accessed January 13, 2023. https://d-nb.info/gnd/122897617.

Drucker, Johanna. *The Digital Humanities Coursebook: An Introduction to Digital Methods for Research and Scholarship*. Abingdon, Oxon, New York: Routledge, 2021.

Es, Karin van, Maranke Wieringa, and Mirko Tobias Schäfer. "Tool Criticism: From Digital Methods to Digital Methodology." In *Proceedings of the 2nd International Conference on Web Studies*, 24–27. New York: Association for Computing Machinery, 2018. https://doi.org/10.1145/3240431.3240436.

Filmoteca Española. "Filmoteca Española estrena la restauración de Doña Francisquita, un clásico de la zarzuela llevado a la gran pantalla en 1934." Accessed February 21, 2023. https://www.cultura.gob.es/cultura/areas/cine/mc/fe/comunes/noticias/2019/10/donafrancisquita.html.

Generalitat de Catalunya. "Popular Film: 1926–1937." Accessed January 13, 2023. https://hemerotecadigital.bne.es/hd/card?oid=0004189218 and https://repositori.filmoteca.cat/handle/11091/8836.

Ginzburg, Carlo. "Morelli, Freud and Sherlock Holmes: Clues and Scientific Method." Translated by Anna Davin. *History Workshop* 9 (1980): 5–36.

Ginzburg, Carlo. "Spie. Radici di un paradigma indiziario." In *Miti emblemi spie. Morfologia e storia*, edited by Carlo Ginzburg, 159–209. Torino: Einaudi, 1986.

GlobalGenSearch. "GlobalGenSearch." Accessed January 22, 2023. https://ggs.spdns.eu/.

Hagener, Malte. "Nationale Filmproduktion und Exil. Zur Produktion und Rezeption des Films Gado Bravo." In *Exil in Portugal*, edited by Stiftung Deutsche Kinemathek, 50–69. Munich: Text + Kritik, 2002.

Heinink, J. B., and Alfonso C. Vallejo. *Catálogo del cine español. Volumen F3, Films de ficción 1931–1940*. Madrid: Cátedra/Filmoteca Española, 2009.

54 Johannes Gutenberg-Universität Mainz, accessed June 30, 2023, https://filmexil.uni-mainz.de.

Hilberg, Raul, and Walter Manoschek. “The Legacy of Simon Wiesenthal for Holocaust Studies.” Lecture in German at the IFK Internationales Forschungszentrum Kulturwissenschaften in Vienna, June 8, 2006. Accessed January 13, 2023. https://www.youtube.com/watch?v=epUoqWxwjj8&t=908s&ab_channel=WienerWiesenthalInstitutf%C3%BCrHolocaust-Studien.

Hilberg, Raul. “Zur Integration des Wissens um die Judenvernichtung in die Gedenk- und Erinnerungsarbeit.” Lecture in German at the 14th Annual Conference of Alemannia Judaica, Breisach, Germany, March 18, 2001. Accessed January 13, 2023. https://www.youtube.com/watch?v=kmpRMaN2RH0.

Klages, Imme, and Alexandra Schneider. “Mapping German Film Migration. Digital Film Historiography Using the Example of the Estate of Günter Peter Straschek.” In *Archives and Museums of Exile*, edited by Bettina Bannasch, Doerte Bischoff, and Burcu Dogramaci, 222–238. Berlin: De Gruyter, 2019.

Klages, Imme, and Fabian Kling. “Data Collection: Crew of DONA FRANCISQUITA (ESP 1934) [Dataset].” *Zenodo*. Accessed March 26, 2024. https://doi.org/10.5281/zenodo.10657958.

Munoz Aunion, Marta, and Fernando Gonzalez Garcia. “El paraiso de los amigos de cine.” In *Faros y torres vigía: el cine español durante la II República (1931–1939)*, edited by Julio Pérez Perucha and Agustín Rubio Alcover, 149–164. Madrid: Asociación Española de Historiadores del Cine, 2016.

N.N. “U.A. handling foreign films outside U.S.” *Film Daily* 68, no. 125 (1935): 10.

Österreichische Nationalbibliothek. “Jahresauswahl – Mein Film.” Accessed January 13, 2023. https://anno.onb.ac.at/cgi-content/anno-plus?aid=mfi.

Österreichische Nationalbibliothek. “Mein Film: Filmproduktion in Spanien.” Accessed January 13, 2023. https://anno.onb.ac.at/cgi-content/anno-plus?aid=mfi&datum=1934&page=716&size=45.

Reigeluth, Tyler. “Why data is not enough: Digital traces as control of self and self-control.” *Surveillance & Society* 12, no. 2 (2014): 243–254.

Ricoeur, Paul. *Memory, History, Forgetting*. Translated by Kathleen Blamey and David Pellauer. Chicago: University of Chicago Press, 2004.

Schöch, Christof. “Aufbau von Datensammlungen.” In *Digital Humanities – Eine Einführung*, edited by Fotis Jannidis, Hubertus Kohle, and Malte Rehbein, 223–233. Stuttgart: Metzler, 2017.

Wisconsin Center for Film and Theater Research. “Doña Francisquita.” Accessed January 13, 2023. https://lantern.mediahist.org/catalog?op=AND&keyword=Doña+Francisquita&second_keyword=&third_keyword=&title=&Author=&subject=&dateString=&publisher=&description=&sort=score+desc%2C+dateStart+desc%2C+title+asc&search_field=advanced&range%5Byear%5D%5Bbegin%5D=1934&range%5Byear%5D%5Bend%5D=1935&commit=Refine.

Zeitschriften Datenbank. “Film.” Accessed January 13, 2023. https://zdb-katalog.de/list.xhtml?t=Film&dig=digitalisiert&asc=false.

Zeitschriften Datenbank. “Mein Film: die österreichische Filmillustrierte.” Accessed January 13, 2023. https://zdb-katalog.de/title.xhtml?idn=1141914123&view=full.

Filmography

Doña Francisquita, dir. Hans Behrendt. Spain, 1934.

Gado Bravo, dir. António Lopes Ribeiro and Max Nosseck. Portugal, 1934.

Nicole Braida and Frauke Pirk

Teaching Small-Gauge Formats with Digital Methods

Introduction

Working with digital methods and data challenges our way of doing Film Studies. Taking as a starting point a specific teaching module that we experimented with in the last two years, this chapter offers some insights into the challenges of teaching digital methods for undergraduates in film history. We asked ourselves what the best methods and tools for tackling small-gauge film data were. What might this methodological shift add to the study of film historiography, and how can we teach students new skills? In what follows, we describe our experience in teaching a generation of students who are the so-called "digital natives."[1] We introduced the students to data criticism, data cleaning, data enrichment, and data visualization. In this chapter, we do not consider Film Studies and Digital Humanities (DH) as separate disciplines; instead, we aim to combine them into Digital Film Studies. And although we characterize Digital Humanities as a discipline, it is important to note that there is not a final understanding of what DH is and what should be included.[2]

We designed and taught an undergraduate course on small-gauge film formats as a starting point in order to familiarize students with issues of doing film history with digital methods. Other experiments, such as those made by Susan Aasman with small-gauge and amateur video material, have shown the difficulty of tackling so-called home movies through digital methods and formalistic ap-

1 Marc Prensky, "Digital Natives, Digital Immigrants Part 2: Do They Really Think Differently?" *On the Horizon* 9, no. 6 (2001): 1–6, https://doi.org/10.1108/10748120110424843.

2 For further information see, for example, Manuel Burghardt, "Big Tent Digital Humanities und die Entwicklung der Computational Humanities. Aktuelle Trends und Fallstudien für Computergestützte Verfahren in den Geistes- und Kulturwissenschaften" (presentation at the Interdisziplinäre Hybrid-Vorlesungsreihe zum Themenschwerpunkt "Digital Humanities" – Wintersemester 2020/2021, Johannes Gutenberg-Universität Mainz, December 10, 2020), accessed February 15, 2023, https://video.uni-mainz.de/Panopto/Pages/Viewer.aspx?id=e5f175a4-eb21-4caa-8ca7-ac8c0108cd24; and, with a focus on interdisciplinarity, see Julie Thompson Klein, "Introduction: Emerging," in *Interdisciplining Digital Humanities: Boundary Work in an Emerging Field* (Ann Arbor: University of Michigan Press, 2015), 1–13, accessed July 21, 2023, http://www.jstor.org/stable/j.ctv65swxd.5.

https://doi.org/10.1515/9783111082486-006

proaches.[3] In our hands-on experience, we highlighted the challenges of working mostly with metadata about small-gauge films. The development of this film-historical exercise aimed at giving both an introduction to amateur and non-theatrical film formats and also a tentative overview of key Digital Humanities practices and concepts. In hands-on sessions, based on specific research questions, students learned to gather and structure datasets, to elaborate and clean data, and finally to present and conceive a visualization.

Although research at the intersection of Film Studies and digital methods has already discussed the challenges of approaching film history from a Digital Humanities perspective, less has been said about how to implement the necessary skills for such methodologies within Film Studies curricula.[4] We decided to lean on the work of Johanna Drucker to frame our approach, following the triad of "materials + processing + presentation," thus learning first about the material and the way we might process it through different tools, and finally how to present it.[5] Two iterations of teaching this course[6] has exposed several implications of applying digital methods as hands-on tools in a Film Studies classroom, beyond the fact that many humanities students bring no prior data literacy with them. In this contribution, we discuss how, although we departed from a specific framework for teaching, we left a large space for creating a heuristic line of attack. A trial-and-error approach helped adjust different needs and expectations from us as teachers and researchers and from undergraduate students from Film Studies. Based on our hands-on experience, we consider which data cleaning and analysis tools we used and how we adjusted them, and which strategies we developed for the specific small-gauge material that we and our students worked on. As an outcome, we propose a series of principles that might be helpful to frame similar courses.

3 Susan Aasman, "Unlocking Multiple Histories of Amateur Media: From Micro- to Macro-Histories," *Screen* 61, no. 1 (2020): 157–166, https://doi.org/10.1093/screen/hjaa011.

4 Taylor Arnold et al., "Introduction: Special Issue on AudioVisual Data in DH," *Digital Humanities Quarterly* 15, no. 1 (2021), accessed July 21, 2024, https://www.digitalhumanities.org/dhq/vol/15/1/000541/000541.html; Aasman, "Unlocking Multiple Histories of Amateur Media"; Rossella Catanese, Adelheid Heftberger, and Christian Gosvig Olesen, "Computer-Based Approaches to Film Archiving, Restoration and Philology," *Cinergie – Il Cinema e Le Altre Arti* 20 (2021): 1–6, https://doi.org/10.6092/ISSN.2280-9481/13948; Julia Noordegraaf, "Computational Research in Media Studies: Methodological Implications," *Tijdschrift Kwalon* 21, no. 1 (2018): 52–59; Christian Gosvig Olesen et al., "Data-Driven Research for Film History: Exploring the Jean Desmet Collection," *The Moving Image* 16, no. 1 (2016): 82–105.

5 Johanna Drucker, *The Digital Humanities Coursebook: An Introduction to Digital Methods for Research and Scholarship*, 1st ed. (Abingdon, Oxon; New York: Routledge, 2021).

6 The course was taught in two successive semesters, first in 2021/2022 and then in 2022/2023.

Re-Scaling Film Studies with Digital Methods

Our course was developed in the context of the project DiCi-Hub – A Research Hub for Digital Film Studies.[7] DiCi-Hub is funded for five years by the Volkswagen Stiftung (2021–2026) in the funding parameter "World Knowledge – Structural Support for 'Rare Subjects'."[8] The project aims to facilitate the structural development of our small discipline,[9] Film Studies. Our specific goal was to address the pressing issue of digital transformation, which calls for a profound reconfiguration of Film Studies. Digitization has not only led to a multitude of new forms and formats of the moving image, calling for an expansion of the meaning of "film," it has also led to an enormous growth of data and meta-data that accompany these new formats. In particular, DiCi-Hub combines hermeneutical and post-hermeneutical analytics of film with new digital methods and tools to reposition Film Studies as a discipline that turns the challenge of the digital into an opportunity for both research and teaching.

The project brings together three universities and key areas of film culture: Marburg, Mainz, and Frankfurt. While the three collaborated on a joint module on data criticism/data literacy, each has a specific thematic for research, providing opportunities to work on a broad variety of digital methods and tools, to implement them in our research and teaching, and to share our experiences to improve our workflows.

The focus in Mainz is on film formats. Drawing from previous research on format studies,[10] we investigate the technological infrastructures and conventions sustaining the flows of moving images. Because the Super 8 format on which we decided to work in the course, particularly what in German are called *Kauffilme*[11] (thus reduction prints), is not yet well researched and present in institutional archives,[12] we collaborated directly with private German collectors. With their sup-

7 See https://www.uni-marburg.de/de/fb09/medienwissenschaft/forschung/forschungsprojekte/dici-hub, accessed March 31, 2024.

8 See https://www.volkswagenstiftung.de/en/funding/funding-offer/world-knowledge-structural-support-rare-subjects-completed, accessed July 18, 2023.

9 See https://kleinefaecher.de/, accessed July 18, 2023.

10 Marek Jancovic, Axel Volmar, and Alexandra Schneider, eds., *Format Matters: Standards, Practices, and Politics in Media Cultures* (Lüneburg: meson press, 2019), https://doi.org/10.14619/1556.

11 Jeanpaul Goergen, "Filmgeschichte im Wohnzimmer. 16mm- und Super8-Kurzfassungen deutscher Filmklassiker," *Filmblatt* 7, nos. 19–20 (2002), accessed July 21, 2023, https://www.filmblatt.de/2002/10/01/filmblatt-7-jg-nr-19-20-winter-fruehjahr-2002/; Alexandra Schneider, "Viewer's Digest: Small-Gauge and Reduction Prints as Liminal Compression Formats," in *Format Matters*, ed. Jancovic, Volmar, and Schneider, 129–146, https://doi.org/10.14619/1556.

12 We first contacted few film archives, but none had sufficient material on Super 8 reductions prints; most Super 8 films are preserved as amateur movies and experimental films.

port and the support of a broader international community, we obtained other resources on Super 8, such as fanzines and film catalogs. We loaned and digitized some material,[13] and developed an OCR workflow to extract relevant information into our own database for further research. Catalogs and other material were not only part of our research, but were incorporated in our teaching module for digital methods. This starting point for re-scaling Film Studies offers a new way of looking at film history that wishes to keep a hermeneutical approach but also to learn how to work with data about film. If Film Studies are facing a transformation, we want to equip students with the right toolkits to take advantage of the growing availability of data and archival material online.

The Digital Native Fallacy

In 2022 we often expect students to be familiar with work within a digital environment, because we assume them to be very aware of the technology they deal with on a daily basis. The term "digital natives" was popularized in the early 2000s by Prensky[14] and indicates a new generation of students who are familiar with the use of technology and getting information through digital media. Nevertheless, our experience suggests that many students have limited experience with software. The misconception has already been demonstrated by Sonia Livingstone[15] but we would like to explain this observation by giving a more detailed insight from our own experience in the preparation of our teaching module.

Before even starting, we informed the students that they needed a laptop to participate in the course. Although we realized this was not the best solution, we did not have the capacity to lend hardware to participants. Due to the increasing number of students that only use a tablet, we decided to work on individual solutions and tried to establish a supportive network through their peers. We chose mainly software that is easy to install and that runs on all three main operating systems (Windows, MacOS, and Linux), since instructors do not know before the course which operating system individual students use. This meant that we had to test the software on different devices to ensure its functionality and to get a

13 The film catalogs of German distributors were given by Andreas Chmielewski who is part of the collector community.

14 Prensky, "Digital Natives, Digital Immigrants Part 2," 1–6.

15 Sonia Livingstone, "Enabling Media Literacy for 'Digital Natives' – a Contradiction in Terms?" in *"Digital Natives": a Myth?* (London: POLIS, London School of Economics and Political Science, 2009), 4–6, accessed July 21, 2023, https://eprints.lse.ac.uk/48944/.

feeling for problems that might occur. In our experience it turned out that the previous knowledge of the students in dealing with hardware and software was very heterogeneous, often determined on the basis of private interests and auto-didactically acquired skills, which made instructors' pre-course exploratory testing for anticipated challenges especially crucial. Finally, we privileged software that was either free to the public or available through an existing university subscription.

In Germany, there are no consistent computer science classes in high schools.[16] A large number of students have reported back to us that they only acquired the skills to work with the most common programs of the Microsoft Office package at school. Unfortunately, this does not seem to include the use of Excel or spreadsheets in a meaningful way. It also showed that the operation of hardware and software could be done very intuitively and hardly needed to be dealt with on a deeper level. So we created very detailed manuals for the installation and use of the software we employed in our course. We adapted and extended such manuals after the first year of our course. In addition, we tried to develop students' awareness of new terminology in order to facilitate their ability to solve problems independently in the future with the help of search engines.

It also became clear that dealing with datasets in the comma separated values (CSV) format was causing problems. In response, we then explained the differences between the formats, their individual areas of usage, and their respective advantages and disadvantages. We also discussed a procedure for importing CSV files into a spreadsheet program. For a faster and more effective way of working, we are currently discussing the possibility of implementing an optional basic computer course directly before the start of term so that we can narrow our focus on the contents of the course.

This experience points to the fallacious assumption that contemporary students, as "digital natives," are very comfortable with basic computer skills. Despite being surrounded by digital technology, the current student generation has the experience of working mostly with easy user-friendly interfaces, leading them to view computers largely as "black boxes," not as tools demanding skill for successful handling. This makes them less aware of the possibility of adjusting and setting basic computer procedures, like installing, managing, and setting the desktop environment. This lack of proficiency, as noted by a report from ICDL Europe,[17] may contribute to a digital divide, which we hope to address by implementing digital methods in Film Studies curricula as well.

16 See Informatik-Monitor 2022/2023, accessed July 18, 2023, https://informatik-monitor.de/.

17 "The Fallacy of the 'Digital Native'," *ICDL Europe* (blog), accessed February 13, 2023, https://www.icdleurope.org/policy-and-publications/the-fallacy-of-the-digital-native/.

New Concepts for Studying Film History

Besides the already mentioned technical requirements, studying film history with digital methods means bringing new concepts and approaches to an established field of studies. The task of combining a focus on small gauges with digital methods has been especially challenging because of a lack of hands-on manuals for working on small-gauge formats with these methods. We decided to lean on different sources: on one hand, literature from format studies, amateur film, and small-gauge cinema and, on the other, texts from Digital Humanities and data visualization.

In the first conception of the exercise in 2021, we focused on the broader topic of small-gauge formats. In our first class session, an expert gave an introductory lecture on 16mm films and on amateur formats,[18] and brought some film material, cameras, and projectors of several different formats. This starting point into the actual material proved to be very fruitful for engaging students in investigating small gauge, especially for those students with little background experience.

This was repeated the next year by inviting another expert, this time not an academic, but an amateur collector[19] and creator of a website specializing in Super 8 film formats and in the German distribution of "reduction prints," the reduced versions of films on smaller gauges.[20] We decided to devote class discussions to academic and popular literature about small gauges, literature on nontheatrical films, *Gebrauchsfilme*, for instance, or about 16mm or Super 8 film.[21] This thematic introduction prepared students with a useful overview on the his-

18 The person in question was Dr. Alexander Stark, who had worked on an amateur/professional German filmmaker (Elizabeth Wilms) who used mostly 16mm film material for her productions.

19 The collector was Joachim Schmidt, who also runs the website https://off2.de/, accessed March 31, 2024.

20 Schneider, "Viewer's Digest," 129.

21 Michael Teubig, "Was ist aus den Super-8-Vertrieben geworden?" off2 – Filme auf Super 8, 1987, accessed July 21, 2023, https://off2.de/publikationen/die-leinwand/wo-sind-die-super-8-firmen-l2-87/; Scott MacGillivray, *Castle Films: A Hobbyist's Guide* (Bloomington, IN: iUniverse, 2004); Vinzenz Hediger, "'Dann sind Bilder also nichts!' Vorüberlegungen zur Konstitution des Forschungsfelds 'Gebrauchsfilm'," *Montage/AV* 14, no. 2 (2005), 11–22, accessed July 21, 2023, https://doi.org/10.25969/MEDIAREP/220; Dan Streible, Martina Roepke, and Anke Mebold, "Introduction: Nontheatrical Film," *Film History* 19, no. 4 (2007), 339–343; Paul Eisloffel, "A Brief History of the 16mm Film Format," *MAC Newsletter* 41, no. 1 (2013), 32–33, accessed July 21, 2023, https://lib.dr.iastate.edu/macnewsletter/vol41/iss1/7; Alexandra Schneider, "Theorie des Amateur- und Gebrauchsfilms," in *Handbuch Filmtheorie*, ed. Bernhard Gross and Thomas Morsch (Wiesbaden: Springer, 2021), 225–242.

tory of the formats and helped them navigate and understand the data and resources they would eventually work on.

To provide an understanding of the tools and significance of the digital methods, it was necessary to acquaint students with several concepts. These would guide them in working on their project, and with a critical approach on tools as well. We drew on some basic literature from Digital Humanities to introduce them to the discipline and where possible selected research that entailed film material or elements from cinema history,[22] or alternative perspectives with a strong humanities perspective.[23] These readings, together with the initial introductory lessons, were aimed at answering the following questions:

- What is film-historical research?
- Which institutional and online resources are available for historical research?
- What is data/source criticism?
- What are digital repositories and databases?
- What is data and metadata?
- What are Digital Humanities and which projects of Film Studies use such approaches?
- What is data structuring and data modeling?
- What is data cleaning and data enriching?
- What is data visualization?

In both years, the course employed primarily a collaborative teaching modality, beginning by inviting experts/scholars to discuss one of the concepts with single contributions from their perspective. We followed a co-teaching model for the remaining topics, splitting the various thematics between us, depending on our own expertise.

Teaching methods for historical research are critical, especially for undergraduates who are often unfamiliar with material resources, the criteria of doing research, and being critical about sources (especially if dealing with online resources). An interesting tool for introducing this topic to familiarize them with film and audiovisual sources was the website from the University of Zürich "Ad Fontes."[24] We also discussed a series of repositories and databases useful for

22 Deb Verhoeven, "Visualising Data in Digital Cinema Studies: More than Just Going through the Motions?" *Alphaville. Journal of Film and Screen Media* 11 (2016): 92–104; Franco Moretti, "Planet Hollywood," *New Left Review* 9 (2001): 90–101.

23 Drucker, *The Digital Humanities Coursebook*.

24 See "Übungen zum Umgang mit Film- und Videoquellen," Universität Zürich, accessed February 7, 2023, https://www.adfontes.uzh.ch/384000/training/filmquellen.

small gauge (although not exclusively), such as the Media History Digital Library[25] or archive.org and distinguished between databases and digital repositories. Although these concepts are essential for historical research, we decided to reduce the amount of time allocated to them in the second year, restructuring the course by compressing the attention given to each individual theoretical topic, and leaving more space for hands-on work. We modeled the new structure around the hands-on sessions, as follows:

- Introduction to Working with Film Data (Program, Aims, Student Performance Requirements)
- Topic or Film Historical Introduction
- Structuring Data
- Enriching Data
- Digital Methods for Film Studies
- Exploring Data through Visualizations
- Presenting/Visualizing Data

For both years, we set a final group project in the form of a presentation to serve as an assessment of student performance, and reserved four to five sessions exclusively for group work. Creating the groups for this project at the very outset of the course, as we did in the second year, allowed students maximum opportunity to collaborate and organize their projects.

The final class sessions were dedicated to their presentations. In the first year, the pandemic forced us into an online modality, so we assigned students a Powerpoint or interactive presentation. In the second year we opted, thanks to the possibility of working together, for the modality of a poster presentation, borrowing a practice well-known in the natural sciences but less familiar in the customary practices of humanities research. The final project would present not only students' results in the form of visualizations, but also an overview of their sources (data sources, and material or literary sources), tools, and methods used. Students were also required to discuss their results and the challenges faced during every phase of the project development.

Compared to more traditional methods of student performance assessment, the hands-on work with software and the practice of working with data emphasizes the importance of processual performance. While a traditional text essay measures students' understanding by their ability to analyze a specific film text with the proper literary references, for instance, hands-on research projects

25 See https://mediahistoryproject.org, accessed March 15, 2024.

gauge understanding of essential concepts like data modeling by forcing students to critically and reflectively apply what they have learned.

Before participation in this course, many students seemed to lack awareness of any of the concepts that we considered essential for understanding the challenges of doing interdisciplinary research in the humanities. Students found an introduction to Digital Humanities concepts, such as data modeling or data cleaning, disorienting even before having had the opportunity to work directly on a project. With this in view, we hoped that the hands-on work provided the necessary experience to reflect on those concepts and a useful introduction to the process of Digital Humanities research.

Tools for Hands-On Work

Having structured the course, we needed to choose the tools we would employ. We were careful to test potential tools beforehand to understand their uses. The choice we made was based on a negotiation between a broader scholarly discussion within Film Studies circles and our own abilities and interests. We paid particular attention to several criteria when selecting tools. First, that they were accessible free or within the university network, second, that they were easy to learn for students with no previous knowledge, third, that they ran on the widely used operating systems, and, fourth, that they worked well with our own data sources.

The core functionality of the software required for the course, depending also on the student performance we required, was to find software that let us structure, analyze, and visualize the data. While numerous tools contain such functionality, we opted for these main three resources:

- Excel for structuring data and sharing datasets
- Open Refine for enriching and linking data, and also for structuring and cleaning data
- Tableau Public for analyzing, visualizing, and presenting data

Choosing a digital tool, ideally, means also to approach it with a critical distance[26] and not to accept it blindly as an instrument which brings unquestionable results. We did not give specific guidelines or criteria for reflection, but instead often dis-

26 Nicole Braida, Isadora Campregher Paiva, and Josephine Diecke, "Was machen wir mit digitalen Tools und was machen sie mit uns?" *Open Media Studies Blog*, February 2, 2022, accessed July 21, 2023, https://zfmedienwissenschaft.de/online/open-media-studies-blog/was-machen-wir-mit-digitalen-tools-und-was-machen-sie-mit-uns.

cussed issues together in class during the hands-on work. What emerged were questions on how to model our datasets depending on the tool and the consequence that that would bring, for example. We noted, for instance, that choosing specific geographical categories for data is often necessary for the visualization, but this demands taking decisions that do not always comply with the historical developments of a specific nation/territory. To work with digital tools and methods from the perspective of Film Studies offers us the opportunity to reflect on how these new approaches can complement and enrich an established hermeneutical practice. Nevertheless, we must admit that we often faced technical and pragmatic limitations, solving which depended on student and teacher preparation and skill, as well as encountering various other more logistical and infrastructural issues, such as hardware access. For example, can we choose a software that is not free or open source because our university provides us with a license, or should we opt for open versions instead of proprietary ones for more fundamental reasons of openness?

For our exercise, we picked Microsoft Excel as a useful tool and format for data sharing since the University of Mainz provides a license to all students. Open Refine is a great tool for data cleaning and enriching. It enables the so-called "reconciliation" of a dataset through an API with Wikidata (or other platforms). In our case, the data enrichment process was particularly profitable: through a list of film titles, we could enrich each entry with additional information about the film, such as director, cast, country of origin, production company, and unique identifiers from IMDb or OMDB.[27] The software, which was originally called Google Refine and published in 2010 by Google, is now open source. Open Refine can clean datasets, "facet"[28] many additional features, cluster data entries and create consistency, and complete datasets by fetching geocoded information. In class, we chiefly used it for the task of reconciling data through Wikidata, an operation that would be more or less easy to automatize depending on the dataset. We gave students different guided assignments to work on with Open Refine to become acquainted with the process of reconciliation. Open Refine also carries with it the possibility of using regular expressions to manipulate data more directly with its own language called GREL (General Refine Expression Language)[29] or Python. Yet the intuitive interface allows beginners to complete simple tasks with the tool without difficulties.

Of great importance of course are the issues of structuring and modeling data. These essential decisions need to be decided beforehand by choosing fields and data entries and, when working with Open Refine, what additional informa-

27 See https://www.omdbapi.com/, accessed February 15, 2023.

28 See https://openrefine.org/docs/manual/facets, accessed February 15, 2023.

29 See "General Refine Expression Language," Open Refine, November 24, 2022, accessed July 21, 2023, https://openrefine.org/docs/manual/grel.

tion to reconcile. If structuring data within a spreadsheet implies selecting different columns and classifying sources with consistency, this sometimes forces us to take decisions on what to discard. For instance, to make our dataset easily legible using the visualization tools, it is better to select one specific genre for each film entry; this also means discarding other genres' definition. A film is often not just a documentary film or a drama, but may also be a comedy, a western, or a horror movie at the same time. Each operation of data modeling and data cleaning should be performed with both critical and pragmatic reflection, following our research question but also taking into account the potentiality of the tools and our own skills.

As the field of feminist studies has shown, in a data project what is left out or what we choose to measure is also a "measure of who [or what] we value."[30] From this feminist perspective, the performance of cleaning data is also a way of disciplining it. Although this reflection is fundamental also from a Film Studies perspective, sometimes choices need to be made to balance one's own skills and software's affordances.

Visualizing data requires us to make the same critical reflections: as Johanna Drucker argues, to visualize is always an act of interpretation.[31] This means that what a visualization tool does is itself an act of interpretation, and how or what a software allows you to visualize has implications for the production of knowledge. The tools we chose for our exercise presented various limits: these stemmed on one hand from individual students' inexperience with data visualization and basic statistics and, on the other, from the technical limitations of the tool itself.

Tableau Public,[32] the free version of the business intelligence software Tableau, combines pleasing visualizations with a powerful engine for data analysis and interpretation within the same interface. In the first year we chose other tools for visualization: the open-source browser-based software RawGraphs[33] or Excel, to create basic graphs and charts with a dataset. These options were either more difficult to handle or less aesthetically pleasing than Tableau Public. We also offered the possibility of using some digital storytelling tools, such as TimelineJS,[34] StorymapsJS,[35] and TimeMapper.[36] These browser-based tools were developed by the

30 Catherine D'Ignazio and Lauren F. Klein, *Data Feminism*, Strong Ideas Series (Cambridge, MA: MIT Press, 2020).

31 Johanna Drucker, *Visualization and Interpretation: Humanistic Approaches to Display* (Cambridge, MA: MIT Press, 2020).

32 See https://public.tableau.com/, accessed June 24, 2021.

33 See https://rawgraphs.io/, accessed July 19, 2021.

34 See https://timeline.knightlab.com/, accessed May 27, 2021.

35 See https://storymap.knightlab.com/, accessed May 27, 2021.

36 See https://timemapper.okfnlabs.org/, accessed May 31, 2021.

Open Knowledge Foundation Labs and Knightlab from Northwestern University and enable the relatively painless generation of maps or timelines. While these tools were used by some students for their project with some excellent outcomes,[37] others found them outdated because of their older-looking interface design. These tools also limit the insertion of references, so it seems they are, alone, insufficient for rigorous scholarly work.

Tableau Public, despite the limits of its free version,[38] was well received by students and chosen by the majority for their own projects (we also presented Datawrapper,[39] but no students used it in their final projects). In class, we also developed assignments to acquaint students with data visualizations and to explore datasets with other tools, such as Google NgramViewer,[40] Project Arclight,[41] and Wikipedia's Pageview Analysis.[42] These tools offer direct and simple access to visualize and compare data and the possibility of looking into the original material of the dataset.

The use of data visualizations as an entry point into a larger dataset as a way to explore the data, has been discussed by Alberto Cairo in his influential book about data visualization, *The Truthful Art*.[43] Cairo borrows from John W. Tuckey the practice of "explorative analysis of data,"[44] and argues that visualizing the dataset first offers us a way to better understand patterns and trends, which are often not explicit in a dataset, and thus to pose the right questions. In our exercise, we realized, in supervising students' projects, that their visualizations and analysis represented a way of getting to know the dataset better and thus raise further and broader questions.

In the case of a dataset on Super 8 reduction prints, some projects highlighted patterns in the genre distribution and the country of origin of films and, at the

37 One student, for instance, tracked the history of Super 8 film format through a timeline and also adjusted the design by using the HTML mark-up language.

38 There are many limits in the free version, such as in the range of tools for cleaning the data, but also on the possibility of choosing layouts for the visualization. In the last course, we encountered technical difficulties we were not able to solve, such as publishing a visualization, in order to download it as an image format.

39 See "Enrich your Stories with Charts, Maps, and Tables," Datawrapper, accessed December 2, 2021, https://www.datawrapper.de.

40 See https://books.google.com/ngrams/, accessed July 21, 2023.

41 See http://search.projectarclight.org/, accessed April 29, 2021.

42 See https://pageviews.wmcloud.org/?project=en.wikipedia.org&platform=all-access&agent=user&redirects=0&range=latest-20&pages=Cat|Dog, accessed February 8, 2023.

43 Alberto Cairo, *The Truthful Art: Data, Charts, and Maps for Communication* (Indianapolis: New Riders, 2016).

44 John Tukey, *Exploratory Data Analysis*, 1st ed. (Reading, MA: Pearson, 1977).

same time, those visualizations offered an overview of the relationship between specific national productions and a genre. This often unclear connection made us question which kinds of film productions the graphics would describe, and thus raised an interest into investigating single film entries in the list more directly. Visualization was thus very productive for knowing our dataset better and suggesting previously unknown connections and relationships between the data. Nevertheless, it must be combined with a close reading of the sources themselves to produce a detailed hermeneutical analysis. This insight shows how the use of digital tools is complementary to other film-historical approaches.

New Resources for Studying Small Gauge Film Formats

This complementarity between tools and hermeneutical methods is particularly evident in the study of small gauges. Small-gauge formats are often presented in archives as negative copies of films that were originally printed in 35mm; other examples of these substandard formats were used for educational films, industrial films, home movies, family films, amateur film practices, or experimental films. The information age has produced new forms of accessibility to these formats through digital repositories and online archives, offering an efficient alternative to in-person visits to archives. Projects such as the Amateur Movie Database[45] or the Amateur Film Platform[46] offer convenient access to amateur films and small-gauge formats.

In the first year of our digital methods course, students were allowed to choose their own research topic, but we also suggested a list of resources. These facilitated access to either datasets and lists of films directly or a repository of film and material about film, such as the MHDL.

In the second year of teaching, we decided to focus directly on Super 8, since we had previously begun gathering catalogs of reduction prints that had circulated in Germany in the 1970s and early 1980s. We discovered an extensive non-institutional database – the so-called Super 8 Database,[47] which is a freely accessible website that collects and publishes data on Super 8 reduction prints distributed by a series of international companies. The project was started in 2018 by Eivind Mork, a Norwegian

45 See https://www.amateurcinema.org/index.php/amdb, accessed February 9, 2023.
46 See https://www.amateurfilmplatform.nl/, accessed February 9, 2023.
47 See https://super8database.com, accessed May 2, 2022.

computer engineer and amateur collector, who keeps the database updated (it now counts up to 8,360 film entries) and provides the datasets freely in different file format extensions, such as .csv, .xlsx, and .ods. It is an outstanding resource, which also links some of the films directly with unique identifiers to IMDb and shows the source catalogs and the film's original covers.

We decided to use this resource with students to provide new insights into this Super 8 collection from a Film Studies perspective. Many films in Super 8 were reduction prints, i.e., only extracts or cut versions of longer films. The film's database entry would therefore frequently have a title not directly ascribable to its "original" movie. Depending also on the country of distribution, the same film might have different versions (in different lengths, in color or black and white, with or without sound) and different titles in each country, further complicating the process of identifying the contents of the database. This represented a challenge in cleaning the data because cleaning could not be reduced to simple operations of making data uniform and formatting the entries following a standard. In Film Studies, especially when working on historical material, as Johanna Drucker has observed, "the trade-off between preserving the specificity of source information and standardizing information for use is always fraught with ethical questions."[48] Nevertheless, this challenge offered a way to make students filmic expertise fruitful.

The Super 8 Database and the other sources mentioned above are some of many institutional and non-institutional sources, which we may find navigating online. On Wikipedia, for instance, we may find additional lists; these and other websites can provide a starting point for further research. This new type of source, especially the non–institutional ones, requires a new way of research, which we want to introduce next.

A Web Philology Workflow for Handling Super 8 Data

As we mentioned, working with Super 8 data[49] presented a challenge in finding the "original" theatrical film release from which the reduction prints had been copied. This operation was further complicated by the fact that we did not work

48 Drucker, *The Digital Humanities Coursebook*, 31.

49 We chiefly refer here to data as an umbrella term for metadata; we actually worked mostly on metadata on film, starting from titles and other general information about the Super 8 film copies.

with digitized or printed film versions, but only with film titles and metadata coming from film catalogs and turned into spreadsheets. To date a Super 8 reduction print title to its theatrical film release (or television release)[50] was of great importance if we wanted to work with Open Refine and reconcile our data entries through Wikidata. For that reason, we developed a workflow and taught it to students.

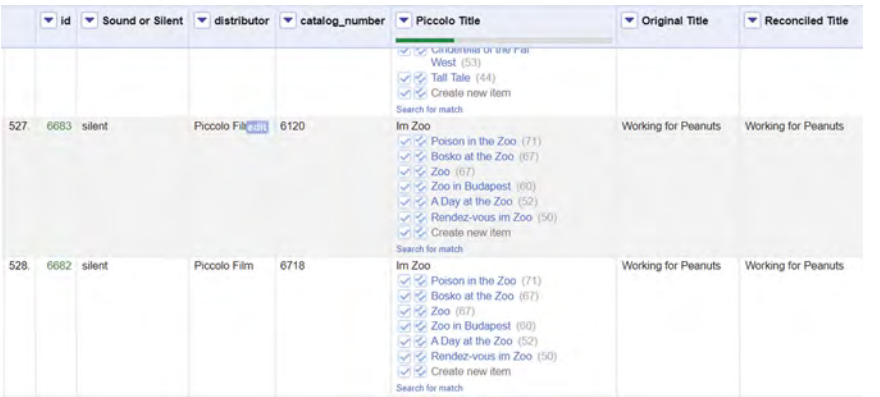


Figure 1: Some film entries in Open Refine from a dataset of Super 8 films distributed by Piccolo Film. Source: screenshot by the authors.

Figure 1, which is a screenshot from the interface of Open Refine, shows two different data entries, representing two films with different serial numbers, but sharing the same title "Im Zoo." Open Refine, unable to reconcile the title to a direct match, provides various suggestions in the final column. The title, in any case, is very vague; one cannot recognize any familiar film title. The serial number and distributor can, however, help researchers extract additional information from the original catalog source behind the Super 8 Database entry.

Figure 2 shows the original source (in the Super 8 Database), a scan of a print catalog of Disney films. Thus, it became clear to us that the film entitled "Im Zoo" was possibly an extract (in two short versions) of a Disney animation, but the Super 8 Database did not contain more specific information.

To solve the problem there were other useful sources: IMDb, and Ebay. Ebay is a great platform to look for Super 8 reduction prints, because collectors are still

50 While working on the catalogs, we realized that the Super 8 libraries were not only reduction prints of theatrical film releases but also of television films and series or specific televisual events, such as sport competitions.

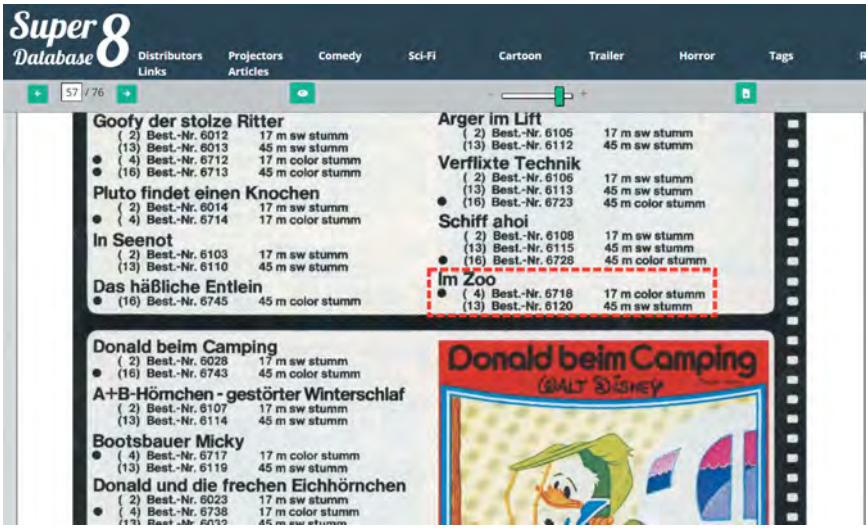


Figure 2: On the Super 8 database some titles are connected to a catalogue. Source: screenshot by the authors.

selling and buying Super 8 films over the platform, often uploading photos and details about the reduction prints. If we searched for the title "Im Zoo" as Super 8 by Piccolo Film, we might find the original cover by Piccolo Film. This told us that the film is an extract of a short animation involving Donald Duck and Chip 'n Dale. By searching further on a possible film with Chip 'n Dale at the Zoo we found a wiki page about the short film Working with Peanuts from 1953. A quick search on IMDb gave us more information and the possibility of adding a unique identifier to our list. By searching directly for a match in Open Refine, we could also reconcile the Super 8 title version with the original film. This allowed us to link our entry with Wikidata and gather further information: director, country of origin, date of first release, etc.

The use of such commercial resources demonstrates the infrastructural problem of working outside the institutional realm and the necessity of making our datasets more interoperable and accessible for further research. Moreover, the possibilities offered by Wikidata emphasize the importance of Linked Open Data and the necessity of adding new layers of data even for small-gauge films.

Of course, this process for identifying and enriching the content of the Super 8 database might vary depending on each entry's genre, distributor, and country of origin. Some of these Super 8 companies would sell fuller versions and thus use the original titles, others would rename or translate their different reduction prints; in the latter case, finding the right match would not be possible without watching the

material directly. In any case, this workflow shows that, in the case of small gauges and especially with reduction prints, automated reconciliation is not always possible, but the features offered by applications like Open Refine can be complemented with web philological workflows and thus with manual research.

This research practice and the issues of reconciliation highlight a broader question about a film's so-called original version. Of course, pragmatically speaking, these Super 8 distributors had to buy the rights of specific titles from production companies, such as Disney or Universal, before selling their reduction prints. Although we may be able to speak of an original copy in this case, the question nevertheless remains whether Super 8's different versions represent a completely different film or merely a manifestation[51] of a previous film. Sometimes they were just extracts, sometimes full versions, but often they were edited for the Super 8 distribution only and that would alter the original film text. Film scholars, then, must determine whether they are still talking about the same "work." This is an interesting outcome of our work with Super 8 prints; further research is needed to compare Super 8 versions with their original theatrical copies, the distribution rights into which Super 8 companies bought in the first place.

The developed workflow also allowed us further insights into these films by revealing other information and other resources, mostly driven by amateur collectors. Another example is a resource about the Super 8 versions of Asterix and Obelix. In the Piccolo Film dataset, there are many different titles coming from different films. Searching online we found a resource called the "Deutsches Asterix Archiv"[52] which has information about many different media of Asterix, and also on all Super 8 versions circulated in German and French by different distributors. They explicitly list the single Super 8 reduction prints by original film. These resources are of course non-institutional ones and cannot be taken as archival information curated by experts. We may talk about these informal archives as a form of "rogue archives,"[53] which are opposing the institutional archives, and represent a form of archiving from below. Nevertheless, because Super 8 reduction prints are often missing from academic film archives, these amateur and fan sources are a great starting point for our further scholarship with Super 8 and other small gauges.

---

51 Natasha Fairbairn, Maria Assunta Pimpinelli, and Thelma Ross, *The FIAF Moving Image Cataloguing Manual* (Brussels: Fédération internationale des archives du film, 2016). For a discussion or difference between work and manifestation, we refer to the FIAF manual.

52 See "Asterix Archiv – Bibliothek – Filme und Hörspiele," Comedix, accessed February 10, 2023, https://www.comedix.de/medien/lit/asterix_super_8.php.

53 Abigail De Kosnik, *Rogue Archives: Digital Cultural Memory and Media Fandom* (Cambridge, MA: MIT Press, 2016).

Food for Thought: Challenges and Lessons Learned

To conclude, we want to share the lessons we learned. In this section, we name our experiences and explain how we dealt with them. The course was developed as part of a historical tutorial module within the Film Studies Bachelor, which posed a first challenge: how to sensibly spread the course over a weekly two-hour time slot. We planned 14 sessions with a duration of 90 minutes each. Taking our hands-on approach into account, we established the following schedule: two introductory sessions (general introduction to the topic and expert lecture); sessions on deepening the topic of small gauge; an introduction to digital methods and working with different tools, source and data criticism; sessions for group work on students' projects and presentations of preliminary results; finally, two sessions with final presentations and wrap-up.

We split the course into three parts: input, practice and output (following the principle of material + processing + presentation mentioned previously). In our experience, this split has been successful, although the weighting of the three parts may need to vary depending on the previous competence of the students.

Another important challenge we faced was a general lack of interest in understanding Digital Humanities (DH) concepts. Teaching students at the beginning of their BA, we consistently observed that their interest in such concepts seemed to be rather weak. We see two main issues. First, we did not take the time to explain DH in depth but gave only an overview[54] so that the students could classify the concepts and methods introduced and used. Second, it may simply have been too overwhelming because students were at the beginning of their second year of study, which meant that they had only just started acquiring basic knowledge in Film Studies and were not at all proficient in digital methodologies in order to understand and be interested in a theoretical approach in this field of interdisciplinary research.

The strong interdisciplinarity was not only difficult for the students to grasp, but also challenged the way we shaped and developed the course. We had the opportunity to invite experts from different fields, and we directly experienced co-teaching to test a fruitful collaboration between Film Studies and DH. By combining our different fields of expertise, we taught students the value of this interdisciplinarity. We had to explore new methods or apply familiar methods to new

54 In the second iteration of the course, we assigned the students a recorded talk from Manuel Burghardt: Burghardt, "Big Tent Digital Humanities und die Entwicklung der Computational Humanities."

research areas, and we could not assume that students were already familiar with digital methods. One strategy to deal with those circumstances was to openly communicate that we were developing new teaching concepts. We tried to raise awareness among students in the course that they were learning new methods and ways of working, and that small-gauge formats were as much a vehicle for learning new methods as they were material to be understood in its own right. From the beginning, we shared an open and critical perspective over the tools. We invited students to address difficulties in applying the new methods to Film Studies and also discussed with the students the limitations, for instance, the challenges with visualization when they started to apply them with Tableau (e.g., hardware limits, software limits and bugs, datasets, and data types). Although we successfully anticipated many problems, students often raised new and unforeseen issues, forcing us to challenge our own understanding of the tools and their limits.

Encouraging criticism of tools and data sources was one of the most important aspects to keep in mind in our course. Although we read different sources[55] to underline the importance of this aspect, it seemed difficult for the students to apply them to their projects. In response to the feedback of our students, who requested additional course time focused on understanding datasets, we hope in future iterations of this course to introduce the practice of "reading datasets" in a way similar to other academic texts. This process is inspired by text comprehension, which is a common practice in Film Studies courses. Just as students are often asked to write a series of questions about a text and its structure, we would hope to develop parallel analytical competencies around datasets.

As our course syllabus continues to develop, we would also like to thematize the future value and significance of digital methods for the students of Film Studies. Although we do not want to diminish the importance of hermeneutical approaches, we want to give the students a starting point for working with digital tools and to support data literacy. Having taken this course, they might in the near future even deepen and put in practice these newly learned techniques in other coursework or research.[56]

We are in a privileged situation of co-teaching, permitting us to rely on our respective fields of expertise and allow each other's strengths to fill in our own

55 Christof Schöch, "Big? Smart? Clean? Messy? Data in the Humanities," *Journal of Digital Humanities* 2, no. 3 (2013), accessed July 21, 2023, https://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/; Drucker, *The Digital Humanities Coursebook*.

56 At the present moment, however, ours is the only course in our department with a distinct focus on combining Film Studies and digital methods, so we have to observe whether there are individual students who show interest in pursuing this approach themselves, as long as we are there to support them.

weaknesses. Unfortunately, the combination of these skill sets is not yet common in academia – both in Digital Humanities and traditional Film Studies. Our experience with Dici-Hub indicates that the majority working in this interdisciplinary field are self-made digital film scholars, mainly with a background in Film Studies. We must acknowledge that, although we hope that we are supporting the aptitude of a next generation of students for using digital methods, we expect administrators to experience difficulty in finding personnel with the necessary interdisciplinary skills who are willing or able to stay in the precarious conditions of academic work. This difficulty seems likely to persist over the midterm.

Despite the difficulties encountered, our experience generated helpful advice for teachers interested in delving into this field. While there were many challenges to be faced, we not only learned strategies to solve concrete problems but, more importantly, we learned some lessons that we treasure. Here are a few solving strategies we started to develop:

– *Encouraging the students in working with digital methods and tools (working with digital natives)*: One of the main aspects we learned from our first course was a feedback routine, providing recurrent moments for discussing and getting feedback on the project. We implemented this within the phase dedicated to presentation and discussion of preliminary results. On the one hand, this meant we could monitor the progress of the groups; on the other hand, the students could also give their feedback and discuss recurring problems in the areas of the research questions and the digital tools. One of the main aspects was also talking about failure. This helped to deepen the understanding of workflows with the tools and minimize the qualms about "crashing" the computer or software. Thus, in the end, we had a routine of explaining and demonstrating the tools, deploying supervised practice with given datasets and free but supervised work with chosen datasets. Due to the feedback routine, the students also became teachers of their peers when they presented their work and troubleshooting strategies.
– *Team & hands-on work (working interdisciplinary)*: In order to strengthen collaborative work, it is best that students form groups at the beginning of the course and work together until the end of the semester. Early formation of groups has the social purpose that the students get to know each other, since not all of them would previously have been acquainted, since they were at the beginning of their studies. Early group formation also helps students assess the strengths and weaknesses of the group at an early stage, so that they can then be more effective in working in teams. We prefer the in-presence format over hybrid learning because it encourages the students to ask more questions when they occur and they can see that we often answer questions jointly. This strengthens the character of interdisciplinary workflows and

gives insight into at least two disciplines. They can also benefit from the individual prior knowledge in their groups, due to their status as "digital natives" and the fact that their own foci in studies are highly heterogeneous. Although we had already increased the amount of hands-on sessions, further iterations of this course may benefit from even more emphasis on practice. Ideally, this could in turn lead towards the use of certain methods or tools beyond the course. This would also include a re-weighting between conventional lecture assignments and hands-on-practices, while maintaining the workload planned for the course. As we learned, hands-on work experience is the ideal starting point also for a critical approach to the tools and sources and reinforces data and tool criticism. This is especially important because we work with tools and methods that were not originally developed for Digital Film Studies but which we try to adapt. This leads to our third point:

– *Understanding the challenges of Digital Film Studies (working in Digital Film Studies)*: The most difficult challenges lie in the selection of a dataset and the tools to work on the data. Therefore, we reduced the methods and tools taught and limited the choice of datasets. In this way, it was possible to guide and support the students in a better way and provide the possibility of a successful first encounter with digital methods and tools. Furthermore, we needed to develop strategies for a deeper data literacy when working with datasets. This cannot only include source criticism. The students need to learn to read datasets as texts: because this course uses particular datasets as its main sources, it is expedient for students to gain exposure to them, and begin working on their final projects from the outset in the course. We should also teach students to understand the structure of datasets, how their data was modeled, and how to deal with specific problems in the data entries. This also includes the experience that not everything can be automated. We need to raise awareness that, before using a tool, one needs a clear idea of the desired results in order to properly design and target an efficient-as-possible workflow. This "clear idea" is a negotiation between knowledge of the tools and experience, so students' research directions, and thus their workflows, may be obvious from the beginning, or may evolve over time, making patience an important virtue to be cultivated.

Our experience thus far, we hope, might offer a way to make data productive for digital film and media studies. Our specific approach to small-gauge formats, and in particular to Super 8 reduction prints, demonstrates how implementing digital tools often implies a negotiation between film-historical knowledge and material and digital skills. Using digital methods creates a space for students to learn how to collaborate towards a goal, and also to learn a different approach for doing

research, not based only on reading and discussing texts, but also on researching, exploring, modeling, and visualizing data. If the online realm presents today's scholars with new opportunities for research, it is fundamental that we continue to keep our critical and hermeneutical approach and learn how to apply it to digital sources and digital tools. Data literacy and digital literacy are not only essential for studying small gauges with digital methods, but as a toolkit for integrating digital methods into Film Studies.

References

Aasman, Susan. "Unlocking Multiple Histories of Amateur Media: From Micro- to Macro-Histories." *Screen* 61, no. 1 (2020): 157–166. https://doi.org/10.1093/screen/hjaa011.

Arnold, Taylor, Stefania Scagliola, Lauren Tilton, and Jasmijn Van Gorp. "Introduction: Special Issue on AudioVisual Data in DH." *Digital Humanities Quarterly* 15, no. 1 (2021). Accessed July 21, 2023. https://www.digitalhumanities.org/dhq/vol/15/1/000541/000541.html.

Braida, Nicole, Isadora Campregher Paiva, and Josephine Diecke. "Was machen wir mit digitalen Tools und was machen sie mit uns?" *Open Media Studies Blog*, February 2, 2022. Accessed July 21, 2023, https://zfmedienwissenschaft.de/online/open-media-studies-blog/was-machen-wir-mit-digitalen-tools-und-was-machen-sie-mit-uns.

Burghardt, Manuel. "Big Tent Digital Humanities und die Entwicklung der computational Humanities. Aktuelle Trends und Fallstudien für Computergestützte Verfahren in den Geistes- und Kulturwissenschaften." Presentation at the Interdisziplinäre Hybrid-Vorlesungsreihe zum Themenschwerpunkt "Digital Humanities" – Wintersemester 2020/2021, Johannes Gutenberg-Universität Mainz, December 10, 2020. Accessed February 15, 2023. https://video.uni-mainz.de/Panopto/Pages/Viewer.aspx?id=e5f175a4-eb21-4caa-8ca7-ac8c0108cd24.

Cairo, Alberto. *The Truthful Art: Data, Charts, and Maps for Communication*. Indianapolis: New Riders, 2016.

Catanese, Rossella, Adelheid Heftberger, and Christian Gosvig Olesen. "Computer-Based Approaches to Film Archiving, Restoration and Philology." *Cinergie – Il Cinema e Le Altre Arti* 20 (2021): 1–6. https://doi.org/10.6092/ISSN.2280-9481/13948.

De Kosnik, Abigail. *Rogue Archives: Digital Cultural Memory and Media Fandom*. Cambridge, MA: MIT Press, 2016.

D'Ignazio, Catherine, and Lauren F. Klein. *Data Feminism*. Strong Ideas Series. Cambridge, MA: MIT Press, 2020.

Drucker, Johanna. *The Digital Humanities Coursebook: An Introduction to Digital Methods for Research and Scholarship*, 1st ed. Abingdon, Oxon; New York: Routledge, 2021.

Drucker, Johanna. *Visualization and Interpretation: Humanistic Approaches to Display*. Cambridge, MA: MIT Press, 2020.

Eisloffel, Paul. "A Brief History of the 16mm Film Format." *MAC Newsletter* 41, no. 1, Article 7 (2013): 32–33. Accessed July 21, 2023. https://lib.dr.iastate.edu/macnewsletter/vol41/iss1/7.

Fairbairn, Natasha, Maria Assunta Pimpinelli, and Thelma Ross. *The FIAF Moving Image Cataloguing Manual*, edited by Linda Tadic. Brussels: Fédération internationale des archives du film, 2016.

Goergen, Jeanpaul. “Filmgeschichte im Wohnzimmer. 16mm- und Super8-Kurzfassungen deutscher Filmklassiker.” *Filmblatt* 7, nos. 19–20 (2002). Accessed July 21, 2023, https://www.filmblatt.de/2002/10/01/filmblatt-7-jg-nr-19-20-winter-fruehjahr-2002/.

Hediger, Vinzenz. “‘Dann sind Bilder also nichts!’. Vorüberlegungen zur Konstitution des Forschungsfelds ‘Gebrauchsfilm’.” *Montage/AV* 14, no. 2 (2005): 11–22. Accessed July 21, 2023. https://doi.org/10.25969/MEDIAREP/.

ICDL Europe. “The Fallacy of the ‘Digital Native’.” Accessed February 13, 2023. https://www.icdleurope.org/policy-and-publications/the-fallacy-of-the-digital-native/.

Jancovic, Marek, Axel Volmar, and Alexandra Schneider, eds. *Format Matters: Standards, Practices, and Politics in Media Cultures*. Lüneburg: meson press, 2019. https://doi.org/10.14619/1556.

Livingstone, Sonia. “Enabling Media Literacy for ‘Digital Natives’ – A Contradiction in Terms?” In *“Digital Natives”: a Myth?*, 4–6. London: POLIS, London School of Economics and Political Science, 2009. Accessed July 21, 2023. https://eprints.lse.ac.uk/48944/.

MacGillivray, Scott. *Castle Films: A Hobbyist’s Guide*. Bloomington, IN: iUniverse, 2004.

Moretti, Franco. “Planet Hollywood.” *New Left Review* 9 (2001): 90–101.

Noordegraaf, Julia. “Computational Research in Media Studies: Methodological Implications.” *Tijdschrift Kwalon* 21, no. 1 (2018): 52–59.

Olesen, Christian Gosvig, Eef Masson, Jasmijn Van Gorp, Giovanna Fossati, and Julia Noordegraaf. “Data-Driven Research for Film History: Exploring the Jean Desmet Collection.” *The Moving Image* 16, no. 1 (2016): 82–105.

Prensky, Marc. “Digital Natives, Digital Immigrants Part 2: Do They Really Think Differently?” *On the Horizon* 9, no. 6 (2001): 1–6. https://doi.org/10.1108/10748120110424843.

Schneider, Alexandra. “Theorie des Amateur- und Gebrauchsfilms.” In *Handbuch Filmtheorie*, edited by Bernhard Gross and Thomas Morsch, 225–242. Wiesbaden: Springer, 2021.

Schneider, Alexandra. “Viewer’s Digest: Small-Gauge and Reduction Prints as Liminal Compression Formats.” In *Format Matters: Standards, Practices, and Politics in Media Cultures*, edited by Marek Jancovic, Axel Volmar, and Alexandra Schneider, 129–146. Lüneburg: meson press, 2019. https://doi.org/10.14619/1556.

Schöch, Christof. “Big? Smart? Clean? Messy? Data in the Humanities.” *Journal of Digital Humanities* 2, no. 3 (2013). Accessed July 21, 2023. https://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/.

Streible, Dan, Martina Roepke, and Anke Mebold. “Introduction: Nontheatrical Film.” *Film History* 19, no. 4 (2007): 339–343.

Teubig, Michael. “Was ist aus den Super-8-Vertrieben geworden?” off2 – Filme auf Super 8, 1987. Accessed July 21, 2023. https://off2.de/publikationen/die-leinwand/wo-sind-die-super-8-firmen-l2-87/.

Thompson Klein, Julie. “Introduction: Emerging.” In *Interdisciplining Digital Humanities: Boundary Work in an Emerging Field*, 1–13. Ann Arbor: University of Michigan Press, 2015. Accessed July 21, 2023. http://www.jstor.org/stable/j.ctv65swxd.5.

Tukey, John. *Exploratory Data Analysis*, 1st ed. Reading, MA: Pearson, 1977.

Verhoeven, Deb. “Visualising Data in Digital Cinema Studies: More than Just Going Through the Motions?” *Alphaville. Journal of Film and Screen Media* 11 (2016): 92–104.

II Data

Sarah-Mai Dang

Managing the Past: Research Data and Film History

Introduction

Given the growing number of data-driven research projects, research data management (RDM) is becoming increasingly important across all disciplines. As I have elaborated elsewhere, it has also become a relevant field in film and media studies, a field which is very likely to grow in the near future.[1] RDM describes the practices of collecting, selecting, modeling, organizing, preserving, and sharing of data in order to facilitate their access and reuse. This involves a sequence of many different steps and phases that may be repeated and vary depending on the research project and discipline. Ideally, RDM involves a management plan that outlines these steps in a research project including data sources, formats, documentation, storage, and access. Today, there are numerous tutorials and guidelines that describe the complex workflows, commonly defined and explained as data lifecycles.[2]

However, even though there are now many initiatives that seek to address the specific humanities requirements of RDM,[3] there is nevertheless much to explore and understand about particular disciplinary challenges. While there are numerous resources that provide general advice, when it comes to specific pro-

1 Sarah-Mai Dang, "Forschungsdatenmanagement in der Filmwissenschaft. Daten, Praktiken und Erkenntnisprozesse," *montage AV* 29, no. 1 (2020): 119–140.

2 For example, for the German-speaking community the website forschungsdaten.info has become a central reference: Forschungsdaten.info, "English Pages. Forschungsdaten und Forschungsdatenmanagement," accessed February 3, 2023, https://forschungsdaten.info/english-pages/. In the Netherlands, the idea of "Data Scopes" (https://data-scopes.github.io/Data-Scopes/) has been discussed as part of broader critical tool discussions. Data Scopes explores how research data may (or may not) be linked and shared to ensure greater transparency and to understand how data shape research. Rik Hoekstra and Marijn Koolen, "Data Scopes for Digital History Research," *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 52, no. 2 (2019): 79–94. https://doi.org/10.1080/01615440.2018.1484676. I thank Christian Gosvig Olesen for pointing out the "Data Scopes" project in his feedback on an earlier version of this chapter.

3 For instance, specially established digital research centers at universities now offer comprehensive counseling on research data management. Additionally, the recently launched National Research Data Infrastructures (NFDIs) in Germany aim to identify central questions in this regard across disciplines, promote cross-project exchange, and provide concrete assistance in the event of problems.

https://doi.org/10.1515/9783111082486-007

ject issues such as terms of use for film historical databases or subject-specific storage options, there are still few examples of best practice in film and media studies. Thus, until now learning by doing has been the most effective approach. In addition to practical and legal challenges that must be addressed, we also need to pay more attention to the cultural and political implications of RDM. Until recently, RDM has often been treated as a purely organizational or technical side task, required by funding organizations or the respective institution in terms of reusability and good scholarly practice. Humanities scholars tend to view it as a nuisance because it is not considered to be part of humanities research.[4] Yet, as more and more scholars are becoming aware, the way RDM is conducted deeply affects our understanding of film culture as well as the scholarship that explores it. For example, what credits we retrieve or include in a database to identify or describe a film does not only determine how we can search for specific criteria, but also reflects what we consider relevant for analyzing film culture. If we list only the title of a film, the year and country of production, and the director, as is common in scholarly works, we cannot analyze the involvement of other professions such as the editor or screenwriter. This demonstrates that ascribing specific metadata to artifacts is not a neutral procedure, but an act of interpretation based on theoretical premises and cultural assumptions. In this respect, interpretation is part of all RDM practices and does not only take place afterwards.[5] Thus, different from what one might associate RDM with, it is neither pure technical or organizational, nor universally valid. It is shaped by intellectual conventions and institutional frameworks, and vice versa; it shapes our concepts and ideas of what we perceive as film or authorship.[6] We should not therefore submit a data management plan just to meet formal funding requirements or institutional practices, but, on the contrary, use it to reflect and help shape research.

In this chapter, I reflect on RDM and its impact on digital scholarship regarding film history. Rather than providing a guideline for implementing RDM in film and media studies, I want to draw attention to the theoretical and political implications of RDM that can serve to conduct more informed data practices. In doing

4 Sophie G. Einwächter, "Forschungsdaten (in) der Film- und Medienwissenschaft – Sophie G. Einwächter über vorurteilsbehaftete Begrifflichkeiten und fruchtbare Momente in der Lehre," *Open Media Studies Blog* (2019), accessed March 21, 2023, https://mediastudies.hypotheses.org/1314.

5 Lisa Gitelman and Virginia Jackson, "Introduction," in *"Raw Data" Is an Oxymoron*, ed. Lisa Gitelman (Cambridge, MA: MIT Press, 2013), 3.

6 See also Julia Noordegraaf, Kathleen Lotze, and Jaap Boter, "Writing Cinema Histories with Digital Databases: The Case of Cinema Context," *TMG Journal for Media History* 21, no. 2 (2018): 106–126, https://doi.org/10.18146/2213-7653.2018.369.

this, I will discuss questions such as: How is the data organized? By whom and for what purpose? What information has been extracted? Which sources have been ignored? How do current premises and ideas inform digital curation practices? How does RDM shape our understanding of film culture?

While data is not only numbers but can also be texts, images, audio, or video,[7] this chapter focuses on film historical metadata. Drawing on critical approaches to data-driven projects and infrastructures, as well as my own experience with film historical databases, I seek to outline a framework that allows us to systematically scrutinize RDM. To this end, I will compare two digital data initiatives: the Women Film Pioneers Project (WFPP) and filmportal.de of the DFF – Deutsches Filminstitut & Filmmuseum. Given their different designs and objectives, a comparison promises to provide interesting insights into the particularities of RDM and thereby emphasize its relevance. Both initiatives are project partners of my current research group "DAVIF – Visualizing Research Data on Women in Film History" (2021–2025). They provide the research data for exploratory data visualizations in addition to their curatorial knowledge and archival expertise.[8] This means I have not been involved in the production and curation of the data in any of the projects. Instead, I look at the data as a humanities scholar in retrospect to understand the objects at the heart of my research.

While the initial aim of DAVIF was to make research data on women in early cinema more visible by the means of data visualizations, analyzing how the research was produced in the form of data has unexpectedly become a significant aspect of my research. Conducting the first case study, it soon became clear that, in order to make sense of the data visualizations we created in our project, I needed to understand the underlying source material as well. My focus shifted from exploring the presentation and reuse of research data to critically scrutinizing the digital source. As emphasized by scholars in the digital humanities, when working with data, the research process becomes a fluent, interconnected series of different steps that depart from the more linear structure of traditional methods.[9] During the course of my study, it became apparent that data visualizations not only enhance existing research, but can also provide valuable insights into data and serve

7 Dang, "Forschungsdatenmanagement in der Filmwissenschaft," 121–122.

8 The project is funded by the German Federal Ministry of Education and Research (BMBF) for four years (2021–2025), https://www.uni-marburg.de/en/fb09/institutes/media-studies/research/research-projects/davif.

9 Andreas Fickers, Juliane Tatarinov, and Tim van der Heijden, "Digital History and Hermeneutics – Between Theory and Practice: An Introduction," in *Digital History and Hermeneutics: Between Theory and Practice*, ed. Andreas Fickers and Juliane Tatarinov (Berlin, Boston: De Gruyter Oldenbourg, 2022), 9–10.

as a method for the analysis of the data corpus. To better understand the underlying premises and practices, we also conducted semi-structured interviews with individuals involved in the management of the databases. Through these interviews, I was able to understand essential features of the sources and workflows that generated long-standing initiatives.[10]

Analyzing Film Historical Metadata

Although the data lifecycle model does not imply that RDM is purely technical, but is intended to simplify the complexities involved, I approach the topic from a different angle in order to emphasize the entangled practices, assumptions, and external factors that one might not necessarily associate with what is commonly referred to as "management." Based on the interviews with our project partners, the data visualizations conducted in our research group, and other critical approaches, I have identified three focus areas as a starting point for further examining RDM practices and their particularities in order to understand film historical metadata: (1) content and context; (2) data modeling and categorizations; and (3) access and reuse.

1 Content and Context

What is the subject area of the database? What time period does the database cover? What are the geographical foci? What is the primary goal? Where is the database affiliated? Who is involved in the work? Where does the funding come from?

RDM varies depending on the goal of an initiative, where it takes place, and the people involved. If databases are ideally preserved beyond a project's lifetime, they become part of a digital research infrastructure. As such, they are often perceived as neutral and far removed from personal influence, although decisions are made by individuals who naturally have a particular point of view that shapes the goal of a project and how it is conducted. Thus, to understand the situatedness of databases and their particular contexts, we must also, as Julia Flan-

10 Sarah-Mai Dang, "The Women Film Pioneers Explorer. What Data Visualizations Can Tell Us about Women in Film History," *Feminist Media Histories* 9, no. 1 (2023): 76–86. The interviews will be published once they have been edited.

ders states, take into account "digital humanities practitioners" – may it be scholars, designers, or archivists – as responsible parties."[11]

The Women Film Pioneers Project (WFPP) is an online platform for research on women in the silent film era, launched in 2013.[12] With more than three hundred career profiles, including filmographic and bibliographic information as well as a collection of film historical resources, its goal is to make women's global work more visible and facilitate further investigation.[13] It is run by co-founder Jane M. Gaines, who is based at Columbia University.[14] Kate Saccone serves as the project manager, who also edits and curates the profiles.[15] In addition, many people from various institutions have contributed to the website, both when it was planned as a printed encyclopedia and since it was converted to a digital format – graduate students, volunteers, web designers, external curators, and editors.[16]

In order to better understand the data collected and curated by the WFPP, the DAVIF research group conducted interviews with Jane M. Gaines and Kate Saccone about the genesis and approach of the initiative. We also conducted interviews with David Kleingers, who is head of the digital department and strategic development at DFF, and Bianca Sedmak, who manages the filmographic data editing there.

The filmportal.de was launched by the DFF – Deutsches Filminstitut & Filmmuseum in 2005.[17] With information on more than 150,000 films and 250,000 people, according to their own statement, it is the most comprehensive publicly published filmography of Europe. Its goal is to document all film productions in Germany from its beginnings to the present day in order to facilitate both academic and popular inquiries. The cinematographic works documented on film-

11 Julia Flanders, "Building Otherwise," in *Bodies of Information. Intersectional Feminism and Digital Humanities*, ed. Elisabeth Losh and Jacqueline Wernimont (Minneapolis: University of Minnesota Press, 2018), 289–304.

12 Jane M. Gaines, Radha Vatsal, and Monica Dall'Asta, *Women Film Pioneers Project* (New York: Columbia University Libraries, 2013), accessed May 2, 2022, https://wfpp.columbia.edu/.

13 Women Film Pioneers Project, "About the Project – Women Film Pioneers Project," accessed May 3, 2022, https://wfpp.columbia.edu/about/.

14 See https://wfpp.columbia.edu/; on the historical context of the WFPP see also Kate Saccone's chapter in this volume, "(Re)Visioning Women's Film History: The Women Film Pioneers Project and Digital Curatorial-Editorial Labor."

15 See Saccone, "(Re)Visioning Women's Film History," this volume. Saccone reflects on her role and the many different tasks her work as project manager involves.

16 Women Film Pioneers Project, "Editorial Team and Acknowledgments – Women Film Pioneers Project," accessed February 17, 2023, https://wfpp.columbia.edu/editorial-team-and-acknowledgments/. See also Saccone, "(Re)Visioning Women's Film History," this volume.

17 Deutsches Filminstitut & Filmmuseum, "Filmportal.de. Alles Zum Deutschen Film," accessed March 21, 2023, https://www.filmportal.de/.

portal.de have mostly been screened publicly as a theatrical release or at a festival, for example.

Relevant sources for new entries include the official release lists of the top organization of the film industry, Spitzenorganisation der Filmwirtschaft (SPIO), decisions of the Voluntary Self-Regulation Body (FSK), and information from national and regional film funding bodies. Other important sources are the catalog publications of national as well as international festivals, on the basis of which new film work records are also created.[18] The focus is on theatrical releases, but over the years more and more television films have been included as well. Although the majority of the films included in the database are German productions and co-productions, the collection does not exclusively focus on national cinematography. It also includes, as Sedmak pointed out in our interview, historical works the DFF considers relevant for film history, such as Battleship Potemkin.

The portal relies on permanent public funding by the Federal Government Commissioner for Culture and Media (Beauftragte der Bundesregierung für Kultur und Medien – BKM) and the Hessian Ministry for Science and Art (Hessische Ministerium für Wissenschaft und Kunst – HMWK). Additional funders are or were at some point the German Federal Film Board (Filmförderungsanstalt – FFA), the Federal Ministry of Economics and Technology (Bundesministerium für Wirtschaft und Technologie), the Initiative Culture & Creative Industries of the Federal Government (Initiative Kultur- und Kreativwirtschaft der Bundesregierung), and the Friedrich Wilhelm Murnau Foundation.[19]

While film scholars, historians, PhD students, film curators, and archivists contribute to the WFPP and there are different editors responsible for individual countries, the pioneer profiles are currently primarily edited and curated by film scholars Jane M. Gaines and Kate Saccone. The content of filmportal.de, on the other hand, is collected by special editorial teams who focus on filmographic data, text, and images. The scope of work is certainly large in both projects, but it is organized differently. Within the WFPP, there is a mixture of official positions and volunteer work. While the manager, director, and research assistant positions are official staff positions, the contributors' research and writing is conducted "on the side" like any other publishing labor. In contrast, for filmportal.de. all work is conducted as part of official positions at the DFF.

18 David Kleingers, interview by Sarah-Mai Dang, Pauline Junginger, and M. Leonie Biebricher, May 5, 2022.

19 David Kleingers, "Filmportal.de. Die zentrale Internetplattform zum deutschen Film," in *Handbuch Kulturportale: Online-Angebote aus Kultur und Wissenschaft*, ed. Ellen Euler et al. (Berlin, Munich, Boston: De Gruyter Saur, 2015), 204, https://doi.org/10.1515/9783110405774-018.

Although both initiatives aim to facilitate further research on film and film history, their goals differ in terms of theoretical, conceptual, and political aspects. The WFPP seeks to foster the visibility of women workers in early cinema and, in doing so, critically reflect the contingencies of film history and "reconfigure film knowledge,"[20] whereas the DFF intends to set standards for a systematic registration of film works.[21] For this purpose, the DFF has created an internal filmographic database (Zentrale Filmographische Datenbank – ZDB), which is the core of filmportal.de. The ZDB was originally the product of a merger of the previous database by the DFF (former DIF) and CineGraph – Hamburgisches Centrum für Filmforschung e.V. Both initiatives collect film historiographical and personal data. However, while the WFPP focuses on individuals and their careers, the DFF pays greater attention to cinematographic works. This is important to keep in mind when searching for specific aspects of film history in these databases, as they reflect these foci in terms of completeness and nuance. For instance, the WFPP provides more details about women's careers, whereas the DFF offers more comprehensive filmographies.

In principle, we must not forget that if a data collection is missing certain information, another may be able to provide it. As the Women Film Pioneers Explorer, a case study of our project, has shown, the United States is the country with the most WFPP profiles for the years 1895 to 1926, with 47.1% (163 entries).[22] It is followed by Great Britain (9.2%, 32 entries), France (5.2%, 18 entries), Germany (4.3%, 15 entries), and Australia (4.0%, 14 entries). There is almost no data on women who worked in Africa, the Middle East, Central Asia, or India. As I have elaborated elsewhere, these figures do not mean that women in the United States were the most active in early cinema.[23] They rather provide information about the content of the WFPP database, and reveal something of its history. These results are understandable insofar as the project started in the United States and is based in New York, as is the main initiator, Jane M. Gaines. In addition, the creation of profiles, from submission to editing to the final posting of new entries, reflects the fact that it requires a great deal of time and coordination work. Yet, a huge gap becomes visible. It is therefore important to be aware of the situatedness of data.[24] One must not lightly assume that there is no information on a specific country anywhere or that persons or films did not exist there at

20 Women Film Pioneers Project, "About the Project."

21 Kleingers, "Filmportal.de," 206.

22 Henri Dickel et al., "Women Film Pioneers Explorer," 2021, accessed March 21, 2023, https://www.online.uni-marburg.de/women-film-pioneers-explorer/.

23 Dang, "The Women Film Pioneers Explorer."

24 Dang, "The Women Film Pioneers Explorer."

all. Instead, we need to remind ourselves that no database can ever be complete. This might seem an obvious fact but can easily be forgotten. Databases are always the result of specific conditions in certain place and time constellations; they are part of particular RDM practices. Like scholarly publications, they represent only partial perspectives, which are nevertheless key to knowledge production.

2 Data Modeling and Categorizations

Which kind of work is involved in data modeling? How is the data organized and structured? What categories are considered relevant? How are political categories such as class, race, and gender addressed?

While it is important to have an interdisciplinary exchange about handling data, it is necessary to first understand RDM from within one's own discipline or in order to understand the implications of data-driven research. Looking at databases in the field of film history, it is clear that, even within a discipline, or even a particular area of a discipline, data can vary to a great extent.

A comparison of the WFPP and the DFF databases reveals the specific approaches of these initiatives. While the WFPP's data modeling can be defined as *research-driven* since it addresses a specific research interest, namely the study of women's work in early cinema, the DFF focuses primarily on the potential reuse of data and seeks to ensure interoperability through standards. Its approach can therefore be described as *curation-driven*.[25] This is certainly due to the institutional situatedness of the two initiatives. As Tim van der Heijden points out, because the WFPP is a research project that is institutionally embedded within a university and a university library, it maintains different data management protocols than the DFF. The latter is an institute within a museum that houses some comprehensive archives and collections of material on all aspects of film, and the filmportal.de website is only one of the DFF's many projects. Because of its funding and cultural policy position, the DFF works on an institutional level and is

25 In the context of data modeling, a distinction is generally made between curation-driven and research-driven. However, these definitions cannot always be clearly separated from each other, as I have already pointed out. See Sarah-Mai Dang, "O.J. – Recherchepraktiken, Datenquellen und Modellierungen," in *Doing Research. Wissenschaftspraktiken Zwischen Positionierung und Suchanfrage*, ed. Sandra Hofhues and Konstanze Schütze (Bielefeld: Transcript, 2022), 330–337; Fotis Jannidis and Julia Flanders, "A Gentle Introduction to Data Modeling," in *The Shape of Data in the Digital Humanities: Modeling Texts and Text-Based Resources*, ed. Julia Flanders and Fotis Jannidis (London, New York: Routledge/Taylor & Francis, 2019), 26–94.

more committed to a wider applicability or reuse of data than the WFPP, which is driven primarily by project-specific research questions on a micro level.[26]

As I have explained elsewhere, it was particularly interesting to learn during the interview with Kate Saccone that the diverse job titles collected in the database are the result of the research done by the authors.[27] As a result the database also contains job titles that one would not normally associate with film production, such as "society matron," "metalworker," or "carpenter." A taxonomy emerged from the first set of pioneer profiles and the authors' archival research, with final decisions made during conversations between the WFPP editors and the contributors.

As Saccone explains in this volume with regard to questions about standardization and interoperability, the taxonomy can be updated as needed.[28] She proofreads and copyedits the profiles and inserts the biographical and occupational metadata suggested by the authors into the website's content management system, WordPress.[29] It is important to note that Saccone's continuous "(re)visioning," as she conceptualizes it regarding editorial labor in the digital era, is not a linear process of RDM with clearly defined steps. Instead, it is complex work that involves, as she defines it, "digital curatorial processes of creation, presentation, preservation, and the ongoing management of digital (textual) materials and film historical knowledge."[30] Even though digital tools and infrastructures offer great support and make workflows much easier, working with and on data is mostly manual work and does not happen automatically. RDM demands intensive, painstaking work that requires specific knowledge and technical skills in addition to a wide range of expertise and responsibilities, including making decisions, organizing tasks, communicating with colleagues and institutions, and monitoring processes. This is why the DFF has a permanent filmographic editorial staff of two full-time employees who deal primarily with the new entry of film works and the maintenance of the existing database.[31] As Kleingers points out, data about a film changes frequently from the first official production announcement to the theatrical release, and its filmography is never complete. The editors are constantly returning to existing entries to add and correct information.[32]

---

26 I thank Tim van der Heijden for these important remarks about an earlier version of this chapter.

27 Dang, "The Women Film Pioneers Explorer."

28 Saccone, "(Re)Visioning Women's Film History," this volume.

29 The final terms are always chosen in collaboration with the authors and based on the existing taxonomy. For a detailed workflow see Saccone, "(Re)Visioning Women's Film History," this volume.

30 Saccone, "(Re)Visioning Women's Film History," this volume.

31 Kleingers, interview, May 5, 2022.

32 Kleingers, interview, May 5, 2022.

Based on the assumption that film history is itself characterized by coincidences, contradictions, and contingencies, the WFPP emphasizes epistemological uncertainties by allowing individual categorizations, even encouraging the authors to take their own approach.[33] This principle, which stems from theoretical and political considerations, reveals the numerous diverse professions that women held in early cinema, and thus their manifold impact in film history.[34] As it turns out, a decade after the launch of the WFPP, Jane M. Gaines contends that this approach has helped to identify an "incredible range of names and titles and types of occupations we never dreamed existed when we first began."[35]

Unlike the WFPP, the DFF models its data in an SQL database, a relational database with structured tables, according to the European standard EN15907. This standard defines a set of metadata to describe cinematographic works, including their variants and manifestations. EN15907 is not a data model itself but a scheme that offers a standardized approach for developing one.[36] As computer science students I work with on data visualization have noted, this standard includes a comprehensive terminology (although terms to be used in the DFF database are not specified) but, interestingly, no details on informational procedures are provided (e.g., that years must be numbers, or that names cannot be numbers). This can be a barrier to further data processing accuracy, as errors that could easily be prevented by technical specifications can creep into the database. This observation illustrates how differently people look at data, and how different backgrounds – in this case computer science and film historical and archival concepts and premises – determine priorities and goals in data processing.

As for the occupations, the DFF works with a terminology that is intended to be as reliable and pragmatic as possible. It is shaped by internal considerations, international film historiographical debates, and authority data of their project partners. The origin of the data is recorded as well as possible, as Bianca Sedmak explained in the interview.[37] All sources used are entered in a database, which can only be accessed via the internal ZDB user interface. If the referenced material is available in digital form, it is archived on an internal server. This includes press materials, scans of FSK cards, censorship cards, etc. The same procedure

33 Women Film Pioneers Project, "Guidelines: Profiles – Women Film Pioneers Project," accessed March 3, 2023, https://wfpp.columbia.edu/guidelines-profiles/.

34 Dickel et al., "Women Film Pioneers Explorer."

35 Jane M. Gaines, interview by Pauline Junginger, February 5, 2023.

36 Deutsches Filminstitut (DIF) e.V, "EN 15907," filmstandards.org, accessed March 1, 2023. http://filmstandards.org/fsc/index.php?title=EN_15907.

37 Bianca Sedmak, interview by Sarah-Mai Dang, M. Leonie Biebricher, and Pauline Junginger, May 19, 2022.

applies to correspondence such as e-mails or, in rare cases, letters.[38] Early film is documented in detail because, Sedmak explains, the DFF has incorporated all the encyclopedias, including censorship maps and secondary sources. Primary sources hardly still exist. It is also quite possible that some sources have just not been discovered yet, or even looked for. Sedmak points out that the DFF very much depends on encyclopedias and other research, and experts in the field of film history.[39]

The DFF has several controlled topic-based vocabularies for their database to establish consistent spellings, as Bianca Sedmak explained.[40] Their occupation vocabulary is divided into two areas. One area includes the credits of film-related work – that is, the relation between the film and the person. A large proportion of these terms is derived from the most commonly used terms in movie credits. The other area includes person-related work, which is related to the GND (Gemeinsame Normdatei) of the German National Library (DNB – Deutsche Nationalbibliothek). This vocabulary was adapted in the course of the collaboration with the GND. Since this vocabulary is synchronized and imported into the GND, it must be standardized and comply with the job descriptions of the GND. Both lists contain controlled vocabulary, so they cannot be manipulated manually. Each list includes around 140 to 150 job titles. The titles in the list of film-related works are in turn divided into 16 top categories. These categories are decisive for the occupation that is displayed on the person page at filmportal.de.

The DFF is far from being able to map every occupation in personal records, Sedmak notes.[41] With regard to interoperability with the German National Library, it is also not necessary to represent every profession in the database. For reasons of clarity, the terms are therefore limited to the most common ones.[42] For "model," for example, there is no term. It is resolved by using "actress" ("Darstellerin") for the occupation if the person did some acting, or "participation" ("Mitwirkung") if it is a documentary film about a model. The DFF seeks to describe the actual work of a person as precisely as possible, which is why non-film-related job titles such as "politician" or fields of work such as "medicine" can also be found among the person-related occupations.[43]

38 Sedmak, interview, May 19, 2022.
39 Sedmak, interview, May 19, 2022.
40 Sedmak, interview, May 19, 2022.
41 Sedmak, interview, May 19, 2022.
42 I thank Bianca Sedmak for the additional insights provided in her feedback on an earlier version of this chapter.
43 Sedmak, interview, May 19, 2022.

The modeling of job titles is quite understandable, although not uncomplicated. When it comes to gender classifications (e.g., "female," "male," "non-binary"), however, it not only gets complicated, but also messy. The DFF deals with these attributions in different ways. Depending on the specific table of the database, whether it is person-related or film-related, internal, or published data, the DFF uses gendered job titles or descriptions that refer to film divisions rather than job titles. For instance, they use "Szenenbild" (scenography) rather than "Szenendesigner:in" (which refers to a scene designer who is female, male, or non-binary). For some professions, however, both variants exist, for example, "Darsteller" (male actor) and "Darstellung" (acting). While the job titles in personal records are predominantly gendered, film-related categories that describe how a person is related to a film are sometimes gendered, and then based on the masculine form. To account for historical and cultural developments, multiple variants, in turn, have been retained, like "Szenenbild" (scenography) and "Bauten" (buildings). These terms appear synonymous but refer to different professions. Sedmak explained in the interview that some distinctions or nuances are not published on filmportal.de, but documented in the internal ZDB.[44] The different handling of gender attributions is remarkable but in a way also understandable. Gendered categorizations are sensitive and pose a political conceptual, and technical challenge.

It becomes apparent that standardizations cannot provide all-encompassing solutions for linking data and ensuring sustainable reuse, let alone comprehensive documentation of film historical aspects. Moreover, the question arises as to what is actually involved in the demand for standards. Which data or procedures should be standardized, and for what purpose? What is lost with standardization; what is gained?[45] While the implementation of standard metadata can foster interoperability and collaboration, it might also reinforce blind spots. Data can help make particular histories visible, but might also obscure certain aspects of the past.[46] The issue of standards is relevant not only in terms of job titles, but also in terms of gender. For example, if a record of film historical occupations does not have a gender assigned to it, either for the job title or the linked person,

44 The current data model of the ZDB is described in a semi-public Wiki, where the DFF also provides information on the controlled vocabulary of occupations. Deutsches Filminstitut (DIF) e.V, "Tätigkeiten – DIF Filmographie Wiki," filmstandards.org, accessed March 8, 2023, https://filmstandards.org/difzf/index.php?title=Tätigkeiten.

45 Dang, "O.J."

46 Sarah-Mai Dang, "Representing the Unknown: A Critical Approach to Digital Data Visualizations in the Context of Feminist Film Historiography," in *How Film Histories Were Made: Material, Methods, Discourses,* ed. Malte Hagener and Yvonne Zimmermann (Amsterdam: Amsterdam University Press, 2023), 467–493.

we cannot analyze, for instance, how many women or men or non-binary people have directed a film, or how this might have changed over time. Standards are crucial not only in terms of data modeling, but also in terms of what data we include or exclude. This raises the question of what information we consider key in the first place – and what we wish to be able to be retrieved in the course of further research.

Since the WFPP features women workers, it does not include male pioneers. The website understandably does not provide an option to search for gender, but nor does filmportal.de. However, the DFF's internal database can be searched by gender. The ZDB contains three gender categories: "female," "male," and "undefined," where "undefined" can refer to both "non-binary" and "unknown": a fourth category would actually be necessary to distinguish between the two. "Unknown" is usually employed by the GND in the sense of "undetermined" and not in the sense of "non-binary."[47] As Sedmak stated in the interview, gender categorization is important to the DFF and internal annotations are made in the ZDB when uncertainties or new findings arise. However, the representation of gender is rather secondary for filmportal.de, where the personal data pages do not say male or female, but only "Darsteller" (actor). The DFF team has been working on this for a long time and are strongly advocating that wording should be more accurate. But that would be an extensive undertaking that requires many resources.

As for databases that explicitly focus on gender representation, in my view the BFI Filmography (2012–2017)[48] can serve as good practice example. In a detailed documentation of their data modeling, which was accessible together with the data online, the curators describe which gender attributions they have made, how, and for what reason. In doing so, they reflect on the advantages and disadvantages of their approach. They rightly point out that, while their method is not perfect as binary categories and external gender attributions leave out nuances, the focus on gender can still foster further discussions about equality in film industries.[49]

Of course, not all the implications of categorization can always be considered, and it cannot be the goal to create a perfect database because, as already stated, there is no such thing, but we need to situate and, in doing so, critically reflect on the underlying processes of databases. Categorizations affect how we evaluate a person's role in history. Assigning specific metadata to people or objects is a powerful act of meaning-making.[50] By ignoring gender or other political categories, or

47 Sedmak, interview, May 19, 2022.

48 See https://filmography.bfi.org.uk/, accessed March 8, 2023.

49 See https://filmography.bfi.org.uk/, accessed March 8, 2023.

50 Miriam Posner, "What's Next. The Radical, Unrealized Potential of Digital Humanities," in *Debates in the Digital Humanities*, ed. Matthew K. Gold and Lauren F. Klein (Minneapolis, London:

by focusing exclusively on one individual, the director of a film, for instance, numerous people and facets of film history are neglected.[51] We also must pay attention, as Flanders argues, to what function attributions have, how they are used, and how they are conceived, whether they are applied as fact-stating descriptions that perpetuate a form of othering, or as a "category of discovery."[52]

3 Access and Reuse

Where is the data stored? How can a database be accessed? How is the interface designed? How can the data be retrieved? How is the data licensed? Who uses the database? How can users contribute?

As research practices change in the wake of digitalization, research results beyond the classical publication become more and more relevant.[53] Whereas five years ago Adelheid Heftberger and Marion Goller were critical of the very low interest in open access and reuse of research data in film studies, this has fortunately changed considerably in recent times.[54] As open science activists rightly argue, sharing research data of any kind and making it reusable without technical, financial, or legal barriers is the prerequisite for fruitful, effective, and inclusive knowledge production.[55] Reuse logically requires findability, accessibility, and interoperability, which is why the FAIR data principles – findable, accessible, interoperable, reusable – were created.[56] Since data reflect political, cultural, and social conditions, and thus the distribution of power, the Global Indigenous Data Alliance (GIDA) has developed the CARE principles (collective benefit, authority

University of Minnesota Press, 2016), accessed March 21, 2023, https://dhdebates.gc.cuny.edu/read/untitled/section/a22aca14-0eb0-4cc6-a622-6fee9428a357.

51 Dang, "Representing the Unknown," 469.

52 Flanders, "Building Otherwise," 296–298.

53 Dang, "Forschungsdatenmanagement in der Filmwissenschaft," 128–130.

54 Marion Goller and Adelheid Heftberger, "Die Öffnung von Forschungsdaten in den Film- und Medienwissenschaften: praktische und urheberrechtliche Herausforderungen," *<intR>²Dok [§] – Fachinformationsdienst für internationale und interdisziplinäre Rechtsforschung* (2018), accessed March 21, 2023, https://doi.org/10.17176/20180515-233758.

55 Heinz Pampel and Sünje Dallmeier-Tiessen, "Open Research Data: From Vision to Practice," in *Opening Science: The Evolving Guide on How the Internet Is Changing Research, Collaboration and Scholarly Publishing*, ed. Sönke Bartling and Sascha Friesike (Cham: Springer International Publishing, 2014), 213–224, accessed March 21, 2023, https://doi.org/10.1007/978-3-319-00026-8_14.

56 FORCE11, "The FAIR Data Principles," accessed March 21, 2023, https://www.force11.org/group/fairgroup/fairprinciples.

to control, responsibility, ethics) to complement the FAIR data principles by drawing attention to the power imbalances and historical contexts that they believe are being ignored by the current open data movement.[57] Who benefits from data-driven projects and who does not is an important question that applies not only to multinational technology companies but also to academia. As Catherine D'Ignazio and Lauren Klein argue, "asking these *who questions* allows us, as data scientists ourselves, to start to see how privilege is baked into our data practices and our data products."[58]

WFPP is regularly archived in the Wayback Machine, and pioneer profiles are deposited as PDF files, tagged with a DOI (digital object identifier), in Columbia's digital repository, Academic Commons. Although it is not clear on the website, the content created by the WFPP – not the images or films of third parties – has a Creative Commons Attribution license, according to Saccone. The biographical dataset I requested for the Women Film Pioneers Explorer case study is now archived on Academic Commons as well.[59] All the information on WFPP is freely accessible and reusable. This is somewhat different from the DFF. As already mentioned, filmportal.de is an output of the ZDB. It is detached from the DFF's central database and operates via a content management system programmed for its specific requirements, based on Drupal.[60] This means that the ZDB and filmportal.de contain slightly different information. Since the DFF is our project partner, the DAVIF research group has the privilege of accessing both datasets. Many of the contents and objects presented online on filmportal.de are protected by copyright and exploitation rights and exclusively for non-commercial use. While the portal is a public platform, the ZDB is for internal use or research projects only.[61] In addition, there is a web service for personal and corporate data such as

57 Global Indigenous Data Alliance, "CARE Principles," accessed January 23, 2023, https://www.gida-global.org/care.

58 Catherine D'Ignazio and Lauren F. Klein, *Data Feminism* (Cambridge, MA: MIT Press, 2020), 26.

59 Jane M. Gaines and Columbia University Libraries, "Women Film Pioneers Project Biographical Data," dataset compiled December 7, 2020, https://doi.org/10.7916/m4dc-n768.

60 Kleingers, "Filmportal.de," 212.

61 The filmographic data of the ZDB can be exported as XML data via an OAI-PMH interface. As Kleingers explained, this is used, for example, by the partners of the DFF, like the federal archive (Bundesarchiv) or the Goethe-Institut. The web service (another interface) is used by the German Film Academy (Deutsche Filmakademie), among others. Access to these interfaces is not generally available to the public, but has to be granted by the DFF. The reuse of the data obtained via this interface is regulated in respective contracts. The website filmportal.de has another interface for the export of object-related metadata (i.e. information on digital photos, videos, texts, etc., that are held in the portal). The data provided via this interface are used by the German Digital

film distribution or production companies.[62] New entries of personal information go directly into the GND and the GND links to the corresponding information on filmportal.de (and vice versa).[63] Backup copies are made regularly by the DFF itself.

Legal issues concerning immaterial goods, their ownership, and conditions of use, are usually highly complex and can only be decided on a case-by-case basis. Thus, the question of which research data can be shared and used is not easy to answer in principle.[64] In this regard, however, I was surprised to learn from a legal consultant that as long as the data obtained is not used for commercial but only for scholarly purposes, web scraping is legal. That is, as long as one does not want to build a competing product to IMDb, for instance, it is legal to download data from the website. In dealing with data usage agreements (Nutzungsvereinbarungen), in which my research assistant Pauline Junginger and I have invested considerable time, I have learned that the threshold of originality (Schöpfungshöhe) of datasets can be reached relatively quickly. On the other hand, however, if data are only facts, they are not protected by copyright. But if someone or an institution has invested a lot of time and money in a database, ancillary copyright (Leistungsschutzrecht) takes effect. Again, as always with legal issues, it depends on the particular case. It is also important to distinguish between data types, metadata, films, texts, or other – digitized or born digital – objects we are dealing with. The legal uncertainties that still prevail in this context may in any case prevent the free use of research data and the advancement of digital scholarship. Data is resource-intensive and thus expensive.[65] This is another reason why, in addition to the promotion of open data, we need to further familiarize ourselves with the legal framework.

Filmportal.de is used for general and scholarly inquiries. In addition to film historians, filmmakers are increasingly asking for information about their work to be updated. Thus the website serves, as Kleingers explains, both as a source of information and as a presentation platform.[66] With regard to the accessibility of the data, it should be noted that the implementation of English will be further

Library, the European Film Gateway, and the Europeana, among others. I thank David Kleingers for the additional insights provided in his feedback on an earlier version of this chapter.

62 Kleingers, "Filmportal.de," 212.

63 Kleingers, "Filmportal.de," 213.

64 Linda Kuschel, "Wem 'gehören' Forschungsdaten?" *Forschung & Lehre*, September 12, 2018, accessed March 21, 2023, https://www.forschung-und-lehre.de/forschung/wem-gehoeren-forschungsdaten-1013/.

65 D'Ignazio and Klein, *Data Feminism*, 41–47.

66 Kleingers, "Filmportal.de," 207.

developed. To date, the interface allows at least a rudimentary search with English terminology. For reasons of resource, the WFPP's content is currently only available in English. Like the DFF, the WFPP serves different interests and users. Remarkably, apart from film historical research, it is also used as a genealogical source. In the interview, Saccone reported that many family members reach out to WFPP. Once a family member asked the editorial team to rewrite a profile of a woman because in their opinion she was only secondarily a filmmaker. In conversation with the author, however, who had done extensive archival research, they decided to continue to represent this woman who made one film and perhaps did not see herself as a filmmaker, and the profile has remained. This was an ethical decision to increase the visibility around such women.[67]

In principle, everyone can contribute to the WFPP. In doing so, authors must adhere to the standards of the project, which are ensured by peer review. Although these specifications are not of a technical nature in the sense that one might associate with RDM, they are part of a standardized approach that has been developed for the project and is continuously adjusted.

As other film and media scholars have shown and I have argued elsewhere, while film historical databases such as the ZDB/filmportal.de or the WFPP provide valuable resources for film historical research, it is important to further explore digital technologies for creative reuse of existing data.[68] For instance, as already indicated, data visualizations can open up new perspectives on data and facilitate critical reflection on historical sources. The way data is presented has a major impact on how we can access, explore, and reuse it.[69] By means of scaling, for instance, data visualizations can provide orientation, as Deb Verhoeven claims. Due to the size of big datasets, anomalies or absences do not become apparent unless experiments are conducted to examine them.[70]

67 Kate Saccone, interview by Sarah-Mai Dang, M. Leonie Biebricher, and Pauline Junginger, March 25, 2022.

68 Dang, "Representing the Unknown."

69 Sarah-Mai Dang, "Unknowable Facts and Digital Databases: Reflections on the Women Film Pioneers Project and Women in Film History," *Digital Humanities Quarterly* 14, no. 4 (2020), accessed March 21, 2023, http://www.digitalhumanities.org/dhq/vol/14/4/000528/000528.html.

70 Deb Verhoeven, "Show Me the History! Big Data Goes to the Movies," in *The Arclight Guidebook to Media History and the Digital Humanities*, ed. Charles R. Acland and Eric Hoyt (Falmer: REFRAME Books, 2016), 172, accessed March 21, 2023, https://projectarclight.org/wp-content/uploads/ArclightGuidebook.pdf.

Conclusion

Since RDM is of great importance to our studies, as I hope to have shown, and this importance will grow rapidly, we should not regard it as an annoying necessity that has to be addressed in the next research proposal, but as an opportunity to help shape critical digital scholarship. RDM should be considered an integral part of our research as data becomes more and more important – not only to film and media studies, but to other disciplines as well. It is therefore crucial for researchers to reveal how the data they apply has been shaped. This is necessary to acknowledge the limitations and messiness of one's own methods.

While new approaches and practices can offer exciting perspectives on film culture and its past, additional expertise and strategies are needed. Data-driven approaches require a deep understanding of the underlying data, how it was created, by whom, and for what purpose. Up to now, a distinction has generally been made between technical and content-related work in data-driven projects. However, in view of the multi-layered entanglements, a closer collaboration between the two areas is required in order to do justice to the numerous aspects of RDM.

Needless to say, the comparison of the WFPP and the DFF is not conclusive. In addition to providing insights into particular RDM practices, their premises and implications, it should, first and foremost, serve as a starting point for further inquiries of other initiatives, not least our own RDM approaches and data-driven research projects. A closer look at how film historical metadata has been collected, selected, modeled, organized, preserved, and shared, has shown that data is shaped by a complex interplay between people, institutions, and infrastructures, as well as practical, technological, and theoretical premises.

As demonstrated, data may vary not only in quality but also in type and model due to different assumptions and goals. This makes it difficult to map data across projects. Since each of the WFPP and DFF databases was developed independently for specific purposes and in different languages, it seems impossible to link these two. However, comparing heterogeneous data, as in this case, can provide many insights into the conditions of historiographic knowledge production. It helps in bringing initiatives together – not necessarily on a technical level, but certainly in terms of film historiographical questions – and to further stimulate critical data discourses.

To understand the RDM of the DFF and the WFPP, the semi-structured interviews were extremely helpful. In addition to discoveries made through the data visualizations, many insights were gained through the conversations with the individuals involved in the work. The exchange has intensified the cooperation with our project partners, but, ideally, all projects should document their data practices. We can learn from other disciplines such as ethnography and social sci-

ences about documentation protocols. For instance, Van der Heijden and Kolkowski took inspiration from history of science in how they document experiments in laboratory settings.[71] Library science could also be helpful: it might be interesting to look more closely at the history and infrastructure of libraries in relation to RDM within academic research practices, not least because of the increasing collaboration between researchers and librarians in the storage, access, and re-use of data in data-driven projects. Systematic documentation through data papers, for example, which is also becoming more widely discussed in film and media studies,[72] would not only help one to better understand one's own premises and practices, but also open up scholarship. By situating relevant datasets like we do with scholarly publications, we can strengthen data-driven projects and further advance digital film history.

References

Dang, Sarah-Mai. "Representing the Unknown: A Critical Approach to Digital Data Visualizations in the Context of Feminist Film Historiography." In *How Film Histories Were Made: Material, Methods, Discourses,* edited by Malte Hagener and Yvonne Zimmermann, 467–493. Amsterdam: Amsterdam University Press, 2023.

Dang, Sarah-Mai. "The Women Film Pioneers Explorer. What Data Visualizations Can Tell Us about Women in Film History." *Feminist Media Histories* 9, no. 1 (2023).

Dang, Sarah-Mai. "O.J. – Recherchepraktiken, Datenquellen und Modellierungen." In *Doing Research. Wissenschaftspraktiken Zwischen Positionierung und Suchanfrage*, edited by Sandra Hofhues and Konstanze Schütze, 330–337. Bielefeld: Transcript, 2022.

Dang, Sarah-Mai. "Forschungsdatenmanagement in der Filmwissenschaft: Daten, Praktiken und Erkenntnisprozesse." *montage AV Zeitschrift für Theorie und Geschichte audiovisueller Kommunikation* 29, no. 1 (2020): 119–140.

Dang, Sarah-Mai. "Unknowable Facts and Digital Databases: Reflections on the Women Film Pioneers Project and Women in Film History." *Digital Humanities Quarterly* 14, no. 4 (2020). Accessed March 21, 2023. http://www.digitalhumanities.org/dhq/vol/14/4/000528/000528.html.

Deutsches Filminstitut (DIF) e.V. "Tätigkeiten – DIF Filmographie Wiki." filmstandards.org. Accessed March 8, 2023. https://filmstandards.org/difzf/index.php?title=Tätigkeiten.

71 Tim van der Heijden and Aleksander Kolkowski, *Doing Experimental Media Archaeology: Practice* (Berlin, Boston: De Gruyter Oldenbourg, 2023), https://doi.org/10.1515/9783110799767.

72 Data papers in journals like *Research Data Journal for the Humanities* or *Social Science* or *Journal of Open Humanities Data* are very intriguing and offer important insights into RDM; see, for example, Thunnis van Oort and Julia Noordegraaf, "Structured Data for Performing Arts History: An Introduction to a Special Issue of Data Papers: Arts and Media," *Research Data Journal for the Humanities and Social Sciences* 5, no. 2 (2020): 1–12. https://doi.org/10.1163/24523666-bja10008.

Deutsches Filminstitut (DIF) e.V. “EN 15907.” filmstandards.org. Accessed March 1, 2023. https://filmstandards.org/fsc/index.php?title=EN_15907.

Deutsches Filminstitut & Filmmuseum. “Filmportal.de – Alles Zum Deutschen Film.” Filmportal. Accessed July 15, 2023. https://www.filmportal.de/.

Dickel, Henri, Matija Miskovic, Kharazm Noori, Christian Schmidt, Atefeh Soltanifard, Sarah-Mai Dang, and Thorsten Thormählen. “Women Film Pioneers Explorer.” Accessed March 21, 2023. https://www.online.uni-marburg.de/women-film-pioneers-explorer/.

D’Ignazio, Catherine, and Lauren F. Klein. *Data Feminism*. Cambridge, MA: MIT Press, 2020.

Einwächter, Sophie G. “Forschungsdaten (in) der Film- und Medienwissenschaft – Sophie G. Einwächter über vorurteilsbehaftete Begrifflichkeiten und fruchtbare Momente in der Lehre.” *Open Media Studies Blog* (2019). Accessed March 21, 2023. https://mediastudies.hypotheses.org/1314.

Fickers, Andreas, Juliane Tatarinov, and Tim van der Heijden. “Digital History and Hermeneutics – Between Theory and Practice: An Introduction.” In *Digital History and Hermeneutics: Between Theory and Practice*, edited by Andreas Fickers and Juliane Tatarinov, 1–20. Berlin, Boston: De Gruyter Oldenbourg, 2022.

Flanders, Julia. “Building Otherwise.” In *Bodies of Information. Intersectional Feminism and Digital Humanities*, edited by Elizabeth Losh and Jacqueline Wernimont, 289–304. Minneapolis: University of Minnesota Press, 2018.

FORCE11. “The FAIR Data Principles.” Accessed March 21, 2023. https://www.force11.org/group/fairgroup/fairprinciples.

Forschungsdaten.info. “English Pages | Forschungsdaten Und Forschungsdatenmanagement.” Accessed February 3, 2023. https://forschungsdaten.info/english-pages/.

Gaines, Jane M., and Columbia University Libraries. “Women Film Pioneers Project Biographical Data.” Dataset compiled December 7, 2020. https://doi.org/10.7916/m4dc-n768.

Gaines, Jane M., Radha Vatsal, and Monica Dall’Asta. *Women Film Pioneers Project*. New York: Columbia University Libraries, 2013. Accessed May 2, 2022. https://wfpp.columbia.edu/.

Gitelman Lisa, and Virginia Jackson. “Introduction.” In *“Raw Data” Is an Oxymoron*, edited by Lisa Gitelman, 1–14. Cambridge, MA: MIT Press, 2013.

Global Indigenous Data Alliance. “CARE Principles.” Accessed January 23, 2023. https://www.gida-global.org/care.

Goller, Marion, and Adelheid Heftberger. “Die Öffnung von Forschungsdaten in den Film- und Medienwissenschaften: praktische und urheberrechtliche Herausforderungen.” *<intR>²Dok [§] – Fachinformationsdienst für internationale und interdisziplinäre Rechtsforschung* (2018). Accessed March 21, 2023. https://doi.org/10.17176/20180515-233758.

Hoekstra, Rik, and Marijn Koolen. “Data Scopes for Digital History Research.” *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 52, no. 2 (2019): 79–94. https://doi.org/10.1080/01615440.2018.1484676.

Jannidis, Fotis, and Julia Flanders. “A Gentle Introduction to Data Modeling.” In *The Shape of Data in the Digital Humanities: Modeling Texts and Text-Based Resources*, edited by Julia Flanders and Fotis Jannidis, 26–94. New York: Routledge/Taylor & Francis, 2019.

Kleingers, David. “Filmportal.de – Die zentrale Internetplattform zum deutschen Film.” In *Handbuch Kulturportale: Online-Angebote aus Kultur und Wissenschaft*, edited by Ellen Euler, Monika Hagedorn-Saupe, Gerald Maier, Werner Schweibenz, and Jörn Sieglerschmidt, 204–214. Berlin, Munich, Boston: De Gruyter Saur, 2015. https://doi.org/10.1515/9783110405774-01.

Kuschel, Linda. “Wem ‘gehören’ Forschungsdaten?” *Forschung & Lehre*, September 12, 2018. Accessed March 21, 2023. https://www.forschung-und-lehre.de/forschung/wem-gehoeren-forschungsdaten-1013/.

Noordegraaf, Julia, Kathleen Lotze, and Jaap Boter. “Writing Cinema Histories with Digital Databases: The Case of Cinema Context.” *TMG Journal for Media History* 21, no. 2 (2018): 106–126. https://doi.org/10.18146/2213-7653.2018.369.

Oort, Thunnis van, and Julia Noordegraaf. “Structured Data for Performing Arts History: An Introduction to a Special Issue of Data Papers: Arts and Media.” *Research Data Journal for the Humanities and Social Sciences* 5, no. 2 (2020): 1–12. https://doi.org/10.1163/24523666-bja10008.

Pampel, Heinz, and Sünje Dallmeier-Tiessen. “Open Research Data: From Vision to Practice.” In *Opening Science: The Evolving Guide on How the Internet Is Changing Research, Collaboration and Scholarly Publishing*, edited by Sönke Bartling and Sascha Friesike, 213–224. Cham: Springer International Publishing, 2014. https://doi.org/10.1007/978-3-319-00026-8_14.

Posner, Miriam. “What’s Next: The Radical, Unrealized Potential of Digital Humanities.” In *Debates in the Digital Humanities*, edited by Matthew K. Gold and Lauren F. Klein. Minneapolis: University of Minnesota Press, 2016. Accessed March 21, 2023. https://dhdebates.gc.cuny.edu/read/untitled/section/a22aca14-0eb0-4cc6-a622-6fee9428a357.

Van der Heijden, Tim, and Aleksander Kolkowski. *Doing Experimental Media Archaeology: Practice*. Berlin, Boston: De Gruyter Oldenbourg, 2023. https://doi.org/10.1515/9783110799767.

Verhoeven, Deb. “Show Me the History! Big Data Goes to the Movies.” In *The Arclight Guidebook to Media History and the Digital Humanities*, edited by Charles R. Acland and Eric Hoyt, 165–183. Falmer: REFRAME Books, 2016. Accessed March 21, 2023. https://projectarclight.org/wp-content/uploads/ArclightGuidebook.pdf.

Women Film Pioneers Project. “About the Project – Women Film Pioneers Project.” Accessed May 3, 2022. https://wfpp.columbia.edu/about/.

Women Film Pioneers Project. “Editorial Team and Acknowledgments – Women Film Pioneers Project.” Accessed February 17, 2023. https://wfpp.columbia.edu/editorial-team-and-acknowledgments/.

Women Film Pioneers Project. “Guidelines: Profiles – Women Film Pioneers Project.” Accessed March 3, 2023. https://wfpp.columbia.edu/guidelines-profiles/.

Filmography

Battleship Potemkin, dir. Sergei Eisenstein. Soviet Union, 1925.

Kate Saccone

(Re)Visioning Women's Film History: The Women Film Pioneers Project and Digital Curatorial-Editorial Labor

Introduction

"One of the first Hispanic women to attain stardom in the silent cinema, Beatriz Michelena appeared in at least a dozen films between 1914 and 1920 and headed her own production company, Beatriz Michelena Features, from 1917 until 1920."[1] So starts the career profile on Michelena, written by MaryAnne Lyons, for the Women Film Pioneers Project (WFPP), a digital publication and scholarly and archival resource hosted by Columbia University Libraries, developed and run by film scholar Jane M. Gaines.[2] On the one hand, in foregrounding Michelena's fluid work as both actress and producer, the profile reflects WFPP's larger mission to recoup what has often been overlooked in traditional film historical accounts of the silent era: women's behind-the-scenes participation, at an international level, as directors, producers, editors, screenwriters, distributors, exhibitors, and more. On the other hand, in highlighting Michelena's status as an early Latina American star (her father was Venezuelan),[3] the profile exemplifies WFPP's desire, through the specificity of individual careers, to construct a wider and more diverse picture of cinema's first two decades.

Michelena's profile was one of approximately 180 similar entries that were featured on the website, alongside eight longer thematic essays and an array of bibliographic and archival resources, when WFPP was launched in October 2013. Since then, one of the project's overarching goals has been to jumpstart further research. The structural inclusion, in 2019, of a "Research Update" box on the back-end of the website, which could be added to any profile as needed, is one of the ways that the project has tried to center the processual nature of film historiographical research, especially as more primary sources are digitized and made accessible through online databases and archival platforms. Thus, it was not un-

1 MaryAnne Lyons, "Beatriz Michelena," in *The Women Film Pioneers Project*, ed. Jane M. Gaines, Radha Vatsal, and Monica Dall'Asta (New York: Columbia University Libraries, 2013), https://doi.org/10.7916/d8-76bd-6466.

2 See https://wfpp.columbia.edu/, accessed April 1, 2024.

3 Some scholars have previously said that Michelena was Mexican-American, but that has since proven to be incorrect.

https://doi.org/10.1515/9783111082486-008

welcome, both logistically and conceptually, when film scholar Laura Isabel Serna got in touch with the editorial team in 2020 with some new information about Michelena. Serna had found an obituary for Michelena's father, the opera singer Fernando Michelena, that called him a "devoted stepfather" to Beatriz and her sister. As the project manager and an editor of WFPP, I decided, after some further research, that this piece of information was worth presenting to readers, even if it was not confirmed elsewhere. I took what Serna had sent me and fashioned it into a short editorial note that I published as a Research Update on the front-end of the profile. Appearing at the bottom of the page, the dated box, which included a link to a copy of the obituary, outlined Serna's findings and the remaining absences (e.g., she had not yet been able to locate a birth record for Beatriz). Following this update, a new PDF version of the profile was also deposited in Columbia's digital repository, Academic Commons, in line with the project's preservation policy. As such, without changing the original profile, the short editorial update that I compiled, presented, and preserved functioned as both a reminder to keep questioning film historical knowledge and as a prompt for potentially new avenues of collective investigation around Michelena (Figure 1).

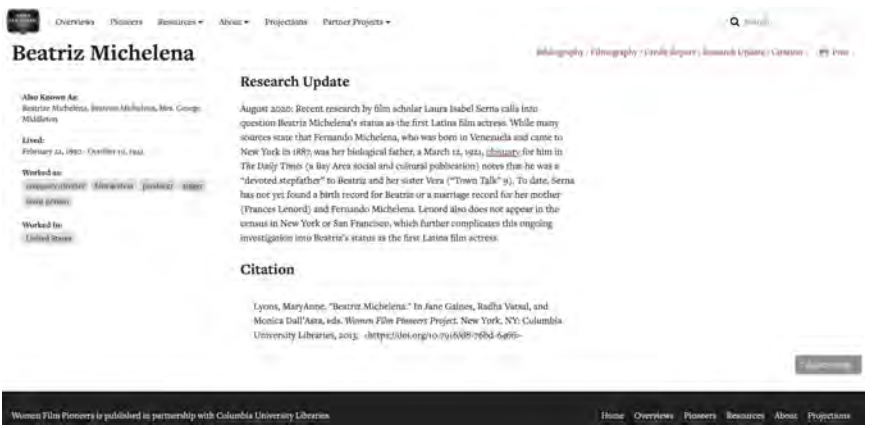


Figure 1: The Research Update on Beatriz Michelena's WFPP profile, screenshot.

I bring up Michelena not to challenge her status as one of the first Latina American film stars and producers. Rather, this anecdote serves to introduce the focus of this chapter, which considers WFPP in relation to both the iterative nature of film historiography and the mutability of digital scholarship and online publishing, as well as the practice of digital curation. Not only does this anecdote show that maintaining a long-running feminist digital humanities resource like WFPP is ongoing and critical film historiographic work, it also reflects the ways in

which editorial labor has expanded in the digital era, revealing its close proximity here to digital curatorial processes of creation, presentation, preservation, and the ongoing management of digital (textual) materials and film historical knowledge. As such, I posit that WFPP relies on and makes visible a digital curatorial-editorial practice that I call "(re)visioning," which draws on the open-ended processes of creating visibility at the heart of feminist film historiography and the practice of versioning at the heart of digital humanities.[4]

While this (re)visioning emerges from a particular feminist case study, it does not exist in a vacuum. In fact, it parallels the recent use of the term "updatism" by Frédéric Clavert and Andreas Fickers in their capacity as editors of the *Journal of Digital History*. Drawing on the field of memory studies, they define updatism as a computational practice reflective of an era "in which the memory of the past is constantly updated," and affirm that to publish in this era means keeping scholarly content continuously readable, explorable, and conceptually relevant.[5] This emphasis on change – or, as I see it, technical or conceptual movement – is an important way to understand what it means to produce, edit, and disseminate (film) historical knowledge in the digital era. In other words, although the idea of updating knowledge of the past in the (historical) present is hardly new to the digital era, with online publishing projects like WFPP, the ability to update our understanding of the past has become a central editorial feature. I use the term (re)visioning rather than updatism for two reasons: first, to hopefully circumvent the risk that the latter carries, to quote Steve F. Anderson, of "all too easily reassure[ing] us that the injustices of the past are being systemically redressed and overwritten by a more enlightened present";[6] and second, to emphasize a non-linear, fragmented, and iterative editorial practice rooted in feminist film historiography.

WFPP is, of course, not the only digital feminist film historical project to deal with updates. For instance, the Importing Asta Nielsen Database is currently at "edition #7" (as of July 2022), with many added features and content, and the editors of the Nordic Women in Film platform regularly make updates to the film-

4 This is slightly different than Adrienne Rich's feminist formulation of "re-vision," or "the act of looking back, of seeing with fresh eyes, of entering an old text from a new critical direction." See Adrienne Rich, "When We Dead Awaken: Writing as Re-Vision." *College English* 34, no. 1 (1972): 18.

5 Frédéric Clavert and Andreas Fickers, "Publishing Digital History Scholarship in the Era of Updatism," *Journal of Digital History* (2022), accessed April 1, 2024, https://journalofdigitalhistory.org/en/article/m7DWqDjY3hoV.

6 Steve F. Anderson, *Technologies of History: Visual Media and the Eccentricity of the Past* (Hanover: Dartmouth University Press, 2011), 168.

maker profiles, especially the ones with active careers.[7] But, as I will show, the specific features for updates implemented on WFPP and the type of work that the editors have fostered are particularly productive for understanding some of the practical and conceptual realities of film historical digital scholarship.

My interest in WFPP's editorial operations is the result of my longstanding position within the editorial team. I became involved with WFPP in 2011 as a graduate student research assistant at Columbia where I was part of the team finalizing the website for the online project launch. In 2013, only a few months before the launch, I became the project manager, a role that I continue to perform a decade later. As project manager, I oversee all aspects of the editing and publishing processes, among other outreach and administrative tasks. Moreover, alongside Gaines and the technical team, I have spent considerable time discussing how to update an ever-expanding academic resource with new profiles and essays, as well as how to make changes to existing scholarship. While many of the initial technological and infrastructural conversations and decisions (e.g., the choice to use WordPress) took place prior to my involvement – making it difficult for me, as a non-technical person who learned some basics "on-the-job," to speak to the project's technological development – my hands-on proximity to the editorial side of the project is unique. This chapter, which draws on public and private information, work documentation, emails with colleagues, and my own experiences, is thus rooted in a desire to make sense of my specific practical and conceptual work at a theoretical level in the context of both WFPP and digital film historiography more broadly.

WFPP has been defined in many ways over time, with "database" and "archive" being the most common terms. For the purposes of this chapter, I have chosen to give attention to its presentation of original scholarship over its database structure by introducing it above as a "digital publication and scholarly and archival resource." Moreover, although I recognize that WFPP could be called an archive, with the profiles and essays collected, preserved, and presented within it, I refrain from using that term here. This is partly because I think that "archive" is currently used too liberally, and partly because, while the notion of the archive is embedded in WFPP in more ways than one (hence making it archival), I think "archive" can also downplay its ongoing textual scholarship. Furthermore, while WFPP does feature some digital copies of historical moving and still images and archival documents, these are not the primary focus on the platform (both in terms of design and editorial labor) or probably the main reason for user-engagement. I

7 See "About Us," *Importing Asta Nielsen Database*, accessed April 1, 2024, https://importing-asta-nielsen.online.uni-marburg.de/; Tove Thorslund, email to author, January 27, 2023. See also https://nordicwomeninfilm.com/, accessed April 1, 2024.

would thus also not call WFPP a "scholarly digital edition," which has been defined by Patrick Sahle as a web-based academic project that offers a "critical representation of historic documents."[8] However, as I hope will become clear, I do see similarities between WFPP and the scholarly digital editions that Sahle discusses, from their shared position within a so-called digital paradigm to the ways that critical editorial labor becomes an ongoing part of the scholarly creation and dissemination process. Thus, to frame WFPP here as a "digital publication and scholarly and archival resource" allows me to home in on the project's editorial and publishing dimensions, while remaining open to the particular film historiographic, archival, and humanistic research impulses that emerge from and are embedded within it.

In the first part of this chapter, I will survey WFPP's development, from its early years as an analog archival research project in the mid-1990s and a planned multivolume book series in the 2000s, to its launch as an online-only institutional project in 2013, and its present manifestation as a well-known academic digital resource. Following that, I will discuss how editorial labor has expanded and shifted in the digital era, including through the practice of versioning. I will then present the concept of digital curation, a data-driven term that follows a so-called lifecycle model of ongoing, active management of digital(ized) materials and data. Although I use it loosely, it is a productive framework for understanding the iterative digital editorial labor – from initial selection, preparation, and publication through any necessary updates – currently carried out on WFPP, especially regarding the profiles. By drawing a connection between iterative and expanded editorial labor and digital curation in this way, I will show how WFPP's integration of the aforementioned Research Update boxes, as well as the implementation of textual versioning via digital object identifiers (DOIs), minor updates, and versioned records in Academic Commons, can all be understood as a part of the broader digital curatorial-editorial practice of (re)visioning, or a critical-feminist perspective on scholarly editorial labor that is open to continued historiographic movement and its ensuing management online.

The Women Film Pioneers Project

In addition to providing some background for what I will discuss thereafter, this survey functions as one of the only published accounts to date of the project's full

8 Patrick Sahle, "What is a Scholarly Digital Edition?" in *Digital Scholarly Editing: Theories and Practices*, ed. Matthew James Driscoll and Elena Pierazzo (Cambridge: Open Book Publishers, 2016), 23.

history.[9] Around 1994, Jane M. Gaines started collecting the names of silent-era women film directors, screenwriters, producers, and editors that she was discovering in her research and through her colleagues' archival investigations. As she later recounted, she had two motives for gathering these archival traces: she was both "fascinate[ed] with collecting – names upon names upon names" and interested in seeing how this labor challenged "a then-feminist orthodoxy," which entailed an "implicit prohibition against empirical work in favor of theory."[10] Gaines' embrace of empirical evidence, part of the historical shift in film studies at that time, reflected a period of renewed attention to silent cinema and the archive as well as women filmmakers. As a result, the historical narrative constructed in the 1970s and 1980s, in which there were virtually "no women" behind the camera during the silent era, was thoroughly upended. As Gaines rightly notes, historians like Anthony Slide and feminist scholars like Sharon Smith published research on some women filmmakers in the 1970s,[11] but the 1990s marked a shift from widespread absence to "a flood of empirical evidence"[12] and, consequently, a much larger awareness, at least in film archival and scholarly domains, that women had worked at all levels of the global silent film industry. Emerging from this dramatic feminist film historiographical shift from lack to presence, WFPP as an act of collecting names not only constituted a challenge to the "theoretical investment in women's 'absence'," which was central for feminist psychoanalytical film theory.[13] It was also a way of capturing (and attesting to) a new historical visibility confronting contemporary film scholars.

Gaines' task of collecting eventually became the Women Film Pioneers Project at Duke University, where she was then a professor. As more women were recovered and more scholars contributed research, the project was conceived as a multi-

9 Such a brief survey cannot do justice to the many people at various institutions who have contributed their intellectual insight and labor to the project (as editors, graduate student research assistants, project managers, web developers, library collaborators, volunteers, and administrators). I do not want to erase the thirty years of labor that has gone into developing, sustaining, and expanding WFPP, both in print and online. See https://wfpp.columbia.edu/editorial-team-and-acknowledgments/, accessed April 1, 2024.

10 Jane M. Gaines, "On Not Narrating the History of Feminism and Film," *Feminist Media Histories* 2, no. 1 (2016): 11.

11 Anthony Slide published *Early Women Directors* in 1977 and Sharon Smith's research was published in 1973 in *Women and Film*. Alongside these first publications was work by Sandy Flitterman-Lewis, Karyn Kay, and Gerald Peary, among others. See Gaines, "On Not Narrating," 19.

12 Monica Dall'Asta and Jane M. Gaines, "Prologue: Constellations: Past Meets Present in Feminist Film History," in *Doing Women's Film History: Reframing Cinemas. Past and Future*, ed. Christine Gledhill and Julia Knight (Urbana: University of Illinois Press, 2015), 14.

13 Dall'Asta and Gaines, "Prologue," 15.

volume academic book series. Volume I was to focus on the United States and Latin America, while a later Volume II would focus on Europe, Canada, Australia, Asia, and the Middle East. Within these national frameworks, the books were designed to emphasize the individual. Comprised mostly of short profiles written by scholars and archivists – many active in the Women and Film History International network and its affiliated conference, Women and the Silent Screen[14] – the goal was to centralize a woman's individual career and its output in order to challenge the established idea that only men shaped cinema in its first two decades. Central to this was the inclusion of an archival filmography in each profile where applicable, which listed any extant film holdings in order to jumpstart further research, preservation inquires, and exhibition activities. In addition to profiles, the manuscript for Volume I, which was written and compiled in the early 2000s, contained a handful of longer overview essays dedicated to national cinemas and specific occupations, as well as several appendices containing bibliographic and archival references. Thus, planned as a standard print academic book, the aim of this second iteration of WFPP was to introduce scholars and students to specific empirical evidence – hundreds of early women filmmakers and their surviving films – historical information and archival materials that had been largely absent from previous film historical narratives.

Around 2007, Gaines relocated to Columbia. While the plan was still to publish Volume I with the University of Illinois Press, she was soon approached by Columbia library staff looking for projects for the university's recently established Center for Digital Research and Scholarship (CDRS).[15] As a result, by 2008–2009, the project comprised the two planned books "with digital on-line components."[16] The planned ancillary website would offer space for supplementary visual materials and relevant event announcements as well as, potentially, additional profiles.[17] However, according to Rebecca Kennison, the founding director of CDRS, "in further discussions, it soon became clear that [the University of Illinois Press] was not so keen on doing the encyclopedia after all, as it didn't make sense financially for them, since fewer and fewer libraries were buying encyclopedic materials and, at that time,

14 Women and the Silent Screen began around the same time as WFPP (the first edition was held in the Netherlands in 1999) and often involves many of the same feminist film historians and archivists. See Gaines, "On Not Narrating."

15 Nancy Friedland, the Librarian for Film Studies and Performing Arts at Columbia, connected Gaines with CDRS. Rebecca Kennison, email to author, December 13, 2022.

16 Mark P. Newton, Jackson Harvell, and Leyla S. Williams, "Women Film Pioneers Project (WFPP): Presentation at Coalition for Networked Information, Fall Forum 2013" (2013), https://doi.org/10.7916/D8GB223R.

17 Kennison, email, December 13, 2022.

Illinois was not positioned to do online products."[18] Thus, by 2010, the project had dropped the planned print books entirely and had reconceived itself as an open-access online-only database that would be published by CDRS. The accompanying in-progress website, which was then hosted on MediaWiki, was consequently redesigned by the CDRS team to account for the inclusion of Volume I's content. New sections that corresponded with the manuscript, such as "Overview Essays," "Women Film Pioneers," and other categories drawn from the appendices ("Archives," "Bibliographies") were added as tabs at the top, as were new categories such as "Contributors" and, for a short time, "Contact and Message Boards" (eventually, "Contact"). As a result of this reconceptualization, the individual profile pages (which closely followed their manuscript model), the overview essays, and the various resources were now placed together on equal terms as WFPP was transformed from a planned print book, with its linear trajectory, into an online database with the potential for open-ended browsing.

In 2011, the content of the in-progress website was migrated to WordPress. I was not involved at the time of this decision but, according to Kennison, WordPress was chosen by the CDRS team for several reasons, including that the lead web designer was most comfortable with that platform, and that the team believed that WordPress would be easier for the non-technical editorial team members to learn compared to other content management systems.[19] For these reasons, and probably also because CDRS led the development phase of this project before taking a back-seat role after the launch, the migration to WordPress was primarily a technical rather than an editorial decision, driven by specific institutional abilities and goals. At the same time, it is worth noting that the migration to WordPress also solidified a direction taking shape at the editorial level: a move away from the possibilities for public editing and community content collaboration on a wiki-based website.

Most importantly, however, WordPress allowed CDRS staff to create a number of customized back-end infrastructural features that they had been unable to employ on MediaWiki. One such customization was the implementation of a back-end taxonomy of industry occupations that functioned as front-end tags on the profiles. Organized alphabetically on the back-end, the taxonomy of occupations listed a variety of relevant terms, such as "director," with corresponding child terms, such as "co-director" or "assistant director." (A simpler taxonomy for the regions in which women worked was also created on the back-end.) The taxonomy was created by the editorial team in collaboration with CDRS colleagues, with the terms themselves coming mainly from the Volume I profiles and the con-

18 Kennison, email, December 13, 2022.

19 Kennison, email, December 21, 2022.

tributors' research.[20] While listing occupations (and regions) at the top of a profile had been an important component of both the print manuscript and the MediaWiki website, these had never been organized into relational fields that could be systemized and hyperlinked across the entire resource. As a result of this work on WordPress, in the finalized version of the website, users could not only move between profiles based on related occupation and region tags,[21] they could also sort via these same occupational and national categories on the profiles landing page, which was organized alphabetically by default.

Between 2011 and 2013, the customized WordPress website went through several front-end aesthetic iterations before everything was finalized in the spring and summer of 2013. The official public launch of the WFPP website, in October 2013, ran in tandem with two complementary film programs at the Museum of Modern Art in New York.[22] In the decade since its launch, WFPP has added many new profiles and (now blind peer-reviewed) overview essays as well as new bibliographic and archival resources.[23] In 2019, the editorial team collaborated with new staff within the Digital Scholarship division of the Columbia libraries (into which CDRS had been folded in 2017) on a redesign of the WordPress website aimed at, among other things, decreasing back-end plugin dependencies and ensuring mobile functionality, as well as creating a better user experience by making the front-end of the website more stylistically and functionally up-to-date and accessible.[24]

This relaunch also marked the moment in which the WFPP editorial and technical teams actively responded to the increasing centralization of data-driven film

20 More should be said about the creation and implementation of WFPP's occupational taxonomy as an editorial entity than space allows. However, following its initial pre-launch creation by the technical and editorial teams in collaboration with a Columbia metadata librarian, the selected tags are now chosen or suggested by the contributors in discussion with the editorial team (and the metadata librarian when warranted). Designed to be an open-ended taxonomy that can be updated as needed, it reflects the diversity of ways that cinematographic work was categorized in the silent era, while also increasingly raising issues about standardization and interoperability given the idiosyncratic nature of this taxonomy.

21 Before a website redesign in 2019, related names were also listed at the bottom of each profile.

22 "Women Film Pioneers Project Launched at the Museum of Modern Art, New York City!" *The Women Film Pioneers Project*, November 25, 2013, accessed April 1, 2024, https://wfpp.columbia.edu/2013/09/25/10192013-women-film-pioneers-project-at-the-museum-of-modern-art-new-york-city/.

23 As of this writing, there are 320 women (from six continents) represented in the published profiles, and fifteen overview essays.

24 "The Women Film Pioneers Project Relaunches!" *The Women Film Pioneers Project*, October 18, 2019, accessed April 1, 2024, https://wfpp.columbia.edu/2019/10/18/the-women-film-pioneers-project-relaunches/.

historiographical practices and discourses within the field of film and media studies. Until this point, as it had been designed to emphasize the text-based scholarship via customized back-end page templates, the WordPress website had effectively codified what could be published. As part of the 2019 relaunch, a new section of the website, called Projections, was added, in which the editorial team hoped to include more digital-friendly and visual approaches to silent film research and feminist scholarship beyond the profile or essay format.[25] Although arguably hampered by the technical limitations of WordPress, along with the fact that any software and design support from the library would take time, given how many other projects it hosts, Projections reflected the editorial team's active interest in digital methods for presenting and disseminating research, alongside its continued investment in textual scholarship. Moreover, the relaunch also coincided with the development of several collaborations between WFPP and external partners who saw the former's film historiographic information as important datasets for developing new narrative and visual approaches to doing women's film history.[26] Previously, data export had not been an internal topic of conversation but, via these collaborations, the Digital Scholarship team made WFPP's biographical and occupational dataset publicly available in Academic Commons to ensure its preservation and to invite further creative (re)use and analysis.[27] As a result, the back-end occupational taxonomy, for example, whose structure had remained largely invisible to front-end readers because of the way that the WordPress website had initially been designed, was now visible in the exported CSV file. Although a small step toward open data and transparency compared to other digital platforms, this embrace of data (re)use introduced the WFPP editorial team to broader discourses around data standardization and interoperability.

On the one hand, this survey highlights that WFPP has gone through many idiosyncratic iterations, straddling both print and digital conceptualizations. Unlike other digital film and media projects that emerged around the same time, such as the Media History Digital Library and the Media Ecology Project,[28] WFPP's central feature – original text-based scholarship bound by a static citation and concepts like "author," "editor," and "publisher" – reflects its specific print legacy, even as it has

25 See https://wfpp.columbia.edu/projections/, accessed April 1, 2024.

26 See, for example, the "Aesthetics of Access: Visualizing Research Data on Women in Film History" (research group), accessed April 1, 2024, https://www.uni-marburg.de/en/fb09/institutes/media-studies/research/research-projects/davif.

27 Jane M. Gaines and Columbia University Libraries, "Women Film Pioneers Project Biographical Data" (2020), https://doi.org/10.7916/m4dc-n768.

28 See https://mediahistoryproject.org/; https://mediaecology.dartmouth.edu/wp/, both accessed April 1, 2024.

opened itself up more to online data-driven endeavors over time. On the other hand, like its contemporaries, it is a film historical project that is rooted in the archival domain: as a digital humanities film project, it is concerned with not only capturing the ongoing "flood of empirical evidence" of women's involvement in the silent film era, it is also about bringing scholars and students around the world closer to archival materials and relevant research data to facilitate further film historical investigation.

"A Living Resource": Scholarly Editing in the Digital Era

WFPP is also an early digital humanities publishing experiment that developed at Columbia at a time when university libraries were increasingly embracing digital publishing, or the online presentation and dissemination of scholarly content, by creating centers and initiatives to further explore different approaches, methodologies, and services, as well as open-access platforms and innovative editorial tools.[29] In fact, WFPP was one of a few early academic press partnerships that CDRS first developed in 2008.[30] Like these and other projects, WFPP was seen as a collaborative investigation into the different ways that scholarship could be presented online, augmented by archival and visual materials, hypertext, and other online resources. This is not to suggest that WFPP or CDRS was the first of its kind, but rather to situate both within a period – roughly the first decade of the twenty-first century – marked by the increasing presence of collaborative interdisciplinary spaces, focused on digital methods and services for research, dissemination, and teaching, within (American) university libraries.

Framing WFPP as an academic publishing project productively situates it within the broader discourses around scholarly editorial labor in the digital era. First, the advent of web-based publishing has changed how we understand what counts as editing, beyond the critical intervention upon a text prior to publication or the "parsing of the cultural record [i.e., manuscripts] in terms of questions of

29 Diana M. Zorich, "Digital Humanities Centers: Loci For Digital Scholarship," in *Working Together or Apart: Promoting the Next Generation of Digital Scholarship* (Washington, DC: Council on Library and Information Resources, 2009), 70–78, accessed April 1, 2024, https://www.clir.org/wp-content/uploads/sites/6/zorich.pdf.

30 Kennison, email, December 13, 2022. See, for example, Rebecca Kennison, Neni Panourgiá, and Helen Tartar, "*Dangerous Citizens* Online: A Case Study of Author-Press-Library Partnerships," *Serials* 23, no. 2 (2010): 145–149.

authenticity, origin, transmission, or production"[31] in the case of scholarly editions, for example. As Anne Burdick et al. wrote in 2012, the then-emerging institutional field of digital humanities was engendering new editorial practices, which could be understood as "productive and generative" work, or a "suite of rhetorical devices that make a work."[32] "Editing," they explain, "is the creative, imaginative activity of making, and as such, design can also be seen as a kind of editing: It is the means by which an argument takes shape and is given form."[33] Thus, while editing the Volume I manuscript in the early 2000s constituted the more traditional practices of selection, organization, and preparation of texts for publication, the online editorial labor around an individual profile, for example, now also involves manually adding the aforementioned regional and occupational metadata on the back-end, which allows the profile to exist within and relate to a broader argument about the global scope and range of women's creative practice that is built into the design of the website.

In addition, scholarly editing's underlying frameworks and assumptions have also changed since the digital turn. Digital platforms – with their potential for collaborative commentary, interactivity, and open navigation, as well as the ability to present more (multimedia) information than a print book – engender processual thinking over notions of a static, fixed publication. As Susan Brown et al. have written regarding their work on the online Orlando Project, the notion of "done" is now a "fragile" and "negotiated" concept, both in terms of ongoing technical needs and content updates.[34] Indeed, WFPP, like many other digital scholarly projects, was always expected to grow: CDRS promoted it as a "living resource"[35] and the editorial team sought to expand and update the website quickly after the launch, initiating "Phase II" of the project immediately (effectively to embrace "everywhere else" after Phase I's primary emphasis on Latin America, the United States, and Canada). Unlike some projects – the Orlando Project, for instance – that make content additions and updates at certain points, WFPP implemented a rolling publication approach. The editorial team quickly began soliciting new contributors and giving new deadlines to authors who were late with submissions, and, by September 2015, the project had published thirteen new profiles, covering more of the United States,

31 Anne Burdick et al., *Digital_Humanties* (Cambridge, MA: MIT Press, 2016), 18.

32 Burdick et al., *Digital_Humanties*, 18.

33 Burdick et al., *Digital_Humanties*, 18.

34 Susan Brown et al., "Published Yet Never Done: The Tension Between Projection and Completion in Digital Humanities Research," *Digital Humanities Quarterly* 3, no. 2 (2009), accessed April 1, 2024, http://digitalhumanities.org:8081/dhq/vol/3/2/000040/000040.html. See also https://www.artsrn.ualberta.ca/orlando/, accessed April 1, 2024.

35 Newton, Harvell, and Williams, "Women Film Pioneers Project (WFPP)."

Europe, and now Asia, and had a handful of thematic and national overview essays assigned and/or out for peer review.

During this period of expansion, we also embraced the opportunity to update existing content. We began making small-scale modifications to the published scholarship, whether it was responding to new film discoveries (or finally confirming archival holdings) by updating a filmography, fixing broken links, and correcting any typos, or updating profiles with new DVD release information or streaming links. In some cases, minor revisions to the profile text were made in collaboration with the author. It should be noted too that many of these corrections and updates came to our attention via emails sent to the "contact the editors" address on the website, which was used, very soon after our official launch, by a wide variety of users – from scholars and archivists to private collectors, relatives of a filmmaker, and casual readers – who all recognized WFPP's capacity to expand and change over time.

Thus, from its launch onward, editing in regard to WFPP has not only involved the "collecting, selecting and preparing [of] texts for publication"[36] and expanded rhetorical and technical labor. It also immediately presumed a certain processual perspective on the resource as a whole and an awareness of the fragmentary and iterative nature of feminist film historiography more broadly. It was, to borrow Patrick Sahle's description of digital scholarly editions, "an open enterprise,"[37] which required ongoing editorial labor. However, while these early updates and corrections to existing content mentioned above – alongside the constant inclusion of new profiles and essays – were necessary for WFPP to remain a reliable feminist film historical database, they were also problematic: there was no means to be transparent about changes in a profile or essay, and we had no editorial policy in place for major textual updates. In other words, "the inherent changeability" of WFPP's content, no longer limited by a static book, "also pose[d] threats to the scholarly ecosystem"[38] within which the project was embedded. Or, to use terminology from the digital humanities and digital publishing, in the immediate post-launch period, the WFPP editorial team did not engage in any versioning practices.

36 Peter L. Shillingsburg, *Scholarly Editing in the Computer Age: Theory and Practice*, 3rd ed. (Ann Arbor: The University of Michigan Press, 1996), 2.

37 Sahle, "What is a Scholarly Digital Edition?" 29.

38 Paul A. Broyles, "Digital Editions and Version Numbering," *Digital Humanities Quarterly* 14, no. 2 (2020), accessed April 1, 2024, http://www.digitalhumanities.org/dhq/vol/14/2/000455/000455.html.

Versioning, or the iterative creation (as necessary for conceptual or technical reasons) and notation of new versions of something,[39] be it numbering digital scholarly editions or software upgrades, is a central practice in the digital humanities, one that embraces transparent process over final product.[40] In the case of digital scholarship and academic publishing, versioning is a way of maintaining the necessary balance between the ongoing reliability and stability of a scholarly resource and its potential variability. Of course textual versions and variants have a long history in academia that pre-dates the digital era, and new editions and supplementary publications have always been a part of scholarly publishing, albeit at a slower pace still bound to the material limitations of the book. But versioning in the computational era totally upends longstanding academic notions of the stable, completed scholarly work, or, at a technical level, the publication itself, which still dominated the previously mentioned forms of scholarly updates and variation. In the case of digital scholarly publications, as Kathleen Fitzpatrick and Paul A. Broyles have respectively argued, versioning reflects the now necessary commitment, on the part of authors and editors, to both the initial creation and editing processes and the "post-publication maintenance of texts,"[41] as well as demands that we remain "attentive to the way textual resources transform in time."[42]

As Fitzpatrick and Broyles also remind us, versioning requires a systematic editorial (or technical) policy. While WFPP did not have one in place at the time of its launch, it soon became clear that it would be necessary to develop a clear workflow for updates, corrections, and new information.[43] In internal conversations, which were intermittent over three years, the editorial team and our library colleagues discussed the benefits and limits of various editorial and design options to responsibly update published profiles and essays. Ultimately, in 2019, we introduced two new features as part of the website relaunch: the Research Update box, already mentioned, which could be added to any profile as necessary as a dated editorial addendum for new information, and the implementation of DOIs for every profile and essay, linked to a PDF record in Academic Commons, which finally allowed for textual versioning. While the former can be understood

39 This chapter focuses on textual versioning and content updates since technical upgrades on WFPP were handled by Columbia library staff.

40 See, for example, Broyles, "Digital Editions"; Burdick et al., *Digital_Humanities*; Kathleen Fitzpatrick, *Planned Obsolescence: Publishing, Technology and the Future of the Academy* (New York: New York University Press, 2011).

41 Fitzpatrick, *Planned Obsolescence*, 119.

42 Broyles, "Digital Editions."

43 Kate Saccone, "The Women Film Pioneers Project: Two Years Later." Paper presented at Women and the Silent Screen VIII, University of Pittsburgh, 2015.

as a casual additive process and the latter implies a more formal iterative one, both function as methods by which information on WFPP could finally be updated in a more transparent and consistent way.[44]

Digital Curation as a Framework for Iterative Editorial Practice

One can read the implementation of Research Update boxes, DOIs, and versioned records through the lens of shifting conceptualizations of editorial labor in the digital era and the necessity to find ways to balance digital textuality's mutability with scholarly ideals of reliability. But one can also read them as curatorial. By curatorial, I am referring specifically to the practice of digital curation, which, much like versioning, emphasizes process over final product, through different levels of ongoing engagement with digital(ized) materials and data. According to Arjun Sabharwal, digital curation, which emerged as a concept in the early 2000s, is a broad framework for understanding the ongoing preservation and promotion of data, datasets, databases, and digital(ized) materials.[45] Within the archival and cultural heritage domains, digital curation often refers to ongoing administrative, management, and preservation actions taken to ensure long-term and meaningful access to digital(ized) material and data. These practices, shaped by the library and information sciences, follow a lifecycle model, which stresses, for example, different related sequential actions, like creation, appraisal, ingest, preservation, storage, access, reuse, and transformation.[46]

While digital curation as I have just described it may not seem an applicable framework for a text-based publication like WFPP, especially since I have explicitly refrained from using "archive" to define the project, I believe its emphasis on ongoing sequential actions is useful for thinking about the profiles specifically as elements now open to research updates and textual versioning, as well as other modifications over time. In other words, while I do not define WFPP as an archive

44 It should be noted that the Research Update boxes and the versioned records are not perfect editorial features; the type of information and documentation that goes into a box is not standardized across the website, and often the notes about changes in the records remain vague. Yet, these are not necessarily finite editorial features, but steps taken, roughly six years into an evolving online project's life, to deal with updates and textual changes.

45 Arjun Sabharwal, *Digital Curation in the Digital Humanities* (Waltham, MA: Chandos Publishing, 2015), 11.

46 Digital Curation Centre, "Curation Lifecycle Model" (2008), accessed April 1, 2024, https://www.dcc.ac.uk/guidance/curation-lifecycle-model.

in order to emphasize the original scholarship on the platform, there are clearly archival operations at play. I see the individual profiles as modules – following Sahle's point about the fluid modularization of scholarly editions in the digital era[47] – within the larger project, which require ongoing active editorial management in order to maintain their reliability as sources for historiographical film research and archival information. As such, I contend that editorial labor on WFPP, in addition to constituting critical textual selection and preparatory work and other forms of expanded rhetorical labor online, can also be understood through a lifecycle model like digital curation.

I am not the first to make a connection between editing and curating. For example, in 2015, W. B. Worthen brought together scholarly editing (of both print and online texts) with more traditional museological conceptualizations of curation as a form of both collections care and public-facing exhibition making.[48] While he outlines several curatorial dimensions of editing (e.g., editorial labor as a form of "*curing*" a text "of its ills," or the compiling of a themed journal issue via submission or solicitation),[49] this is not his main concern. Rather, he is interested in the shift toward discourses of curation in the digital era itself, which he contends reflects "the changing socialization of academic labor, and perhaps [a] changing sense of what we value about it."[50] On the one hand, he sees the turn toward discourses of curation as indicative of the expanded methods for knowledge production emerging in the digital humanities.[51] On the other hand, he worries that "the stylish veil of *curation*" not only obfuscates editing's "disciplined attention to the formal, rhetorical, contextual, and conceptual presentation[s] of an argument,"[52] but that it also makes this work – which is often time consuming, tedious, and detail-oriented labor – "seem harmlessly irrelevant, a melancholy byproduct of the long, withdrawing roar of print culture."[53] So while digital humanities scholars like Kathleen Fitzpatrick have written with excitement about the ways in which the remix culture of the internet, for example, can contribute to an awareness of curation as a

47 Sahle, "What is a Scholarly Digital Edition?" 36.

48 W. B. Worthen, "Fashion(ing)/Formation," *Contemporary Theory Review* 25, no. 1 (2015): 90–93, https://doi.org/10.1080/10486801.2015.992251. While Worthen looks at the academic editor as curator, others have also looked at the reverse, or a museum curatorial function as editor. See, for example, Dušan Barok, Julia Noordegraaf, and Arjen P. de Vries, "From Collection Management to Content Management in Art Documentation: The Conservator as an Editor," *Studies in Conservation* 64, no. 8 (2019): 472–489, https://doi.org/10.1080/00393630.2019.1603921.

49 Worthen, "Fashion(ing)/Formation," 91, 92–93 (emphasis in original).

50 Worthen, "Fashion(ing)/Formation," 90.

51 Worthen, "Fashion(ing)/Formation," 93.

52 Worthen, "Fashion(ing)/Formation," 92 (emphasis in original).

53 Worthen, "Fashion(ing)/Formation," 93.

valuable scholarly endeavor,[54] Worthen worries that such a perspective could also contribute to a limited view of scholarly textual editorial labor, one in which such work is seen as merely cosmetic final steps. Thus, in using the term digital curation, which also echoes notions of care and stewardship, as a way to understand the ongoing editorial labor on WFPP, this chapter offers a way to engage with both Fitzpatrick's excitement and Worthen's hesitancy. It not only provides a framework through which to consider forms of curatorial labor in the digital era as valuable scholarly (in the case of WFPP, specifically film historiographic) work; it also allows for a continued emphasis on iterative editorial labor – from textual to online design and continued updates – as time consuming and detail-oriented disciplinary work in support of authors and their research.

In this section, I will not transpose the exact stages of the digital curation lifecycle model, first developed by the Digital Curation Centre (DCC) in 2008,[55] to WFPP, since a one-to-one comparison is impossible given the latter's specificities – preservation actions for the profiles were not part of the editorial workflow until 2019, and technical processes of disposal and migration, occasional actions in the standard model, are not central to the editorial labor on WFPP. Rather, I am inspired by the framework that such a concept and practice provoke from an editorial perspective: the emphasis on sequential actions over time and an awareness that digital environments require constant management (curation and preservation) of information. Thus, loosely adapting the standard DCC model, I have created a simplified high-level graphical overview that outlines the sequential actions that comprise the current editing and publishing workflows for the WFPP profiles. This editorial lifecycle not only functions as a useful way to break down ongoing editorial labor into finite stages as part of a broader process; it is also particularly germane to WFPP, due to both its subject matter and the fact that long-term web hosting and financial resources are currently not pressing challenges (Figure 2).

At the center of the model is the WFPP profile, the digital textual entity that is assigned (selected by the editorial team), submitted, and eventually published on the platform. Around it, are the full lifecycle actions – curation and preservation – which, taken together, constitute the ongoing management of these scholarly entities over time. These ongoing actions can be broken up into specific sequential actions, which I loosely categorize as: conceptualization, creation, preservation, access and promotion, assessment, and (potential) change for updates.

54 Fitzpatrick, *Planned Obsolescence*, 79.

55 The DCC model has since been used, adapted, and discussed at length within the digital humanities and information sciences. For a detailed critical discussion of the DCC model, see Hea Lim Rhee, "A New Lifecycle Model Enabling Optimal Digital Curation," *Journal of Librarianship and Information Science* (2022), https://doi.org/10.1177/09610006221125956.

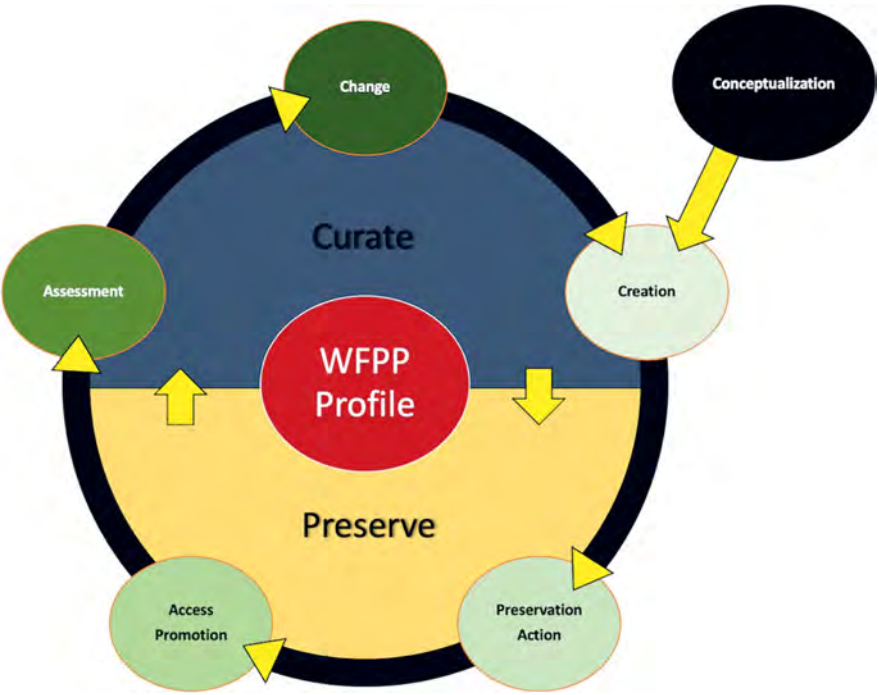


Figure 2: The WFPP editorial lifecycle model for profiles. Adapted from the Digital Curation Centre's "Curation Lifecycle Model," https://www.dcc.ac.uk/guidance/curation-lifecycle-model (accessed April 1, 2024).

In this model, conceptualization of the profiles covers the assignment and editorial review stages, which involve intellectual, administrative, and legal activities. Some profiles are assigned after a potential contributor has reached out to the editors and pitched a profile – often for a woman featured on our "unhistoricized" list[56] – but many are assigned through standard editorial commissions. For example, the editorial team stays abreast of new archival film discoveries, restorations, and scholarly publications that relate to women's film history, and reaches out, either via email, or in person at conferences or film festivals, to scholars or archivists who could contribute a given profile. Following the submission of the initial draft (Microsoft Word) – which can take anywhere from six months to many years after assignment – a rigorous review takes place, involving

56 See https://wfpp.columbia.edu/resources/unhistoricized-women-film-pioneers/, accessed April 1, 2024.

editorial feedback and suggestions at both conceptual and textual levels. This review stage usually involves several rounds of editorial feedback and revisions and often further research, and can take anywhere between a few months and several years.

Throughout this phase, both the author and the editorial team are working toward finalizing the profile as a specific scholarly entity in a format and scope structured by certain agreements. In addition to the style guidelines, which also urge contributors to be self-reflective about their research,[57] this stage also includes handling any rights clearance for images. Moreover, at the time of the first submission, authors must now sign a standard agreement stipulating the (re)use policy and various stakeholder rights that will go into effect should the article be published.[58] For example, the author agreement clarifies that the author retains copyright of the profile and that all content on WFPP is published under a CC BY "Attribution" license. Thus, throughout this stage, processes of selection, administrative and legal practicalities, and collaborative critical and intellectual labor constitute the conceptualization of the WFPP profile more broadly.

Once a profile text is finalized and all relevant accompanying images have been submitted by the contributor, the profile page is created on the website. The text is then placed on the back-end of WordPress, using the customized template originally designed by CDRS. The biographical and occupational metadata is also added, thus integrating the profile into the database structure. The images, as high resolution JPEG files, are also added to WordPress' back-end media library, with caption and rights information, and then placed on the profile page. Further copyediting and the adding of relevant hyperlinks also occurs during this creation period.

Once the author has previewed and signed off on a final version of the profile online, the first preservation actions are taken: a staff member in the library assigns the page a DOI and deposits a PDF version in Academic Commons, after which the profile can officially be published (made visible on the profiles landing page), where it can now be internally tracked via Google Analytics and regularly archived by library staff via the Wayback Machine. There is also some necessary promotional work at this stage, from adding the new profile to the "featured pioneer" space on the WFPP homepage to signaling the new addition on social media and in our quarterly newsletter.

These different editorial and publishing stages – from conceptualization and creation to preservation, and access and promotion – constitute important sequential actions to present and preserve the WFPP profile online. As this brief

57 See https://wfpp.columbia.edu/guidelines-profiles/, accessed April 1, 2024.

58 This policy has been in effect for all new assignments since 2019.

summary also highlights, the editorial labor on WFPP comprises both the standard selection, organization, and preparation of texts and also administrative and technical labor as part of finalizing the work for online publication. It also signals that when a profile is published on WFPP it is not a work-in–progress, but the finalized output of a contributor's scholarly and archival research. But at the same time, as I have already discussed, the project has always been open to that work progressing as needed in order to reflect the iterative and fragmentary nature of feminist film research and archival investigation more broadly. With the implementation of the Research Update boxes, the DOIs, and the versioned records in Academic Commons, this progression, as part of the profile's lifecycle as it were, can now be understood as key phases in more concrete terms.

As part of being freely accessible online entities, the profiles are open to assessment, from readers, contributors, and members of the editorial team, who all recognize them as unfixed modules that can change over time in response to new archival discoveries, for example, or the recognition of previous inaccuracies. Assessment can therefore lead to meaningful change within the profile page, in terms of both research updates and textual versioning and at the level of hyperlinks and metadata. In all cases, these updates lead to new phases of content creation on the profile page, albeit mostly minor to date, as new text and images are ingested, documented, preserved, and made accessible for further ongoing assessment.

As previously mentioned, the editorial team currently uses the Research Update boxes to present new information separate from the profile. As evidenced by the Beatriz Michelena update in my introduction, these editorial addenda act as informal spaces to note new questions, research discoveries, and changes that do not require textual rewrites. Since 2019, I have added a Research Update to a handful of other profiles, including the French actress, director, producer, and archivist Musidora. Published in 2013, her profile could not account for the (future) fact that a fragment of VICENTA (1919), which she directed, was discovered in 2016. I have since updated the filmography and created a Research Update box to highlight this change, sending an updated PDF of the profile to the library for preservation in Academic Commons (where it now exists with the following record note: "A research update box was added to the profile, and the PDF associated with this record was updated, in January 2023").[59]

The implemented DOIs, on the other hand, allow for more formal and extensive textual revisions within the profiles. For example, in 2016, we published a

59 Annette Förster, "Musidora," in *The Women Film Pioneers Project*, ed. Jane M. Gaines, Radha Vatsal, and Monica Dall'Asta (New York: Columbia University Libraries, 2013), https://doi.org/10.7916/d8-f4nt-4j92.

profile on May Watkis, a Canadian woman who was believed to have been the administrative head of a provincial government film agency. As I have written elsewhere, in 2018 previously unseen archival materials were brought to the editorial team's attention, which challenged Watkis' presumed position of director of this organization.[60] When the completely revised profile was published in 2020, it now reflected Watkis' more complicated – and still unclear – position in the early cinematographic field as a clerk at the agency, a projectionist for a local film censor, and a theater inspector.[61] At the top of the new profile, which was assigned a new DOI, there is a brief note indicating that the original version can be accessed at the previous DOI, which links to the archived PDF in Academic Commons and allows for comparison if necessary (the new Academic Commons record also links to the previous version) (Figure 3).

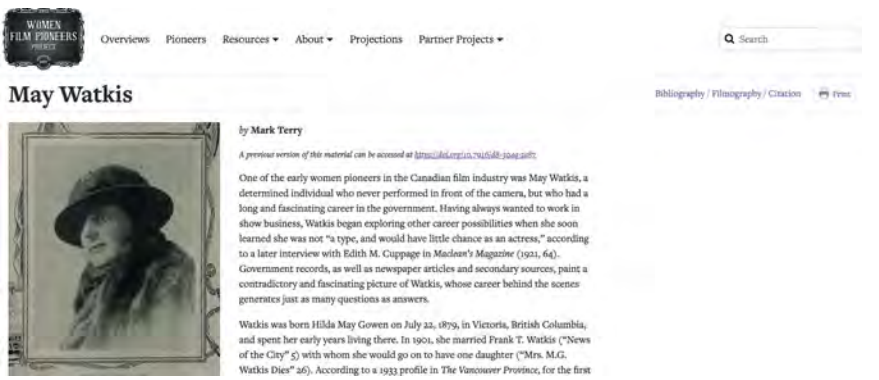


Figure 3: The revised WFPP May Watkis profile with a new DOI, screenshot.

These examples of the Research Update and textual versioning – two different types of website updates that are both connected to stable records in Academic Commons – can thus be understood as part of the ongoing active and critical editorial

60 Kate Saccone, "Digital (Re)Visions: May Watkis and the Women Film Pioneers Project," *Modernism/modernity* 5, cycle 5 (August 17, 2020), accessed April 1, 2024, https://modernismmodernity.org/forums/posts/saccone-digital-revisions.

61 Mark Terry, "May Watkis," in *The Women Film Pioneers Project*, ed. Jane M. Gaines, Monica Dall'Asta, and Radha Vatsal (New York: Columbia University Libraries, 2020), https://doi.org/10.7916/d8-4ac7-fq81.

management of the profiles.[62] These editorial features not only allow the texts "to live and breathe,"[63] to quote Fitzpatrick, in ways that are only possible in the so-called digital paradigm; loosely framed by the concept of digital curation, research updates and textual versioning also become a potential part of a broader editorial lifecycle comprising a wide range of processes, or sequential actions of creation, preservation, and the ongoing assessment of the profiles over time. While I have not (yet) had to add a second new DOI to a profile that already contains some textual versioning, I have returned to some Research Update boxes several times.[64]

Smaller changes can also occur on a profile once it has been made accessible online and open to ongoing assessment. For example, in the case of linked open data practices, the challenge of linking to consistent resources and maintaining reliable connections on the profiles has been a long-term editorial project. At the time of our launch in 2013, linked open data was only briefly discussed internally; the focus when designing and finalizing the WordPress website was to link as much as possible within the website – to other profiles, essays, and resources – rather than connect with outside ones.[65] At the time of the website relaunch in 2019, the editorial team made a conscious effort to better connect the profiles to external archival resources, and I manually went through every profile and linked each archival paper collection to a relevant collection page, online finding aid, or general archive website. (A similar task of linking to the archives in the filmographies was done automatically.) As more institutional resources are digitized and described online, these links also require continual management to avoid link rot or to provide better, more reliable connections. Similarly, the ongoing management of the biographical metadata has become a central part of this iterative editorial workflow as I have, as a result of the ongoing mass digitization of archival documents and historical records, had to update many women's dates of birth or death, for example. Making these technical or metadata changes to the profiles – always with an update to the Academic Commons record as well – not unlike the research updates and textual versioning, can thus be read as part of

62 It should be noted that these PDFs are "reduced" versions of the profiles, without the moving or still images. This runs counter to traditional preservation practices with digital(ized) media where what is preserved is often the high quality, uncompressed version, while what is presented to the public is a compressed, lossy version.

63 Fitzpatrick, *Planned Obsolescence*, 69.

64 See, for example, Angel Miquel, "Mimí Derba," in *The Women Film Pioneers Project*, ed. Jane M. Gaines, Radha Vatsal, and Monica Dall'Asta (New York: Columbia University Libraries, 2013), https://doi.org/10.7916/d8-pw36-1k24.

65 There were early efforts, however, on the part of CDRS, to add links *to* WFPP on Wikipedia, a process that was, unfortunately, undone when the website relaunched in 2019 with a new URL ("cdrs" was removed by the library since the center no longer existed).

the ongoing management and preservation of the profiles to ensure their reliability as scholarly and archival resources over time.

The concept of digital curation, again taken as a loose inspiration here, underscores how editorial labor on WFPP – in addition to being textual support – can be understood as a form of ongoing active management, in terms of textual scholarship, archival information, and data. In other words, if editing in the digital era can be understood as the "suite of rhetorical devices that make a work," to recall Burdick et al., the WFPP editorial team's active and critical management, its digital curation of the profiles – as modules within the larger project that are never quite "done" – supports the project's broader argument about the iterative nature of women's film history. Of course, not every profile has required changes, but that does not mean that they do not exist within this broader framework, as entities that can be changed, at varying levels, when necessary.

(Re)Visioning, or Toward a Critical-Feminist Digital Curatorial-Editorial Practice

In describing WFPP's digital editorial labor as curatorial – where creation and (the potential for) versioning, for example, exist within a lifecycle model – I am interested in the relationship between scholarly editorial labor and notions of mutability and historiographic movement. While a concept like "updatism," for instance, certainly signals digital scholarship's move away from the one-text paradigm of print toward multiplicity and open, iterative online editorial labor, it does not explicitly highlight this tension between fluidity and reliability or, in the context of WFPP, the continuous revisiting of a profile over time as new information must be made visible and new interventions are warranted. I therefore advance the concept of "(re)visioning" instead, which plays on the practices of versioning, as well as on ideas of scholarly revisions and research updates, the iterative nature of feminist film historiography, and the ongoing management of digital environments over time. My conceptualization of "(re)visioning," shaped by WFPP's specific historiographical origins, ongoing institutional funding, and online editorial practices, such as its embrace of a rolling editorial and publishing approach, is thus inherently feminist, reflecting the project's longstanding emphasis on, to quote Sarah-Mai Dang, the "transformative nature of knowledge produc-

tion."[66] Similarly, paralleling the field of data feminism, which argues that labor must be visible,[67] I use "(re)visioning" to highlight the work itself, to turn attention to what is being done on platforms like WFPP at the editorial level and how that work can be framed.

In the way that "(re)visioning" reflects and emerges from WFPP's history and subject matter, one could argue that it is a very limited concept, demonstrating only how the "inexhaustible"[68] nature of feminist (film) historiography and (silent era) archival research finds a productive match in the mutability of digital scholarship and online publishing. But while other feminist websites, such as the Importing Asta Nielsen Database and the Nordic Women in Film platform, are certainly inscribed within this framework, I contend that "(re)visioning" does not only concern feminist projects. In fact, I use "(re)visioning" to encourage the adoption of a critical-feminist approach to digital film historiography more broadly. In addition to centralizing both the iterative and fragmented historiographical processes and the curatorial-editorial work that manages, maintains, and expands the digital resources we develop and use, "(re)visioning" is conceived as part of the broader "interrogation of notions of 'completeness' and 'done'" in the domain of scholarly publishing in the digital era.[69] In other words, the updates (research and metadata) and examples of textual versioning that I have discussed do not reflect, in the words of Joanne Tucker, research "failure[s] which ought to be quietly rectified."[70] Rather, to make (re)visions is a critical-feminist act, a reminder that most digital scholarly projects remain "undone," both in terms of (modularized) content updates and platform-wide enhancements over time. Most importantly, it is a celebration of our collective generative and iterative labor.

Conclusion

Ultimately, my notion of (re)visioning – a concept that frames editorial labor as a form of digital curation encompassing, in the case of WFPP, practices of creation,

66 Sarah-Mai Dang, "Unknowable Facts and Digital Databases: Reflections on the Women Film Pioneers Project and Women in Film History," *Digital Humanities Quarterly* 14, no. 4 (2020), accessed April 1, 2024, http://digitalhumanities.org/dhq/vol/14/4/000528/000528.html.

67 See Catherine D'Ignazio and Lauren F. Klein, *Data Feminism* (Cambridge, MA: MIT Press, 2020), accessed April 1, 2024, https://data-feminism.mitpress.mit.edu/.

68 Brown et al., "Published Yet Never Done."

69 Brown et al., "Published Yet Never Done."

70 Joanna Tucker, "Facing the Challenge of Digital Sustainability as Humanities Researchers," *Journal of the British Academy* 10 (2022): 104, https://doi.org/10.5871/jba/010.093.

assessment, and making various updates – is intended to continue the conversation about what it means to present, disseminate, and preserve scholarship online, especially when that scholarship, and the archival research sustaining it, is neither finite nor complete. What a project like WFPP offers to the broader field of digital film historiography is an example of how editorial labor has expanded and transformed. Most importantly, it models how a critical-feminist framework for doing digital film history can continue to activate an iterative and reflective curatorial-editorial practice.

While this chapter focused on manual editorial labor around the individual profiles, the idea of digital curation as a framework for understanding the editorial work on the website is open to further development. It remains to be seen, for example, how the external (re)use of the WFPP biographical and occupational dataset will feed back into the profiles and the information they present, and what (other) curatorial actions further efforts around linked open data could engender on the website. The value of digital curation as a framework for positioning the research updates and textual versioning as part of the profiles' editorial lifecycle lies not only in how it highlights expanded scholarly editorial labor online, but also in the fluid, flexible perspective of the project as a whole, where further actions – such as thinking more concretely about long-term preservation and digital sustainability – could still be incorporated into editorial workflows.

October 2023 marked the ten-year anniversary of WFPP's online launch. As I have shown, this past decade has been a time of steady growth, both in terms of content and internal practices. Looking ahead, in addition to encouraging scholars to explore the dataset, I hope the project will continue to expand, adding new profiles and overview essays as well as Projections posts that engage with digital methods and data visualization approaches. This may require rethinking the technological limitations of the WordPress website, in collaboration with our library colleagues, and finding ways to deal with larger media files and different applications and tools. I also hope that the editorial team can revisit aspects of the project, from critically returning to our taxonomy of occupations, with an eye toward more standardization across the website and external controlled vocabularies, to thinking more consciously about linked open data and documentation. If the project's history can anticipate its future, there has always been room to expand and transform, to consider and develop slightly different versions of this editorial project in response to both shifts in scholarly dissemination practices and the iterative nature of feminist film historiography. WFPP has come a long way since Jane M. Gaines began collecting names – and there is always more to do.

References

"About Us." *Importing Asta Nielsen*. Accessed April 1, 2024. https://importing-asta-nielsen.online.uni-marburg.de/.

"Aesthetics of Access (DAVIF)." Accessed April 1, 2024. https://www.uni-marburg.de/en/fb09/institutes/media-studies/research/research-projects/davif.

Anderson, Steve F. *Technologies of History: Visual Media and the Eccentricity of the Past*. Hanover: Dartmouth University Press, 2011.

Barok, Dušan, Julia Noordegraaf, and Arjen P. de Vries. "From Collection Management to Content Management in Art Documentation: The Conservator as an Editor." *Studies in Conservation* 64, no. 8 (2019): 472–489. https://doi.org/10.1080/00393630.2019.1603921.

Brown, Susan, Patricia Clements, Isobel Grundy, Stan Ruecker, Jeffry Antoniuk, and Sharon Balazs. "Published Yet Never Done: The Tension Between Projection and Completion in Digital Humanities Research." *Digital Humanities Quarterly* 3, no. 2 (2009). Accessed April 1, 2024. http://digitalhumanities.org:8081/dhq/vol/3/2/000040/000040.html.

Broyles, Paul A. "Digital Editions and Version Numbering." *Digital Humanities Quarterly* 14, no. 2 (2020). Accessed April 1, 2024. http://www.digitalhumanities.org/dhq/vol/14/2/000455/000455.html.

Burdick, Anne, Johanna Drucker, Peter Lunenfeld, Todd Presner, and Jeffrey Schnapp, *Digital_Humanties*. Cambridge, MA: MIT Press, 2016.

Clavert, Frédéric, and Andreas Fickers. "Publishing Digital History Scholarship in the Era of Updatism." *Journal of Digital History* (2022). Accessed April 1, 2024. https://journalofdigitalhistory.org/en/article/m7DWqDjY3hoV.

Dall'Asta, Monica, and Jane M. Gaines. "Prologue: Constellations: Past Meets Present in Feminist Film History." In *Doing Women's Film History: Reframing Cinemas. Past and Future*, edited by Christine Gledhill and Julia Knight, 13–26. Urbana: University of Illinois Press, 2015.

Dang, Sarah-Mai. "Unknowable Facts and Digital Databases: Reflections on the Women Film Pioneers Project and Women in Film History." *Digital Humanities Quarterly* 14, no. 4 (2020). Accessed April 1, 2024. http://digitalhumanities.org/dhq/vol/14/4/000528/000528.html.

Digital Curation Centre. "Curation Lifecycle Model" (2008). Accessed April 1, 2024. https://www.dcc.ac.uk/guidance/curation-lifecycle-model.

D'Ignazio, Catherine, and Lauren F. Klein. *Data Feminism* (Cambridge, MA: MIT Press, 2020). Accessed April 1, 2024. https://data-feminism.mitpress.mit.edu/.

Fitzpatrick, Kathleen. *Planned Obsolescence: Publishing, Technology and the Future of the Academy*. New York: New York University Press, 2011.

Förster, Annette. "Musidora." In *The Women Film Pioneers Project*, edited by Jane M. Gaines, Radha Vatsal, and Monica Dall'Asta. New York: Columbia University Libraries, 2013. https://doi.org/10.7916/d8-f4nt-4j92.

Gaines, Jane M. "On Not Narrating the History of Feminism and Film." *Feminist Media Histories* 2, no. 1 (2016): 6–31.

Gaines, Jane M., and Columbia University Libraries. "Women Film Pioneers Project Biographical Data" (2020). https://doi.org/10.7916/m4dc-n768.

Kennison, Rebecca, Neni Panourgiá, and Helen Tartar. "*Dangerous Citizens* Online: A Case Study of Author-Press-Library Partnerships." *Serials* 23, no. 2 (2010): 145–149.

Lyons, MaryAnne. "Beatriz Michelena." In *The Women Film Pioneers Project*, edited by Jane M. Gaines, Radha Vatsal, and Monica Dall'Asta. New York: Columbia University Libraries, 2013. https://doi.org/10.7916/d8-76bd-6466.

Miquel, Angel. "Mimí Derba." In *The Women Film Pioneers Project*, edited by Jane M. Gaines, Radha Vatsal, and Monica Dall'Asta. New York: Columbia University Libraries, 2013. https://doi.org/10.7916/d8-pw36-1k24.

Newton, Mark P., Jackson Harvell, and Leyla S. Williams. "Women Film Pioneers Project (WFPP): Presentation at Coalition for Networked Information, Fall Forum 2013" (2013). https://doi.org/10.7916/D8GB223R.

Rhee, Hea Lim. "A New Lifecycle Model Enabling Optimal Digital Curation." *Journal of Librarianship and Information Science* (2022). https://doi.org/10.1177/09610006221125956.

Rich, Adrienne. "When We Dead Awaken: Writing as Re-Vision." *College English* 34, no. 1 (1972): 18–30.

Sabharwal, Arjun. *Digital Curation in the Digital Humanities*. Waltham, MA: Chandos Publishing, 2015.

Saccone, Kate. "Digital (Re)Visions: May Watkis and the Women Film Pioneers Project." *Modernism/modernity* 5, cycle 5 (August 17, 2020). Accessed April 1, 2024. https://modernismmodernity.org/forums/posts/saccone-digital-revisions.

Saccone, Kate. "The Women Film Pioneers Project: Two Years Later." Paper presented at Women and the Silent Screen VIII conference, University of Pittsburgh, 2015.

Sahle, Patrick. "What is a Scholarly Digital Edition?" In *Digital Scholarly Editing: Theories and Practices*, edited by Matthew James Driscoll and Elena Pierazzo, 19–40. Cambridge: Open Book Publishers, 2016.

Shillingsburg, Peter L. *Scholarly Editing in the Computer Age: Theory and Practice*, 3rd ed. Ann Arbor: The University of Michigan Press, 1996.

Terry, Mark. "May Watkis." In *The Women Film Pioneers Project*, edited by Jane M. Gaines, Radha Vatsal, and Monica Dall'Asta. New York: Columbia University Libraries, 2020 https://doi.org/10.7916/d8-4ac7-fq81.

Tucker, Joanna. "Facing the Challenge of Digital Sustainability as Humanities Researchers." *Journal of the British Academy* 10 (2022): 93–120. https://doi.org/10.5871/jba/010.093.

"Women Film Pioneers Project Launched at the Museum of Modern Art, New York City!" *The Women Film Pioneers Project*, November 25, 2021. Accessed April 1, 2024. https://wfpp.columbia.edu/2013/09/25/10192013-women-film-pioneers-project-at-the-museum-of-modern-art-new-york-city/.

"The Women Film Pioneers Project Relaunches!" *The Women Film Pioneers Project*, October 18, 2019. Accessed April 1, 2024. https://wfpp.columbia.edu/2019/10/18/the-women-film-pioneers-project-relaunches/.

Worthen, W. B. "Fashion(ing)/Formation." *Contemporary Theory Review* 25, no. 1 (2015): 90–93 https://doi.org/10.1080/10486801.2015.992251.

Zorich, Diana M. "Digital Humanities Centers: Loci For Digital Scholarship." In *Working Together or Apart: Promoting the Next Generation of Digital Scholarship*, 70–78. Washington, DC: Council on Library and Information Resources, 2009. Accessed April 1, 2024. https://www.clir.org/wp-content/uploads/sites/6/zorich.pdf.

Alexandra Schneider and Yvonne Zimmermann

Data Cleaning and Diversity in Digital Film Historiography

Introduction

This chapter engages with what we would like to call the "Drucker-Dilemma" – a dilemma that resonates with the challenges that confront us within our own digital projects. In *The Digital Humanities Coursebook* from 2021, Johanna Drucker writes:

> Digital humanities projects often begin with strong convictions about the interpretative requirements of their study and the need for highly customized data and metadata [. . .]. But when the time for interoperability or cross-project functionality arises, the intellectual investment in elaborate data constructions is often jettisoned, sacrificed to expediency.[1]

The Drucker-Dilemma results from the very necessity to make data machine-readable. The first step of doing so is usually described as the phase of data preparation or data cleaning. While we are not the first to stumble upon the notion of "cleaning" in this context, let us remind ourselves: the cleaning sector is still highly segregated – racially, and based on gender and class. Cleaning is often underpaid or not paid at all, and cleaning is more often than not invisible labor, regardless of whether at home or when it comes to data. At the same time, cleaning is the precondition for the production of surplus or added value. Yet "data cleaning" is the least discussed type of operation in our working with data, even though it takes up 80% to 90% of the research process.[2] It is labor made invisible by our not speaking about it. And as Rashida Richardson, Jason Schultz, and Kate Crawford argued in 2019, the data cleaning discourse also obfuscates the "real dirt" in data practices. In "Dirty Data, Bad Predictions," they show that, although data providers were willing to acknowledge biased data as a problem in law enforcement and jurisdiction, they would only try "to isolate or segregate it from what is presumably 'clean' data instead of seeing it as an indicator of the poten-

1 Johanna Drucker, *The Digital Humanities Coursebook: An Introduction to Digital Methods for Research and Scholarship*, 1st ed. (Abingdon, Oxon; New York: Routledge, 2021), 31.

2 See, for example, Ronald K. Pearson, *Mathematics in Industry: Mining Imperfect Data: With Examples in R and Python*, 2nd ed. (Philadelphia: Siam, 2020), https://doi.org/10.1137/1.9781611976274; and Andy Oram, *Managing the Data Lake* (Sebastopol: O'Reilly Media, 2015), accessed March 21, 2024. https://learning.oreilly.com/library/view/managing-the-data/9781492049876/.

https://doi.org/10.1515/9783111082486-009

tial unreliability of the entire dataset from that jurisdiction."[3] Or as an indicator of injustice, as Catherine D'Ignazio and Lauren F. Klein argue in their 2020 book *Data Feminism*.[4]

In the following, we invoke some of the tropes of the cleaning discourse in Data Science and Digital Humanities before we take up Katie Rawson and Trevor Muñoz' thought-provoking 2019 article "Against Cleaning" to think about how their intervention can be made productive for our own current film historical projects that engage with computational methods.[5]

"Cleaning" as a Problem

In his 2020 book *Mathematics in Industry: Mining Imperfect Data*, Ronald K. Pearson holds that data preparation includes "reconciling and merging data from different sources, identifying and interpreting various data anomalies, and selecting and implementing appropriate treatment strategies for the anomalies that are found."[6] His book is devoted to "the identification and treatment of data anomalies, including examples that highlight different types of anomalies, their potential consequences if left undetected and untreated, and options for dealing with them."[7] Because data may be missing, entered in diverse formats, contain errors, or be lost or corrupted, others in the field, such as Andy Oram, speak of "data that does not conform."[8] Within industrial and applied mathematics, such "data that does not conform," or "data anomalies," need to be "treated," that is erased for the sake of economization.

In the humanities, data cleaning is often described as data transformation. Lev Manovich in his 2020 book *Cultural Analytics*, for example, holds that computational analysis of cultural data means to "[. . .] choose a single format and translate all dates into this format."[9] Christof Schöch in his 2013 essay on data in the

3 Rashida Richardson, Jason Schultz, and Kate Crawford, "Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice," *94 N.Y.U. L. REV. ONLINE* 192 (2019): 199–200, accessed March 21, 2024, https://ssrn.com/abstract=3333423.

4 Catherine D'Ignazio and Lauren F. Klein, *Data Feminism* (Cambridge, MA: MIT Press, 2020).

5 Katie Rawson and Trever Muñoz, "Against Cleaning," in *Debates in the Digital Humanities 2019*, ed. Matthew K. Gold and Lauren F. Klein (Minneapolis: University of Minnesota Press, 2019), 279–292, https://doi.org/10.5749/j.ctvg251hk.26.

6 Pearson, *Mathematics in Industry*, n.p.

7 Pearson, *Mathematics in Industry*, n.p.

8 Oram, *Managing the Data Lake*, n.p.

9 Lev Manovich, *Cultural Analytics* (Cambridge, MA: MIT Press, 2020), 130.

humanities, to take another example, considers data to be "clean" and hence "smart," when "imperfections of the process of capture or creation have been reduced as much as possible."[10] Data cleaning is understood here as a process of standardizing data in which data is improved and "up-smarted." But then, to stick to Manovich's metaphor of translation, if translation is not a neutral process of de- and recoding, what happens with data that gets translated in the process of cleaning.[11] Or, in other words, what is lost in translation?

Against Cleaning

In the last few years, several pertinent publications in critical data studies have appeared, most importantly from the perspective of Critical Race and Black Studies and intersectional feminism. Examples are Safiya Umoja Noble's *Algorithms of Oppression: How Search Engines Reinforce Racism* from 2018, Ruha Benjamin's *Race after Technology: Abolitionist Tools for the New Jim Code* from 2019, and the already mentioned *Data Feminism* by D'Ignazio and Klein, to name just a few central books.[12] As Markus Stauff, Pauline van Romondt Vis, and Karin van Es have recently argued: "After all, the impression that data are universal and the enticing power of 'big data' only result from the 'cleaning' of data that eradicates their local embeddedness and heterogeneity."[13] In his 2019 book *All Data Are Local: Thinking Critically in a Data-Driven Society*, Yanni Alexander Loukissas suggested how to counter the "myth of digital universalism" by thinking data as local.[14]

On a larger scale, this resonates with one of the seven core principles of data feminism that Catherine D'Ignazio and Lauren Klein formulate in their book,

10 Christof Schöch, "Big? Smart? Clean? Messy? Data in the Humanities," *Journal of the Digital Humanities* 2, no. 3 (2013): n.p., accessed March 21, 2024, https://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/.

11 See Dilek Dizdar, "Translation und Grenze. Versuch einer translationswissenschaftlichen Neufiguration," in *Übersetzung*, ed. N. Engel and S. Köngeter (Wiesbaden: Springer VS, 2020), 57–74.

12 See Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York: NYU Press, 2018); Ruha Benjamin, "Race After Technology: Abolitionist Tools for the New Jim Code," *Social Forces* 98, no. 4 (2020): 1–3, https://doi.org/10.1093/sf/soz162; D'Ignazio and Klein, *Data Feminism*.

13 Markus Stauff, Pauline van Romondt Vis, and Karin van Es, "Coffee Roasters' Data Vernacular: On the Entanglement of Digital Data and Craft," in *Situating Data: Inquiries in Algorithmic Culture*, ed. Karin van Es and Nanna Verhoeff (Amsterdam: Amsterdam University Press, 2023), 32–33.

14 Yanni Alexander Loukissas, *All Data Are Local: Thinking Critically in a Data-Driven Society* (Cambridge, MA: MIT Press, 2019).

namely to *embrace pluralism*: "Data feminism insists that the most complete knowledge comes from synthesizing multiple perspectives, with priority given to local, Indigenous, and experiential ways of knowledge."[15] Based on Loukissas and others, Karin van Es and Nana Verhoeff suggest in their 2023 critical data studies reader, *Situating Data: Inquiries in Algorithmic Culture*, that we understand "[s]ituating data as cultural inquiry," which implies not only localizing "data both in and as culture," but also situating "our perspective on, and knowledge about, this culture."[16] This double culturality of data is an important concept for critical data studies.

Published in the same year as Loukissas' book, the article by Rawson and Muñoz, "Against Cleaning," resonates with the above-mentioned critical understanding of data cultures, or with what they describe as "an unresolved conversation about data and reductiveness in the humanities." Their contribution to critical data studies intends to promote "the development of new discourses and the practice of critically attuned data work."[17]

Though not explicitly, Rawson and Muñoz stumbled upon the Drucker-Dilemma in their project "Curating Menus," a research project that intended "to curate and analyze the open data from New York Public Library's What's on the Menu?":[18] either the dataset is not really processable *or* it gets normalized to become machine-readable. As Drucker reminds us, cultural data resists the requirements of data science's understanding of good data in the sense of being "valid, accurate, complete, consistent, and uniform."[19] Or, in the words of Rawson and Muñoz: "What became evident was that cleaning up or correcting values was a misleading – and even unproductive – way to think about how to make the data more useful for our own questions [. . .]."[20] One possible alternative to preserve diversity in their dataset (or data setting to use Loukissas' suggestion) was to engage with anthropologist Anna Lowenhaupt Tsing's critique of scalability. As Rawson and Muñoz point out, Tsing reminds us that "scalable projects are articulations between scalable and nonscalable elements, in which nonscalable effects can be hidden."[21] In relation to their own dataset, Rawson and Muñoz argue that non-scalable elements should not be ignored because they can impact our understanding and handling of the scalable

15 D'Ignazio and Klein, *Data Feminism*, 22.

16 Karin van Es and Nanna Verhoeff, eds., *Situating Data: Inquiries in Algorithmic Culture* (Amsterdam: Amsterdam University Press, 2023), 13, https//:doi 10.5117/9789463722971_intro.

17 Rawson and Muñoz, "Against Cleaning," 281.

18 Rawson and Muñoz, "Against Cleaning," 281.

19 Drucker, *The Digital Humanities Coursebook*, 28, 23.

20 Rawson and Muñoz, "Against Cleaning," 281.

21 Rawson and Muñoz, "Against Cleaning," 284.

data.[22] Hence, nonscalability theory, according to Rawson and Muñoz, "encourages us to grapple with this dynamic at each point of articulation in the process of making scalable objects."[23]

To make scalability explicit, they suggest working with what they call "indexing" instead of cleaning.[24] From the many possible meanings and uses of the word in the Oxford English Dictionary, "index" can be a synonym for the forefinger, as it is used for pointing. But an index can also be the material object in scientific instruments that is used as a pointer that moves on a "graduated scale," indicating movements or measurements. In this understanding an index is not only a "sign, token, or indication *of* something" (OED) but "an information structure designed to serve as a system of pointers between two bodies of information, one of which is organized to provide access to concepts in the other," as Rawson and Muñoz point out in their understanding of an index.[25] By indexing we do create scalability, but it is a scalability that is consciously made, always resonates with the unscalable, and keeps traces of the locality of data. Via indexing, data retain their locality as they only become machine-readable with a reference to the specific context in which they were created. They are not just cleaned or filtered but, ideally, the mistakes are kept, for they are relevant as traces.[26]

While thought-provoking on a conceptual level, we want to consider how this idea of indexing can be translated into practice in the following discussion, taking our own data-driven research projects as examples.

22 Rawson and Muñoz, "Against Cleaning," 285.
23 Rawson and Muñoz, "Against Cleaning," 284.
24 Rawson and Muñoz, "Against Cleaning," 288.
25 Rawson and Muñoz, "Against Cleaning," 288.
26 Jasmijn Van Gorp, "Interstitial Data: Tracing Metadata in Archival Search Systems," in *Situating Data: Inquiries in Algorithmic Culture*, ed. Karin van Es and Nanna Verhoeff (Amsterdam: Amsterdam University Press, 2023), 212.

Beyond Cleaning: How to Make Rawson and Muñoz' Proposition Productive for Our Own Projects?

The Straschek Collection and "Mapping German Film Migration 1930–1950"

Our first example is based on the collaborative research project "Mapping German Film Migration 1930–1950" by Imme Klages, Saeideh Safat Zadeh, and Alexandra Schneider (funded by the German Research Foundation, 2021–2024). The project is film historical in at least two ways: it revisits a specific film historical period and, at the same time, it is based on a data collection that is also already somewhat historical, as it was created a few years ago by someone else – a collection of legacy data if you will.[27]

From 1976 to 1982 the DFG funded a project by Günter Peter Straschek and Thomas Koebner (who was the institutional host for the project). For his research, Straschek had systematically sent out questionnaires to emigrants. After his death in 2009, the collection included a total of over 4,000 paper files of persons related to historical film exiles. Apart from an essay, the only publication was a list of 1,523 names of exiled film personnel. Unlike artistic branches such as theatre, literature, or music, there is no biographical database available about the many scattered professional biographies of those who worked in the German film industries and were forced to emigrate or were even killed by the Nazis during the Second World War.

In our project we revisit the Straschek collection, not trying to complete his work but rather to explore ways of making his unfinished, often unsystematic, and sometimes even idiosyncratic collection accessible for contemporary film historical research. Our particular interest is related to the broad and diverse data in the collection about the professional backgrounds and job biographies of the exiled. The Straschek collection offers insights into the many iterations of Weimar film cultures that are yet to be reconstructed. At the same time, this data about jobs and professional activities in the collection is rather heterogenous: Straschek did not use a controlled vocabulary and worked with a variety of sources for listing jobs people had worked in. The job information that can be found in the per-

27 For a recent suggestion of how to work with legacy catalogues in DH, see James Baker, Andrew Salway, Cynthia Roman, "Detecting and Characterising Transmission from Legacy Collection Catalogues," *Digital Humanities Quarterly* 16, no. 2 (2022), https://www.digitalhumanities.org/dhq/vol/16/2/000615/000615.html.

sonal files unfortunately also comes without any indication of provenance. We do not even know whether they were self-generated or external descriptions, since Straschek collected much of his information via the above-mentioned questionnaires, which were sometimes filled in by relatives, if a person had already died, and sometimes Straschek might have added or changed data if he considered it important.

One might argue that this ephemeral job data should not be used in any scholarly study or at least that it should be "cleaned" by mapping it with other sources. Although we tried to clean and verify the data, we realized that cleaning came at the cost of narrowing the collection down too much. This reduction would be problematic, as other available resources, such as Wikipedia, IMDb or the German filmportal.de, contain rather heterogenous, sometimes even contradictory and insufficient information about many exiled film workers.[28] And even more problematic was the point that reducing the Straschek collection to those persons who could be cross-checked because they were already credited in other databases would mean that many individuals and their contribution to the film industry would be forgotten again as their film jobs are not consistently listed in film archives. This is true for many below-the-line jobs and also for people whose career in the film industry came to an end in exile, often directly because of their forced emigration. So instead of getting rid of ambiguous data, "Mapping German Film Migration" tries to deal with its "porous" data settings productively by finding ways to engage with mistakes, gaps, and ambiguities. In this sense, it engages with the collection's locality, to use the term suggested by Loukissas.[29]

But what does that mean? Let us briefly zoom into two biographies from a case study about female writers in the film industry, which is a rather tricky field. Scripts are often written collectively or by different persons during different stages in the development of a film; sometimes a script is based on an original idea, but it could also be based on a published play, a short story, or a novel. Writing for the screen is a field of practices rather than one specific job. Though guilds have established clear definitions, a lot of writing was and still is undocumented and/or uncredited. This is particularly true for the artists we are researching – the historical film exiles from Nazi Germany. As we all know, exiled writers often face very precarious working conditions, have no work permits at all, or face difficulties in getting accepted by screenwriting guilds. And female laborers of an

28 See also Imme Klages and Alexandra Schneider, "Mapping German Film Migration – digitale Filmgeschichtsschreibung am Beispiel des Nachlasses von Günter Peter Straschek," in *Archive und Museen des Exils*, ed. Sylvia Asmus, Doerte Bischoff, and Burcu Dogramaci (Berlin, Boston: De Gruyter, 2019), 222–238, https://doi.org/10.1515/9783110542103-013.

29 Loukissas, *All Data Are Local*.

industry well known for its gender-bias tend to be forgotten. This is yet another reason why the Straschek collection remains a rich starting point for identifying female writers who worked in the movies. If we assume that he tried to list as many people as possible who were associated with the German film industry of the 1930s, his collection possibly offers the most extensive data on female writers available. Close to 100 names in his list are in some or another way related to female writers (in a very broad understanding).

The following terms can be found in relation to female names in the collection that can be attributed to "writing":

Autorin
Dramatikerin
Dramaturgin
Drehbuch (x)
Drehbuchautor (x)
Drehbuchautorin (x)
Filmjournalistin
Filmkritik
Filmkritikerin
Filmliterat
Journalistin
Librettistin
Novellenautorin
Prosaistin
Publizistin
Romanschriftstellerin
Screenplay (x)
Scenario (x)
Szenario (x)

Six categories (those with an "x") would immediately relate to "scriptwriting"; others might be clustered under the umbrella term of "authors," and yet others would cluster as "journalists." Evaluating some of the controlled vocabularies (the FIAF Glossary and IMDb) showed that the challenge of mapping our list is manifold: apart from some spelling issues, the language, and even the historical variations of jobs of an industry in constant transformation such as the film business, the main issue is that our job information is attached to persons and not to films, so that we actually do not know precisely what writing role someone had in relation to a specific title. Yet another problem is the historicity of job terms that cannot be represented by using a contemporary glossary. A similar challenge applies to the translation of job titles which – both linguistically and industrially – are not universal. Independent filmmaking crediting culture is different from

Hollywood, etc. For all these reasons, indexing the job terms is important because it allows us to keep the local traces of the dataset.

An additional possibility would be to enrich/map our data with other sources, such as IMDb, Wikipedia or the German Filmportal.de. But they all use different vocabularies and, in the case of female scriptwriting, the information from each source differs from every other. On one hand, enriching/mapping is creating more "mess" if you will. At the same time, if we "clean" too much, we risk losing the traces the Straschek collection might offer. This would be the case if, for example, we take out of the collection those persons for whom we do not have "enough" evidence that they really worked in the film industry. This would be the case for two women: Anna Rosa Bernstein and Anna Maria Jokl. Anna Rosa Bernstein (born 1897 in Munich, died 1938 in Zurich) was, according to Straschek, a "Dramaturgin, Lektorin" at the Berliner Drei Masken Publishing house and later worked for MGM. According to Straschek, she arrived via Austria and Hungary in the U.S. in 1933. So far, we cannot find any material trace of her having worked for MGM. Anni Bernstein – as she was also known – killed herself in Zurich Switzerland at the age of 40 in 1938. This trace that can be found via a postcard from 1955,[30] and connects her to a presumably former lover, Gottfried Benn, known not only for his lyrical work but also his antisemitism. According to this trace, her suicide is mentioned in Klaus Mann's diary. To be clear: so far, we have no proof beyond the Straschek collection that Anni Bernstein ever worked in the film industry and also no traces that connect her to the broad field of screenwriting.[31]

A slightly different story is that of Anna Maria Jokl (born 1911 in Vienna, died in Jerusalem in 2001), a known author and psychotherapist who is credited in the Straschek collection as "Cineast, Filmliterat, Drehbuchautor." Though she does have a credit in IMDb, this credit is connected to a novel that was turned into a movie in 2007. In Wikipedia and other sources, she is connected to a film TRATSCH, which she had apparently written but which was not credited to her at its release in 1933 because of her being Jewish.[32] She seemed to have worked for the UFA but, so far, it is mainly her therapeutic and literary work that is known, not her professional activities within the film industry.[33] Anna Maria Jokl was an emerging talent of the German film industry that was forced to change her career. Jokl and Bernstein might not qualify as scriptwriters in a clear-cut sense, but they

30 See https://www.autographen.org/fileadmin/user_upload/5_Kataloge/179.pdf, accessed March 18, 2023.

31 See https://kalliope-verbund.info/de/eac?eac.id=116147237, accessed March 18, 2023.

32 See https://de.wikipedia.org/wiki/Anna_Maria_Jokl, accessed April 14, 2023.

33 See https://www.fembio.org/biographie.php/frau/biographie/anna-maria-jokl/, accessed April 14, 2023.

were part of a broad and diverse community of female talents connected to the publishing and filmmaking world of Weimar Germany.

The Importing Asta Nielsen Database

Our second example is from the Importing Asta Nielsen Database (IANDb).[34] This database features more than 16,000 sets of data – enriched with metadata – on the global distribution and exhibition of the 27 long feature films starring the Danish actress Asta Nielsen, released before the First World War. The Database includes facsimiles of articles and reviews of Asta Nielsen and her 27 films, as well as facsimiles of promotional material of more than 20 countries that are enriched with metadata.

The Importing Asta Nielsen Database is connected to the DFG funded research project "Asta Nielsen – the International Film Star and the Emergence of the Star System 1911–1914" that Martin Loiperdinger and Yvonne Zimmermann conducted from 2018 to 2022. The project takes Asta Nielsen as a case study to research the introduction of the star system in connection with the transition from a program of shorts to feature film exhibition as the standard format of cinema entertainment.[35]

The database is not the result of a structured data model designed at the beginning of a project, based on existing or newly collected data. Instead, it is the provisional outcome of an evolution that began very modestly in 2011 as a rather homegrown MySQL database set up to share researched advertisements on the global distribution of the three Asta Nielsen star series before the First World War with conference attendees in 2011, and later with readers of the subsequent book *Importing Asta Nielsen: The International Film Star in the Making 1910–1914*, edited by Martin Loiperdinger and Uli Jung in 2013.[36]

34 See https://importing-asta-nielsen.online.uni-marburg.de/, accessed April 14, 2023.

35 For the project outcome see Yvonne Zimmermann, ed., "Asta Nielsen, the Film Star System and the Introduction of the Long Feature Film," special issue of *Early Popular Visual Culture* 19, nos. 2–3 (2021), accessed March 21, 2024, https://www.tandfonline.com/toc/repv20/19/2-3. On data-based research with the help of the Importing Asta Nielsen Database, see Martin Loiperdinger, "Early Film Stars in Trade Journals and Newspapers: Data-Based Research on Global Distribution and Local Exhibition," in *The Routledge Companion to New Cinema History*, ed. Daniel Biltereyst, Richard Maltby, and Philippe Meers (London: Routledge, 2019), 138–146.

36 Martin Loiperdinger and Uli Jung, eds., *Importing Asta Nielsen: The International Film Star in the Making 1910–1914* (New Barnet: John Libbey, 2013).

Initially, the database had only two purposes: to store and access data. Processing data was not planned from the outset.[37] "Cleaning" in the sense of standardizing data and metadata had been done along the way, quite intuitively, based on the increasing amount of data collected and the increasing metadata sections implemented, but with no – or at least no explicit – underlying data model or "masterplan." Interoperability was also not considered as a part of the initial design.

In retrospect, two strategies of data cleaning can be identified that were applied to IANDb. The first strategy was to erase "anomalies," that is to dismiss heterogeneity and diversity in favor of cleanliness or "smartness," as Schöch would have it. The second strategy is more in line with Rawson and Muñoz' suggestion of indexing, as an information structure that offers mutual references between "two bodies of information."[38] The second (implicit) strategy then was to keep diversity and the local in the data and to make traces visible.

In what follows, we briefly illustrate the two strategies based on concrete examples. In the trade and local press in English speaking countries, the name Asta Nielsen is often misspelled. However, the database can only be searched using the correct spelling "Nielsen, Asta." The name variants "Neilsen, Asta," "Neilson, Asta," and "Nielson, Asta" that we found in the ads and reviews were eliminated in the IANDb for the sake of standardization. But what if these "misspellings" were not misspellings, but name variants, or, to be more precise, phonetic transcriptions of Asta Nielsen's name in English speaking countries in order to help readers pronounce the Danish star's name correctly? If this assumption is valid, the spelling variants are traces of a conscious adaptation of foreign language names to local cinema audiences and an expression of a historical strategy of film distributors and cinema managers to localize global film stars. Had we kept these spelling variants as index and searchable traces, they could tell us something about the geographical and temporal distribution of the "misspellings" of Asta Nielsen's name. In Loukissas' terms, they would have lent themselves to a situated and local reading of data. Or, going with D'Ignazio and Klein, they would have added situated and local ways of knowing to the dataset. Of course, this problem would not exist if the facsimiles in the database were OCR readable, for that would allow full text searches for the spelling variants. It is planned to implement OCR recognition of the facsimiles in IANDb in the near future, but graphi-

37 We refer here to Lev Manovich's notion of data bases as a basic computer technology that have three basic functions, namely to store, to access, and to process (Manovich, *Cultural Analytics*, 132).

38 Rawson and Muñoz, "Against Cleaning," 288.

cally designed cinema ads and texts that span columns and are interrupted by lines pose a serious challenge to this endeavor.

A good practice example of what Rawson and Muñoz describe as index might be the treatment of data on film distribution titles in the database. Assuming you were interested in JUGEND UND TOLLHEIT, which was released in the 1912/1913 season as the fifth film of the second Asta Nielsen series in Germany, you could first go to the Filmography section of the database where, in addition to the original title, some distribution titles in different languages are listed as follows: JUGEND UND TOLLHEIT, BOHÓ IFJUSÁG, IN A FIX, LAS BATALLAS DEL AMOR, JEUNESSE ET FOLIE, GIOVENTÙ E SPIENSERATEZZA, UNGDOM OG DAARSKAB, MOCIDADE E LOUCURA, Шалости юности.[39] Information on production, script, director, camera, cast, original length, release dates, first screenings, plot description, and access for each of the 27 films can also be found in the Filmography section.

The list of film distribution titles given above is a cleaned and curated selection of the most widespread titles. But IANDb also provides more historical diversity regarding film distribution titles that we found mentioned in the trade and local press. If searching for "Jugend und Tollheit" and "All languages" in the Database section, all the titles that we came across in our data can be found, including the "misspellings" of Asta Nielsen's name. As of March 2023 (the database is constantly being expanded), 54 title variants from 18 countries/empires/colonies are listed.

For Australia, for example, we can find the following titles:

"A Fix"
"Asta in a Fix"
"Asta Neilsen in a Fix"
"Asta Neilson in a Fix"
"Asta Neilson a fix"
"Asta Nielsen in a Fix"

For Great Britain, the following titles are listed:

"Asta in a Fix"
"Asta Neilsen in a Fix"
"Asta Neilson in a Fix"
"Asta Nielsen in a Fix"
"Asta Nielson in a Fix"
"In a Fix"
"Miss Asta Nielsen in a Fix"
"Youth and Frivolity"

39 See https://importing-asta-nielsen.online.uni-marburg.de/filmographie, accessed April 10, 2023.

In addition to the five countries for which we have collected the respective data more or less systematically (Germany, Austria-Hungary, Great Britain, Australia, and New Zealand), title variants are listed from:

Brazil: "Mocidade a Loucura," "Mocidade e Loucura"
Denmark: "Ungdom og Daarskab"
France: "Jeunesse et Folie"
Iceland: "Æskubrek"
Ireland: "Asta Nielsen in a Fix"
Italy: "Giovinezza e pazzia"
Netherlands: "Dolle jeugd," "Stormen der jeugd"
Netherlands Indies: "Dolle Jeugd," "Een les in de liefde"
Russian Empire: "Jugend und Tollheit," "Wybryki Mlodości," "*Шалости юности*"
Singapore: "The Springtime of Youth"
Spain: "Las batallas del amor"
United States: "Lady Madcap's Way"

This list is intentionally detailed to be overwhelming and potentially off-putting. Still, we perceive it as a good practice example for at least two reasons. First, it makes visible the labor that went into data collection and metadata preparation. To make labor visible is the seventh core principle of the data feminism that D'Ignazio and Klein have developed, arguing that "[t]he work of data science, like all work in the world, is the work of many hands."[40] This work can only be recognized and valued if made visible. In the case of IANDb, students and student assistants from the Universities of Trier and Marburg did much of the work of collecting and formatting data and entering metadata, and they are acknowledged in the Content section of the Database.[41]

The second reason can be illustrated by a comparison of the Importing Asta Nielsen Database with IMDb. In IMDb, JUGEND UND TOLLHEIT is listed under the title "Lady Madcap's Way." "Jugend und Tollheit" is given as the original title.[42] From IANDb, we have learned that LADY MADCAP'S WAY was indeed the U.S. distribution title, but that this was not, as the IMDb entry suggests, the only English language title by far, let alone the only distribution title. With IMDb and IANDb, we have two extremes: on the one hand, a reduction of data made to compute and, on the other hand, data heterogeneity that does not compute but that allows one to look at the local in data when studying the global circulation of films. The comparison between IMDb and IANDb also resonates with D'Ignazio and Klein's first and second

40 D'Ignazio and Klein, *Data Feminism*, 22.
41 See https://importing-asta-nielsen.online.uni-marburg.de/content, accessed April 10, 2023.
42 See https://www.imdb.com/title/tt0132258/?ref_=ext_shr_lnk, accessed April 10, 2023.

core principles of data feminism, namely to *examine power* and to *challenge power*. According to the authors, data feminism "begins by analyzing how power operates in the world" and "commits to challenging unequal power structures and working towards justice."[43] The entry for "Lady Madcap's Way" in IMDb gives the impression that this data is universal – which is its power. IANDb with its non-standardizing of data preserves its cultural particularism, local embeddedness, and heterogeneity, and thus challenges the (supposed) supremacy of the U.S. distribution title.

Beyond Cleaning

Although Rawson and Muñoz' idea of the index might not really resolve what we called the "Drucker-Dilemma," we still find their and others' suggestions to critically engage with the concept of "cleaning" useful on a conceptual level. It reminds us of the importance of preserving diversity and understanding scalability as both a promise *and* a predicament. It is a promise because it does allow one to re-think and re-connect dots between often isolated research questions. Their idea also makes it evident that data and the tools with which we manipulate them do not reduce complexity, but "add complexity to the relation between researchers and their objects of study," as Schöch has argued.[44] But scalability, as we have seen, can come with the prize of streamlining data (or realities for that matter) for the sake of interoperability. Interoperability is anything but a technical issue. From our perspective, interoperability should strive to capture complexity and respect differences in standards to the greatest possible degree, and reflect the epistemological implications of the concept. At the same time, we might also want to question our ethical standards, as Deb Verhoeven has suggested, regarding the FAIR principles when she states: "Herein lies the inherent unfairness of FAIR. FAIR treats data and information as if it is loosed from its context and is therefore purely exploitable [. . .]."[45] New boundaries of inclusion and exclusion are set. The same goes for the challenge to translate locality into data models and preparation practices, as both our projects illustrate. Instead of disciplining film studies in the wake of computational humanities, one might

43 D'Ignazio and Klein, *Data Feminism*, 21.

44 Schöch, "Data in the Humanities," n.p.

45 Deb Verhoeven, "Scholarship in a Clopen World," *Pop! Public. Open. Participatory* 4 (2022): n.p., https://doi.org/10.54590/pop.2022.002.

rather rethink the potential of the undisciplined or resisting potential of what could be a critical computational film studies stance.

References

Baker, James, Andrew Salway, and Cynthia Roman. "Detecting and Characterising Transmission from Legacy Collection Catalogues." *Digital Humanities Quarterly* 16, no. 2 (2022). Accessed December 16, 2023. https://www.digitalhumanities.org/dhq/vol/16/2/000615/000615.html.

Benjamin, Ruha. "Race After Technology: Abolitionist Tools for the New Jim Code." *Social Forces* 98, no. 4 (2020): 1–3. https://doi.org/10.1093/sf/soz162.

D'Ignazio, Catherine, and Lauren F. Klein. *Data Feminism*. Cambridge, MA: MIT Press, 2020.

Dizdar, Dilek. "Translation und Grenze. Versuch einer translationswissenschaftlichen Neufiguration." In *Übersetzung*, edited by N. Engel and S. Köngeter, 57–74. Wiesbaden: Springer VS, 2020. https://doi.org/10.1007/978-3-658-20321-4_4.

Drucker, Johanna. *The Digital Humanities Coursebook: An Introduction to Digital Methods for Research and Scholarship*, 1st ed. Abingdon, Oxon; New York: Routledge, 2021.

Es, Karin van, and Nanna Verhoeff, eds. *Situating Data: Inquiries in Algorithmic Culture*. Amsterdam: Amsterdam University Press, 2023. https://doi10.5117/9789463722971_intro.

Klages, Imme, and Alexandra Schneider. "Mapping German Film Migration – digitale Filmgeschichtsschreibung am Beispiel des Nachlasses von Günter Peter Straschek." In *Archive und Museen des Exils*, edited by Sylvia Asmus, Doerte Bischoff, and Burcu Dogramaci, 222–238. Berlin, Boston: De Gruyter, 2019. https://doi.org/10.1515/9783110542103-013.

Loiperdinger, Martin. "Early Film Stars in Trade Journals and Newspapers: Data-Based Research on Global Distribution and Local Exhibition." In *The Routledge Companion to New Cinema History*, edited by Daniel Biltereyst, Richard Maltby, and Philippe Meers, 138–146. London: Routledge, 2019.

Loiperdinger, Martin, and Uli Jung, eds. *Importing Asta Nielsen: The International Film Star in the Making 1910–1914*. New Barnet: John Libbey, 2013.

Loukissas, Yanni Alexander. *All Data Are Local: Thinking Critically in a Data-Driven Society*. Cambridge, MA: MIT Press, 2019.

Manovich, Lev. *Cultural Analytics*. Cambridge, MA: MIT Press, 2020.

Noble, Safiya Umoja. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press, 2018.

Oram, Andy. *Managing the Data Lake*. Sebastopol: O'Reilly Media, 2015. Accessed March 21, 2024. https://learning.oreilly.com/library/view/managing-the-data/9781492049876/.

Pearson, Ronald K. *Mathematics in Industry: Mining Imperfect Data: With Examples in R and Python*, 2nd ed. Philadelphia: Siam, 2020. https://doi.org/10.1137/1.9781611976274.

Rawson, Katie, and Trever Muñoz. "Against Cleaning." In *Debates in the Digital Humanities 2019*, edited by Matthew K. Gold and Lauren F. Klein, 279–292. Minneapolis: University of Minnesota Press, 2019. https://doi.org/10.5749/j.ctvg251hk.26.

Richardson, Rashida, Jason Schultz, and Kate Crawford. "Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice." *94 N.Y.U. L. REV. ONLINE* 192 (2019): 199–200. Accessed March 21, 2024. https://ssrn.com/abstract=3333423.

Schöch, Christof. “Big? Smart? Clean? Messy? Data in the Humanities.” *Journal of the Digital Humanities* 2, no. 3 (2013). Accessed March 21, 2024. https://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/.

Stauff, Markus, Pauline van Romondt Vis, and Karin van Es. “Coffee Roasters’ Data Vernacular: On the Entanglement of Digital Data and Craft.” In *Situating Data: Inquiries in Algorithmic Culture*, edited by Karin van Es and Nanna Verhoeff, 31–48. Amsterdam: Amsterdam University Press, 2023.

Van Gorp, Jasmijn. “Interstitial Data: Tracing Metadata in Archival Search Systems.” In *Situating Data: Inquiries in Algorithmic Culture*, edited by Karin van Es and Nanna Verhoeff, 207–222. Amsterdam: Amsterdam University Press, 2023.

Verhoeven, Deb. “Scholarship in a Clopen World.” *Pop! Public. Open. Participatory* 4 (2022). https://doi.org/10.54590/pop.2022.002.

Zimmermann, Yvonne, ed. “Asta Nielsen, the Film Star System and the Introduction of the Long Feature Film.” Special issue of *Early Popular Visual Culture* 19, nos. 2–3 (2021). Accessed March 21, 2024. https://www.tandfonline.com/toc/repv20/19/2-3.

Kerstin Herlt and Julia Welter

Critically Curating Data in Cultural Heritage Collections

Introduction

Film archives, like other cultural heritage institutions (CHIs), play a crucial role in shaping our audiovisual memory and understanding of the past. However, these archives are not neutral. They make choices about what to include, inadvertently privileging some narratives while excluding others. Historically, these institutions have catalogued objects from a perspective that favors privileged groups, leaving out the perspectives and knowledge of minorities, such as BIPOC and LGBTQ+ people. As a result, archives and museums often become sites of exclusion for these communities, where their stories and artefacts receive limited or no visibility.[1]

In the field of film heritage, the creation of national film canons or lists of "masterpieces" tends to favor certain narratives. However, search platforms such as EFG – The European Film Gateway, which was developed in 2011 under the leadership of the DFF – Deutsches Filminstitut & Filmmuseum and 15 other European film archives and now provides access to collections from over 50 film archives,[2] have broadened access across Europe to previously neglected and ephemeral formats, such as newsreels, short documentaries, travelogues, commercials, and so-called usage films. With its wide range of audiovisual collections covering a variety of topics in different formats, EFG is a rich resource for research into European film production but also a large pool of mostly uncurated material that can make it difficult for users to grasp what exactly is available. Notable exceptions are the EFG's extensive First World War collection of over 3,000 titles from 25 European film archives,[3] as well as its comprehensive thematic focus on some 700 non-fiction films depicting life and reconstruction efforts in post-World War II Europe from 15

1 Dagmar Brunow, "Curating Access to Audiovisual Heritage: Cultural Memory and Diversity in European Film Archives," *Image and Narrative* 18, no. 1 (2017): 101. See also Dang's chapter in this volume, "Managing the Past: Research Data and Film History." Dang reflects on the role of metadata and archival practices for our understanding of film history.

2 Many of them are members of the European Association of Cinematheques, www.ace-film.eu, accessed April 19, 2024.

3 Films related to WWI were digitized and aggregated by 25 European film archives between 2012 and 2014 in the scope of the EFG1914 project. See https://www.europeanfilmgateway.eu/content/efg1914-project, accessed April 1, 2024.

https://doi.org/10.1515/9783111082486-010

film archives.[4] They provide a transnational and multi-perspective approach to two specific events and time periods affecting a large number of (European) countries.

To aid discoverability of the films in both thematic collections, EFG provides browsing entry points by topic, based on a common set of keywords from thesauri such as the Library of Congress Subject Headings. In addition to the thematic collections, the platform provides access to online exhibitions and learning materials that use selected film clips to provide curated access and promote film as a historical source.[5]

While the digital transformation and the creation of platforms such as EFG and the digital library Europeana[6] have enabled a broader perspective on historical events and experiences through a more diverse representation of film and film-related material – and in the case of Europeana, cultural heritage in general – there is an increasing awareness of social injustice and insensitivity in cultural heritage institutions and websites like these. The debate on decolonizing archives and museums shows that providing a wide range of access to content, transnational approaches, and curatorial efforts alone is not sufficient to respond to the call of communities for increased polyvocality, diversity, and equity in collections and collection descriptions.

The content of EFG and Europeana, as well as on Europeana's other aggregator portals,[7] is mainly provided by institutional archives, libraries, and museums. These collections often emphasize national and regional heritage, which sometimes leads to the promotion of conventional national narratives. There are few non-

4 Films from 15 archives focusing on reconstruction efforts after WWII were brought together on EFG as part of the EU-funded research project ViCTOR-E – Visual Culture of Trauma, Obliteration and Reconstruction in Post-WWII Europe, which was led by the Goethe University in Frankfurt, https://www.europeanfilmgateway.eu/content/VICTOR-E-project, accessed April 1, 2024.

5 Online exhibition developed as part of the EU-funded ViCTOR-E project on reconstruction efforts in post-war Europe, including e-learning exercises, https://www.frames-reconstruction.eu, accessed April 1, 2024.

6 Europeana, initiated and funded by the European Union, is an online portal of digitized cultural heritage collections from over 3,000 institutions across Europe. It holds records of 50 million cultural and scientific artefacts and presents these resources on a single platform tailored to modern user needs. Initially inspired by the European Digital Library Network (EDLnet), Europeana was launched in 2008. The Europeana Foundation oversees the operation of the service.

7 The EFG is one of 40 so-called aggregators that provide data from their partner archives to Europeana. Many also run their own portal. Aggregators ensure that the data from their partner archives conforms to Europeana's quality standards and frameworks. All objects searchable through Europeana are hosted locally by the data-providing institutions. Only the metadata and preview thumbnails are stored centrally. Any changes to the metadata must be made at the source.

institutional archives, such as activist archives or community archives, which focus on and reflect stories and experiences of marginalized groups, among the data providers to Europeana. In recent years we have seen a shift in the approach of cultural heritage institutions, spurred on by civil society movements and marginalized communities calling for greater and more accurate representation.

In the film archival sector, discussions on decolonial strategies have surfaced as well. The International Conference "Archives Assembly,"[8] the Global Audiovisual Archiving Conferences,[9] and The FIAF symposium 2024 "Film Archives in the Global South"[10] are just three events that aim to explore and expand decolonial practices within the global audiovisual archival community.

Efforts to decolonize archives and diversify collections must address harmful representations. For example, reviewing EFG collections through the lens of minority communities reveals blind spots in metadata, including offensive language in descriptions, sometimes dating back decades, which have not been revised due to lack of awareness or limited funding. While archival work has focused on the development of common standards, interoperable metadata schemas, and controlled vocabularies, there has been little effort made to revise catalogue records. From today's perspective and with an advanced level of sensitivity to minority issues, there is a clear understanding of European (film) archives to revise catalogue entries and address the dominant Eurocentric gaze that often portrays marginalized groups in stereotypical ways. Some examples of this are the problematic cinematic representations of soldiers from colonized countries during the First World War, Roma people in silent films, and African people in ethnographic documentaries, which often portray these groups as "other."

The initiative to reflect on derogatory representations and to revise descriptions and language in archival descriptions increasingly comes from the communities themselves. An example from the EFG relates to Das Mädchen ohne Vaterland

8 Archives Assembly is a bi-annual festival organized by Arsenal Berlin in cooperation with the Goethe University Frankfurt and other partners. Archive Assembly #1 was organized in 2021: https://www.arsenal-berlin.de/en/archive-distribution/archive-projects/archive-ausser-sich-2017-21/program/, accessed April 1, 2024; Archive Assembly #2 in 2023: https://www.arsenal-berlin.de/news/mark-the-date-archival-assembly-2/, accessed April 1, 2024.

9 7th Eye International Conference on "Global Audiovisual Archiving: North-South Exchange in Knowledge and Practices," accessed April 1, 2024, https://www.eyefilm.nl/en/programme/eye-international-conference-2022/563352; organized by Eye Filmmuseum, the University of Amsterdam, and the Association of Moving Image Archivists in May 2022. The second conference, presented by Archive/Counter-Archive, Eye Filmmuseum, and the Toronto International Film Festival, will take place in 2024 in Toronto, https://counterarchive.ca/gava-2024, accessed April 1, 2024.

10 Symposium of the 2024 FIAF Congress "Film Archives in the Global South," accessed April 1, 2024, https://www.fiafnet.org/pages/Events/2024-fiaf-symposium-call-for-papers.html.

(Urban Gad, GER 1912), in which Asta Nielsen plays Zidra, the "gypsy girl." The film picks up on the stereotypes of child abduction associated with the Roma and shows the "unpatriotic gypsy girl" against the backdrop of the nationalistically charged pre-war period.[11] The Cultural Centre of the Anna Schwarz RomnoKehr Association in Oldenburg, Germany, contacted the DFF – Deutsches Filminstitut & Filmmuseum, which provided the dated synopsis from its catalogue, and prompted a change in the content description, highlighting and contextualizing the negative gypsy stereotypes present in the original synopsis.[12]

The synopsis was rewritten to remove outdated language and mention that the film contained gypsy stereotypes. The word "gypsy girl" was put in quotation marks, as requested by RomnoKehr. However, the approach of putting quotation marks around potentially offensive terms, while a first step and an easy solution in some cases, may not fully address the problem in the long run. What would be an appropriate strategy for cultural heritage institutions to detect and update terms and contextualize content that is considered harmful and inappropriate?

As coordinators of the EFG and contributors to the Europeana database, we at the DFF – Deutsches Filminstitut & Filmmuseum were keen to learn from the experiences of other archives and aggregators. On our initiative and that of the Netherlands Institute for Sound & Vision, we set up the Europeana Aggregators Working Group in autumn 2021 to gain a deeper insight into the challenges of identifying and enhancing the diversity of Europeana's collections. The findings of this working group made it clear that in order to effectively address these challenges we needed a robust solution capable of managing and processing large datasets. From January 2022 to December 2023, the DFF, together with eleven European partner institutions, had been working on this very issue as part of the DE-BIAS project.[13]

---

11 Mädchen ohne Vaterland, accessed April 1, 2024, https://www.europeanfilmgateway.eu/search-efg/m%26auml%3Bdchen%20ohne%20vaterland%20?page=0%2C1%2C0. Original content description: "In den Gebirgswäldern am Balkan liegt eine kleine aber wichtige Grenzfestung, die der Feind zur Eroberung ausspionieren möchte. Als Spionin von durchschlagender Wirkung wird eingesetzt Zidra, die schönste Blüte eines Zigeunerstammes." The content description has been changed to: "In den Gebirgswäldern eines Lands auf dem Balkan liegt eine kleine aber wichtige Grenzfestung, die der Feind zur Eroberung ausspionieren möchte. Dafür wird die 'Zigeunerin' Zidra engagiert, die den diensthabenden Leutnant bezirzen und so an die Pläne der Festung gelangen soll. Im Film schwingen für die Zeit des Ersten Weltkriegs übliche nationalistische Stimmungen mit, zudem werden antiziganistische Stereotype verwendet."

12 See also https://www.romno.de/film-geschichte/, accessed April 1, 2024.

13 DE-BIAS: Detecting and Cur(at)ing Harmful Language in Cultural Heritage Collections, accessed April 1, 2024, https://pro.europeana.eu/project/de-bias. The project started in January 2023 and will run for two years. It supports the Deployment of the Data Space for Cultural Heritage, led by the Europeana Foundation.

Analyzing Bias in Europeana Collections

The Europeana Aggregators Working Group (AGG Working Group) was formed to investigate the issue of problematic descriptions or visual representations, and to develop recommendations for cultural heritage institutions providing data to Europeana. The group's approach is to conduct case studies, using examples provided by AGG WG members, to understand the challenges of managing diversity in cultural heritage data. In addition, the group is reviewing existing guidelines to gain insights into how to deal with offensive language or content in archival collections.

Methodological Approach – User Stories and Case Studies

Initially, the AGG Working Group focused on clarifying the definitions of "harmful" and "offensive" content, recognizing the influence of language and context on different user perspectives. The overall aim is to diversify collections by addressing gaps and promoting inclusivity through collaboration with partners and stakeholders. To facilitate this, members of the AGG Working Group have contributed examples from their respective backgrounds, highlighting problematic elements in content, terminology, and overlooked narratives to enable a comprehensive exploration of the issues at hand.[14]

As a methodology, the AGG Working Group created user stories from the perspective of different types of users – CHI professionals, aggregators, history teachers, students.[15] The template for these user stories was formulated as follows: As [type of user], I want to [perform a task] so that I can [achieve a goal]. The aim was to understand the mindset of different users in order to:

14 Members of the AGG WG were operations and projects manager from the domain and thematic aggregators on Europeana: APEF – Archives Portal Europe (archival finding aids and inventories), DFF/EFG (film heritage), EFHA – European Fashion Heritage Association (fashion related content), Photoconsortium (photographic heritage), MUSEU (museum collections), Open-Up (natural history content), Netherlands Institute for Sound & Vision/EUscreen (audiovisual heritage), Europeana Sounds (music related content). All members were white people, seven women and two men, aged between 35 and 55, from five European countries, see https://pro.europeana.eu/page/aggregators?utm_source=share-your-data%2Fprocess&utm_medium=Find%20an%20aggregator&utm_campaign=internal_link, accessed April 1, 2024.

15 The choice of methodology for the user stories and case studies was mainly informed by Lauren Vargas – Your Digital Tattoo (https://yourdigitaltattoo.com/, accessed April 1, 2024) – who is an external consultant in digital transformation and community management.

- Define WHO could be harmed by the descriptions and representations;
- Understand the WHY by sharing context, highlighting inequalities and stereotypes, and challenging dominant narratives;
- Identify WHAT information is needed to fill gaps and make collections more polyvocal;
- Explore HOW to address (historical) practices of colonialism, racism, sexism, and other forms of domination that have influenced archival practices.[16]

Following this, the user stories were complemented by seven case studies covering different types of media (sound recordings, fashion items, moving images, photographs, botanical drawings, text documents). Six of the case studies dealt with problematic content, content descriptions, and the lack of metadata related to the colonial past, while the seventh case study was based on Roma-related content.[17]

In this article, we want to focus on the case study provided by DFF/EFG to illustrate the practical application of the AGG WG's principles within the audiovisual heritage context.

Improving Metadata and Contextualising Colonial Narratives: Dämonisches Afrika

Our case study began with one of the user stories we had created with the Working Group before:

> As a cultural heritage professional, I want to search EFG for films and film-related material about Africa with a colonial context to see how this content is presented and described in our collection.

Searching the EFG using various English and German keywords related to "Africa" and "colonialism" yielded 25 film stills from the documentary Dämonisches Afrika (Hans Leuenberger, CH 1952/1953). The stills were originally published on DFF's platform filmportal.de and aggregated to Europeana via EFG.[18] However, the images

16 Charlotte Lellman, "Guidelines for Inclusive and Conscientious Description," *Center for the History of Medicine: Policies and Procedures Manual* (2020), accessed April 1, 2024, https://wiki.harvard.edu/confluence/display/hmschommanual/Guidelines+for+Inclusive+and+Conscientious+Description.

17 Kerstin Herlt, Kristina Rose, and Kerstin Arnold, "Final Recommendations for Identifying and Managing Diversity in Cultural Heritage Data," for internal distribution only, n.d.

18 Dämonisches Afrika, 25 film stills, accessed April 1, 2024, https://www.europeanfilmgateway.eu/search-efg/d%26auml%3Bmonisches%20afrika?page=0%2C1%2C0.

lacked comprehensive descriptions, providing only basic metadata, such as object type, title, author, year of production, rights information, provenance/collection, and providing institution. Going back to filmportal.de, we found filmographic information, including director, country, year of production, screenwriter, photographer, editor, composer, production company, and length.[19] However, crucial contextual information such as a synopsis, where the film was shot (country or region), and details of the people portrayed in the film were missing.

In order to be able to fill these gaps as far as possible, we went back to watching the film in its entirety in DFF's film archive. The DFF preserves two analogue prints and a trailer of Dämonisches Afrika.[20] Interestingly, the trailer differs significantly from the film in that the illustrations and text used in the introductory intertitles are much more sensationalist than the full-length film. In order to grab the audience's attention, it propagates colonial and racist stereotypes about Africa, such as exoticism: "The fight between the black magic of the dark continent and the white demons of modern civilization" (Figure 1) and (sexual) excitement: "The obsessed women of the Kunama people, driven by ghosts of dead warriors to wild war games of unique appeal" (Figure 2).

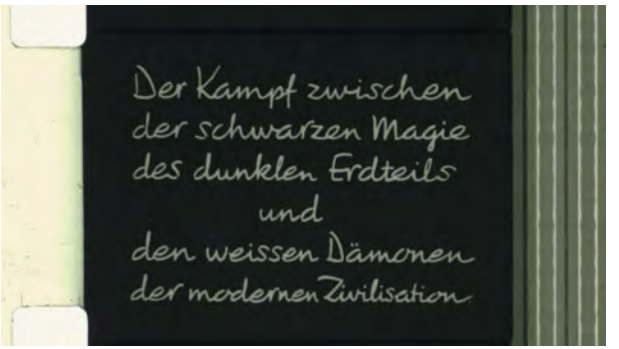


Figure 1: Prologue from the trailer Dämonisches Afrika. Source: DFF – Deutsches Filminstitut & Filmmuseum.

19 Dämonisches Afrika on Filmportal, accessed April 1, 2024, https://www.filmportal.de/en/movie/damonisches-afrika_ea43d4a717e95006e03053d50b37753d.

20 DFF has created a Quadriga Quadscan for internal research purposes.

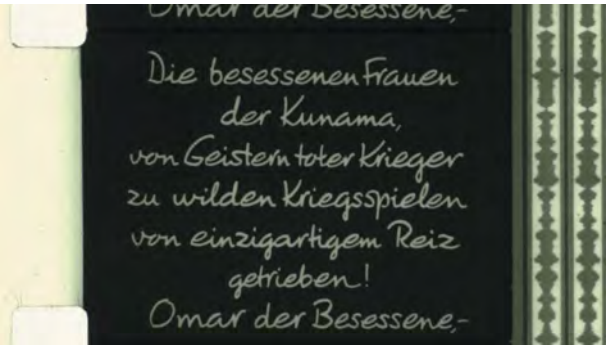


Figure 2: Prologue from the trailer DÄMONISCHES AFRIKA. Source: DFF – Deutsches Filminstitut & Filmmuseum.

Watching the full-length film provided information about the people portrayed and the locations. DÄMONISCHES AFRIKA is an ethnographic journey through Ethiopia along the borders of Sudan, Eritrea, and Egypt.

It portrays people from the region such as the Anouak, Kunama, Betschuk, and Rashaida. It alternates between generic rural scenes (fishing, farming, water fetching, food preparation) and glimpses of urban life in Asmara (traffic, schooling, local industries). The film's "demonic" motif refers to the exploitative technology brought by the White "demon" and the portrayal of "African people's" innate abilities and mystical rituals. The film embodies a European craving for "primitivism," projecting racialized and "primitive" images onto its subjects, especially women and children. This representation is in keeping with old ethnographic filmmaking traditions rooted in colonialism and described by Shankar Arjun "as a process of racialization and the production of 'primitive' subjects over time."[21] Like many of the films that look at Africans from a colonizer's perspective, DÄMONISCHES AFRIKA perpetuates racism and, in its latter part, sexism, when the camera follows a group of young Kunama women as they go about their daily tasks. The male voice-over comments on these scenes in a blatantly sexist way, mocking the women and not only reducing them to "objects of the male gaze,"[22] but creating a story that is completely disconnected from the cultural identity of these young women.[23]

21 Arjun Shankar, "Primitivism and Race in Ethnographic Film: A Decolonial Re-Visioning," *Oxford Bibliographies: Anthropology* (2020): 1–25, https://doi.org/10.1093/obo/9780199766567-0245.

22 Laura Mulvey, "Visual Pleasure and Narrative Cinema," *Screen* 16, no. 3 (1975): 6–18.

23 "so schleppen sich diese Kunama-Damen schlecht und recht durchs Leben, sich und die Lasten, die ihnen ihr Hausfrauendasein aufbürdet. Sie schleppen und schleppen und es wäre kaum

The case of DÄMONISCHES AFRIKA highlights how the interplay of moving images, text and voice-over, editing and soundtrack creates a film that is (not only) highly problematic from today's perspective, especially for the people represented in the film and those communities and people who might identify with the Africans portrayed in the film. On the other hand, ethnographic films like DÄMONISCHES AFRIKA can be valuable testimonies, especially for the countries in which they were made, as they are often the first traces of a country's audiovisual memory, even if seen through the eyes of the colonizer.

Adding a synopsis to DÄMONISCHES AFRIKA would be the first step to improve the discoverability of the content. The film provided rich information about the location and the ethnic groups represented, which should be added to the database. Using specific and accurate terminology for geographic locations and for the description of persons or groups not only increases the discoverability of the content, it also counteracts a homogenizing colonial discourse and generalizations such as "Africans" or "Africa," which often seems to be reduced to a single country rather than a continent of more than 50 countries with a great diversity of languages and identities.[24]

All case studies carried out by the AGG Working Group showed, on the one hand, the presence of harmful content and offensive language rooted in colonialism, sexism, and racism, as well as stereotyped representations of people with different ethnic backgrounds from African countries or Roma people and, on the other hand, a lack of descriptive metadata that would allow the content to be contextualized. The working group defined action points and made recommendations on how cultural heritage institutions can mitigate the harm caused by these shortcomings, and bring collections from, about, or for underrepresented groups to the fore.

mitanzusehen, wären sie nicht so reizend anzusehen"/"so these Kunama ladies drag themselves through life for better or worse, like the burdens that their housewifery imposes on them. They drag and drag and it would be hard to watch if they weren't so lovely to look at" (Time code 00:54:15).

24 Annette Schmidt, "Africa Is Not a Country," in *Words Matter, Work in Progress I*, ed. Wayne Modest and Robin Lelijveld (National Museum of World Cultures, 2018), 44, accessed April 27, 2024, https://www.materialculture.nl/en/publications/words-matter.

Critical Cataloguing: Recommendations from the Europeana Aggregators Working Group

Create Knowledge on the Language to be Used

One approach is publishing a statement about harmful language on the website that explains the biases that are inherently linked to catalogue entries and legacy data. The Cataloguing Lab Initiative has compiled a list of these statements from libraries, archives, and museums in the U.S., Australia, and Canada.[25] In many cases, such as the DPLA Black Women Suffrage, these statements also explain where the content comes from, why the content and the descriptions may be harmful, how the institution intends to address the issue, and how users can report harmful content.[26] Publishing a statement is only useful and credible if it explains how institutions intend to tackle harmful content and terminology.

Understanding the mindset of the different stakeholders and users might help to overcome a "gatekeeper's" attitude. Building a team by bringing together cataloguers, archivists, and members of communities to review how language is used to catalogue and describe collection items is a first step. The working group also recommends consulting inclusive language glossaries for improving descriptions. Useful glossaries already exist as work in progress, such as *Words Matter*[27] or the *Inclusive Terminology Glossary*,[28] which provide guidance on non-discriminatory language for cultural heritage professionals. The *Inclusive Terminology Glossary* has been developed in the form of Google documents and is conceived as a collaborative and crowd-sourcing work. It is divided into different domains or "areas of cultural sensitivity," such as African American history, Indigenous people from different countries and regions, Travelling Communities, LGBTQ+, History of Antisemitism, and more.

Words Matter has a different approach as it was not compiled through crowd sourcing, but it is also conceived as work in progress. For the National Museum of

25 The Cataloguing Lab Initiative, "List of Statements on Bias in Library and Archives Description," last modified September 2023, accessed April 27, 2024, https://cataloginglab.org/list-of-statements-on-bias-in-library-and-archives-description/.

26 DPLA Black Women's Suffrage, Harmful language statement, accessed April 1, 2024, https://blackwomenssuffrage.dp.la/harmful-language-statement.

27 Wayne Modest and Robin Lelijveld, eds., *Words Matter, Work in Progress I* (National Museum of World Cultures, 2018), accessed April 1, 2024, https://www.materialculture.nl/en/publications/words-matter.

28 Inclusive Terminology Glossary, accessed April 1, 2024, https://culturalheritageterminology.co.uk/glossary/.

World Cultures, the question of how to comprehensively represent the heritage of different cultures is central, and hence the need to take a self-critical look at the descriptions of the collections. The publication brings together different perspectives on dealing with discriminatory language in museums and archives, from the perspective of museum staff, community representatives, and interest groups. Starting from the knowledge that "words matter" and that language influences whether groups or communities feel a sense of belonging, the publication gathers positions on whether it is opportune to change words, and how to update them without wanting to change history as such. Language always represents an "ideologically colored worldview," as Esther Peeren explains. In 2016, the Dutch Scientific Council for Government Policy recommended replacing the word "allochtoon" (immigrant) with "migrant." However, according to Peeren, "migrant" has negative connotations and is often linked to "Muslim" in the wake of growing Islamophobia: "That [these terms] are anything but neutral becomes clear when we take into account that, in the current European context, both these terms are intricately bound up with the rise of Islamophobic and more generally xenophobic worldviews. Language, then, is never neutral or objective, nor does it lend itself to a 'big clean-up'."[29]

The key question for members of the working group was: What are appropriate approaches to "fixing" harmful terminology, while recognizing that language is never neutral and changes over time?

Critical Cataloguing

Descriptive metadata which have been created decades or even centuries ago are part of the history of a cultural heritage object. They are living documents as they will evolve throughout their lifecycle in accordance with the holding institutions' current policies. Hence, there are various reasons in favor of keeping the original records because, as they are historical data, replacing the original record seems like hiding past practices. At the same time, cultural heritage institutions have the responsibility to describe their collections accurately and in a way that is respectful to the communities that are represented in their collections.

In the case of Dämonisches Afrika, the majority of the working group members agreed that original (film) titles shouldn't be changed, which is basically in

29 Esther Peeren, "Language Cannot Be 'Cleaned Up'," in *Words Matter, Work in Progress I*, ed. Wayne Modest and Robin Lelijveld (National Museum of World Cultures, 2018), 44, accessed April 1, 2024, https://www.materialculture.nl/en/publications/words-matter.

line with the cataloguing rules of most CHIs. In this case an alternative title could be added: if it's not an original title, but a descriptive title given by the institutions themselves, it can be replaced. But, in order to track the changes in the record, it is recommended to keep the former title in the database.

As regards synopses, some institutions distinguish between fictional and documentary material. In case of documentary films, for example, inappropriate or outdated terminology could be replaced by more inclusive terms. In the case of fictional films, however, replacing "Gypsy" by "Roma people" or "Indians" by "First Nations people" would deny their original reality in that reference to an existing fictional stereotype in cinema of that time would be lost.[30]

When choosing or building a vocabulary for cataloguing and indexing, thesauri functionalities such as preferred terms can be installed to create references between terms that are in use. If a user searches the catalogue for "Eskimo" or "Gypsy," the search result would display "Inuit" or "Roma," so the offensive terms wouldn't be visible, but the outdated term would be still stored in the database.[31] In this way, the database continues to include all the historical and problematic terms that have been used to describe objects over time, but without displaying them to the public.

As a general recommendation, it should be noted that vocabularies that support multilingualism and/or are available as linked open data (LOD) are especially valuable. *Homosaurus*, for example, is a linked data vocabulary of LGBTQ+-specific terms intended to supplement existing thesauri, such as the Library of Congress Subject Headings. By adding additional controlled vocabulary in the catalogue of CHIs, access and the discoverability of LGBTQ+ resources can be improved.[32]

While there is a variety of approaches, guidelines, and glossaries to support the work of inclusive description, two aspects have not been addressed so far: (1) Many institutions cannot update legacy metadata in their databases due to the sheer size and scope of their collections and the complexity of the endeavor; and (2) how can these institutions collaborate with communities and community allies (researchers, critical friends, civil society organizations, counter archives) in the creation of inclusive glossaries?

These two aspects are at the core of the DE-BIAS project, which will be presented in the next section.

---

30 Recommendations of the Catalogue Project Group – Eye Filmmuseum, July 20, 2021. Unpublished Paper.

31 Marijke Kunst, "Being True to the Catalogue," in *Words Matter, Work in Progress I*, ed. Wayne Modest and Robin Lelijveld (National Museum of World Cultures, 2018), 29–34, accessed April 1, 2024, https://www.materialculture.nl/en/publications/words-matter.

32 Homosaurus Vocabulary Terms, last modified June 2023, accessed April 1, 2024, https://homosaurus.org/v3.

Detecting Offensive Language in Cultural Heritage Collections through Data-Driven Tools. Introducing the DE-BIAS Project

Updating metadata in legacy databases is a complex undertaking due to the sheer size and scope of the collections and lack of resources. Members of the AGG Working Group have decided to submit a proposal for developing a tool that automatically detects and flags offensive terms in large datasets. The DE-BIAS project was eventually selected for funding. It involves eleven partners and is coordinated by DFF – Deutsches Filminstitut & Filmmuseum. Its primary goal is to assist cultural heritage institutions in promoting a systematic, more inclusive, and non-discriminatory approach to describing cultural heritage collections, in particular those published on Europeana. To achieve this goal, the DE-BIAS consortium is working on three main tasks: the creation of the DE-BIAS vocabulary in collaboration with communities, the development of the bias detecting tool, and capacity building.[33]

Creating the DE-BIAS Vocabulary

DE-BIAS recognizes the need for cultural heritage institutions to critically engage with their power as gatekeepers of knowledge and to distribute that power to the communities they represent and that are represented – or not – in metadata and objects these institutions manage. This includes the institutions actively seeking knowledge from underrepresented communities and recognizing them as experts rather than passive audiences. Building relationships with communities will lead to richer, more diverse collection descriptions – and possibly more diverse collections – and a more inclusive society.

The DE-BIAS vocabulary will be built around three main themes: (1) migration and the colonial past; (2) gender and sexual identity; (3) ethnicity and ethno-religious identity. To train the algorithm and make the DE-BIAS tool[34] both performant and relevant for this purpose, the establishment of a conscientiously

33 Capacity building for CHIs, Europeana aggregators, and policy makers is crucial for the uptake of the project results, but we will only briefly touch upon this task in this chapter. Capacity building materials and the concept for webinars and training events will be developed at a later stage in the project, mainly based on the results and methodologies applied in the co-creation events.

34 Technical partners in the DE-BIAS project are Datoptron and Thinkcode.

compiled vocabulary will be key. It will be created as a multi-layered process by (1) including scientific research on patterns of bias; (2) building on existing glossaries of harmful terms; and (3) involving the communities themselves in co-creation and validation activities. This vocabulary is intended to remain "dynamic" throughout the project, and will be expanded, adapted, and reworked continuously following the insights and experiences gained.

In the field of cultural heritage, the previously mentioned *Words Matter* publication is a seminal reference on issues of diversity and representation in museums and archives. The glossary also serves as a reference for DE-BIAS, because it has already been transformed into a structured vocabulary (see section *Building the DE-BIAS Tool* below).

Another resource that we are considering for building the DE-BIAS vocabulary, which relates to the thematic background of "ethnicity and ethno-religious identity," is Ronen Steinke's "Antisemitismus in der Sprache" (Antisemitism in Language). Steinke draws attention to historically charged expressions and shows how antisemitic thinking is transported through everyday language. For example, Yiddish expressions such as "mischpoke" (mischpóche) have found their way into German but, unlike the originally neutral meaning of "Familie" (family), "mischpoke" often has a negative connotation when used in German.[35]

An objective of the project is to build the vocabulary in five languages.[36] The initial (and rather naive) idea was to provide a certain number of critical terms and then translate them into the other languages. Taking the Yiddish expressions as an example, straightforward translation into other languages is not possible. Another issue the project is facing is that a term can be harmful in one language, but not in another one. The term "handicapped" is considered detrimental in some languages, but not in French, for example.

Community Engagement and Co-Creation Workshops

Co-creation events with communities constitute the methodological backbone of the whole project.[37] Starting in October 2023, consortium partners will organize workshops with the LGBTQ+ community in Italy and the UK (European Fashion Heritage Association); the Congolese community in Belgium and the Republic of

35 Ronen Steinke, *Antisemitismus in der Sprache: warum es auf die Wortwahl ankommt* (Berlin: Dudenverlag, 2022).

36 The vocabulary will cover terms in English, French, Dutch, German, and Italian.

37 Katholieke Universiteit Leuven (KU Leuven) is the leading partner for co-creating the DE-BIAS vocabulary and community engagement.

Congo (KU Leuven); the Jewish cultural studies and Jewish film studies research community in Germany (DFF – Deutsches Filminstitut & Filmmuseum); and the Surinam community in the Netherlands (Netherlands Institute for Sound & Vision).

Possible scenarios for these workshops are: (1) selecting potentially harmful terms from descriptive metadata, either directly from Europeana collections or from the collections of one of the institutions responsible for organizing the workshops; (2) sharing thoughts and stories about whether these terms are problematic and in what context; (3) exploring the provenance, history, and semantics of a term, and reflecting on the context in which a particular word appears and how its meaning and use has changed over time. Another approach followed by KU Leuven and the Netherlands Institute for Sound & Vision is to invite the communities to describe the content from their perspectives and knowledge. One can assume that the descriptions and terms used by the communities will differ from those recorded in the institution's database to describe the content. There is no one-size-fits-all approach to working with communities, so the collaboration scenarios will be different for each of the workshops.

These considerations will translate into the contextualization of the DE-BIAS vocabulary, explaining the bias/harm of a particular word, addressing the use of the term in its original historical and social context, and suggesting appropriate alternatives, not only for the end users of the site, but also to inform better practices in Cultural Heritage Institutions wishing to update their metadata approach, or to help them *debias* existing descriptions. The community work will be based on existing best practices and methodologies for such participatory activities, mainly informed by the project partner ECCOM.[38] The methodology is work in progress and will be refined during the course of the project, with input from the experiences and outcomes of the co-creation activities.

Building the DE-BIAS Tool

To turn the vocabulary into a machine-readable format, it will be structured as a knowledge graph. In the example below (Figure 3) the term "dwarf" is identified as a contentious label; the appropriate label would be "someone with dwarfism." Knowledge graphs capture relationships and nuances between terms, such as how a term might be problematic in one context but neutral in another. For ex-

38 Cristina da Milano, Elisabetta Falchetti, and Maria Francesca Guida, eds., *Intercultural Rehearsals* (Milano: Editrice Bibliografica, 2019).

ample, the term "exotic" is appropriate when describing plants or species, but not for describing human beings.

Knowledge graphs are multidimensional, they represent multiple facets of a term: synonyms, antonyms, related biases, and more. They are scalable and flexible to incorporate new information or data sources without disrupting existing structures. Via the knowledge graph the DE-BIAS vocabulary can be connected to other sources such as Wikidata or thesauri that are widely used in the cultural heritage sector, such as the Getty Art and Architecture Thesaurus (AAT). The DE-BIAS knowledge graph largely follows the *Words Matter* structure which is available as Linked Open Data.[39] The DE-BIAS tool combines techniques from Natural Language Processing (NLP) and Knowledge Graphs to create a data-driven workflow for detecting and flagging the use of outdated and offensive terms in the collection descriptions of Europeana.[40] It will function in such a way that it parses metadata records and identifies these terms, based on the DE-BIAS vocabulary.[41]

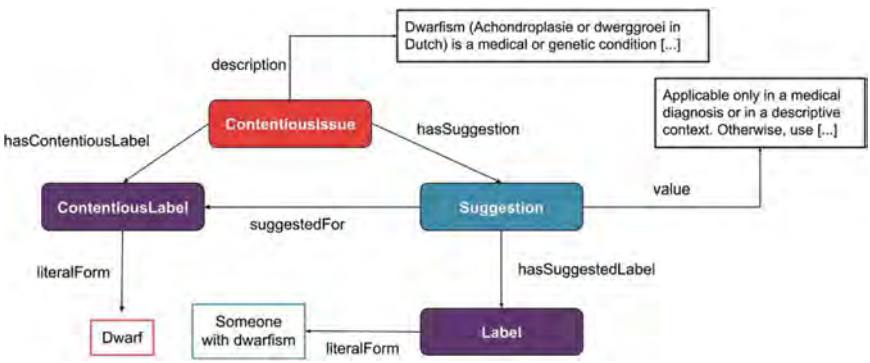


Figure 3: Example taken from the Words Matter Knowledge Graph, https://github.com/cultural-ai/wordsmatter/blob/v1.0.0/glossary.ttl, CC-BY-SA (accessed April 11, 2024).

The bias detection tool can also be used as a standalone tool by cultural heritage institutions and Europeana data providers before publishing their data in Europeana. By doing so, they can use the DE-BIAS vocabulary to analyze their datasets

39 Andrei Nesterov et al., "Cultural-Ai/Wordsmatter: Words Matter: A Knowledge Graph of Contentious Terms," *Zenodo* (2023), https://doi.org/10.5281/zenodo.7713157.

40 NLP uses algorithms such as lemmatization, smart string matching, named entity recognition, and disambiguation to enable context related analysis of words and the exclusion of terms that are named entities (e.g., "Gypsy Kings") and not indicative of bias.

41 The DE-BIAS tools will support the languages of the DE-BIAS vocabulary: English, French, Dutch, German, Italian.

and see whether they contain biased terms. It will generate a report that includes a statistical overview of the bias analysis and the flagged records. A user interface component (tooltip) on the Europeana Collections website will display the detected terms and provide contextual information explaining to the user why these terms have been flagged. This UI component will be designed in several iterations involving the validation by users of the Europeana Website, the community of experts involved in the project, aggregators, and CHIs representatives.

Validation of the developed tool will be conducted through crowdsourcing events on *CrowdHeritage*.[42] The open platform invites cultural heritage institutions to share their collections' metadata for enrichment, and everybody can contribute to improve them. The validation process for the DE-BIAS terms will involve Europeana aggregators and community representatives. Not only will this make it easier and faster to locate relevant terms, but it will also allow cultural heritage professionals and communities to discuss possible alternatives that adhere more appropriately to contemporary societal scenarios and are respectful of different sensibilities.

By doing so, it enables the human validation of the results of the automatic bias detection on metadata records. Users can not only validate automatically detected problematic terms; but also indicate terms/phrases in metadata records they consider problematic. This feedback collected by crowd workers can provide useful insights about the perception of problematic terms by communities. Finding expressions that have not been automatically tagged can contribute to further enriching the bias vocabulary and improving the DE-BIAS tool. The human insights into contextual bias can also be helpful as training and test data for building more efficient bias detection algorithms.

It is important to include "humans in the loop" community representatives as alternative voices, both in the co-creation events, the design of the UI/UX component, and the validation, in order to effectively address the issue of biased language in collection descriptions. By doing so, DE-BIAS contributes to the development of AI technology that takes cultural and ethical values into account.

42 CrowdHeritage platform, accessed April 27, 2024, https://crowdheritage.eu/en.

Conclusion

As language is constantly changing, so is the concept of what cultural heritage means in different societies. Institutions need to rethink traditional approaches to the cultural heritage collections they preserve and work with in order to meet the expectations of communities calling for a more accurate and respectful representation.

In this chapter we have looked at different approaches to representing diversity in cultural heritage collections and, more specifically, to critically reflect on the language that has been used to describe these collections – whether at the level of the item, as in the case study of Dämonisches Afrika, or at the level of large datasets, as in the DE-BIAS project. Guidelines, glossaries, and tools already exist or are in the making to support CHIs in this endeavor. Raising awareness and building capacity in this area is crucial to sensitize cultural heritage professionals to the potential harm caused by language. Training programs equip professionals with the necessary skills to critically assess and reflect on the potential biases, stereotypes, or discriminatory language present in collections and enable them to make informed decisions about their engagement with cultural heritage. In order to ensure the implementation of activities that promote diversity in cultural heritage collections in the long term, CHIs need to be provided with adequate resources, supported by policy frameworks.

References

Brunow, Dagmar. "Curating Access to Audiovisual Heritage: Cultural Memory and Diversity in European Film Archives." *Image and Narrative* 18, no. 1 (2017): 101.

Herlt, Kerstin, Kristina Rose, and Kerstin Arnold. "Final Recommendations for Identifying and Managing Diversity in Cultural Heritage Data." For internal distribution only, n.d.

Kunst, Marijke. "Being True to the Catalogue." In *Words Matter, Work in Progress I*, edited by Wayne Modest and Robin Lelijveld, 29–32. National Museum of World Cultures, 2018. Accessed April 1, 2024. https://www.materialculture.nl/en/publications/words-matter.

Lellman, Charlotte. "Guidelines for Inclusive and Conscientious Description." *Center for the History of Medicine: Policies and Procedures Manual* (2020). Accessed April 1, 2024. https://wiki.harvard.edu/confluence/display/hmschommanual/Guidelines+for+Inclusive+and+Conscientious+Description.

Milano, Cristina da, Elisabetta Falchetti, and Maria Francesca Guida, eds. *Intercultural Rehearsals*. Milano: Editrice Bibliografica, 2019.

Modest, Wayne, and Robin Lelijveld, eds. *Words Matter, Work in Progress I*. National Museum of World Cultures, 2018. Accessed April 1, 2024. https://www.materialculture.nl/en/publications/words-matter.

Mulvey, Laura. "Visual Pleasure and Narrative Cinema." *Screen* 16, no. 3 (1975): 6–18.

Nesterov, Andrei, Laura Hollink, Marieke van Erp, and Jacco van Ossenbruggen. “Cultural-Ai/Wordsmatter: Words Matter: A Knowledge Graph of Contentious Terms.” *Zenodo* (2023). https://doi.org/10.5281/zenodo.7713157.

Peeren, Esther. “Language Cannot Be ‘Cleaned Up’.” In *Words Matter, Work in Progress I*, edited by Wayne Modest and Robin Lelijveld, 43–45. National Museum of World Cultures, 2018. Accessed April 1, 2024. https://www.materialculture.nl/en/publications/words-matter.

Schmidt, Annette. “Africa Is Not a Country.” In *Words Matter, Work in Progress I*, edited by Wayne Modest and Robin Lelijveld, 40–42. National Museum of World Cultures, 2018. Accessed April 27, 2024. https://www.materialculture.nl/en/publications/words-matter.

Shankar, Arjun. “Primitivism and Race in Ethnographic Film: A Decolonial Re-Visioning.” *Oxford Bibliographies: Anthropology* (2020): 1–25. https://doi.org/10.1093/obo/9780199766567-0245.

Steinke, Ronen. *Antisemitismus in der Sprache: warum es auf die Wortwahl ankommt*. Berlin: Dudenverlag, 2022.

The Cataloguing Lab Initiative. “List of Statements on Bias in Library and Archives Description.” September 2023. Accessed April 27, 2024. https://cataloginglab.org/list-of-statements-on-bias-in-library-and-archives-description/.

III Analysis

Christian Gosvig Olesen

Timelines of Scholarly Video Annotation: For a Tool Critical History of Digital Film Historical Scholarship

Introduction

In recent years, tool criticism has gained traction as an approach for evaluating the perspectives and potential biases of digital scholarship's tools. Emerging as a corrective to source criticism as the primary basis for evaluating historical scholarship's truth claims and empirical foundations, tool criticism argues for the necessity of also considering the epistemological foundations of the tools with which sources are processed.[1] Gradually, this perspective has evolved to identify key focus areas, such as interfaces and algorithms, as epistemic devices of which the operations, procedures, and forms of interaction reflect specific historical knowledge regimes.[2]

Tool criticism has been key in developing video annotation tools and related teaching materials for use in film history and theory teaching and research, in the context of the Dutch research and teaching infrastructure project the CLARIAH Media Suite – the central university-level access point for digitized audiovisual collections. The Media Suite is an integrated environment that combines access to digitized audiovisual collections and tools in a single place, allowing scholars and students to build their own corpora of archival materials and add commentary through annotation. Tool critical perspectives have been integrated into the video annotation tool's development at different stages throughout the process. It has entailed comparing the Media Suite's video annotation functionalities with other software by segmenting the same film with different software, proprietary as well as scholarly, to gain insights into how different tools afford different workflows and analytical engagements with digitized films.[3] It has involved evaluating the interoperability of research data resulting from different video annotation software, by

1 Marijn Koolen, Jasmijn Van Gorp, and Jacco van Ossenbruggen, "Toward a Model for Digital Tool Criticism: Reflection as Integrative Practice," *Digital Scholarship in the Humanities* 34, no. 2 (2018): 368–385.

2 Pelle Snickars, Andreas Fickers, and Mark J. Williams, "Editorial Special Issue Audiovisual Data in Digital Humanities," *VIEW Journal of European Television History & Culture* 7, no. 14 (2018): 1.

3 Liliana Melgar Estrada, Eva Hielscher, Marijn Koolen, Christian Olesen, Julia Noordegraaf, and Jaap Blom, "Film Analysis as Annotation. Exploring Current Tools and Their Affordances," *The Moving Image* 17, no. 2 (2017): 49. The film in question was MENSCHEN AM SONNTAG (Germany, 1930).

https://doi.org/10.1515/9783111082486-011

doing data cross-walks to critically consider the extent to which scholars may exchange data between tools.[4] Likewise, it has involved user considerations of how working in an integrated environment confines researchers to an archival perspective, and the extent to which scholars can challenge archival data and freely devise analytical approaches tailored to the complex specificities of the materials they engage with.[5]

Having been involved in these initiatives, I have realized that tool criticism primarily takes a presentist perspective, by almost solely comparing tools that still function and are in use. Seldom, if ever, does tool criticism involve comparing current digital scholarship with tools developed decades ago, in order to understand the shifting epistemologies of film history's digital tooling. In presenting and co-developing the Media Suite, however, I have routinely placed the video annotation tool within a longer history of hypermedia and database-driven film scholarship, highlighting for users how it builds on decades of digital scholarship.[6] In this regard, when highlighting commonalities and differences between contemporary and older digital scholarship, I feel frustrated that I cannot ground this in actual tool critical comparisons, as it is not yet common for tool criticism to include historical perspectives and hands-on engagement with older formats. On the one hand, this has to do with digital film historical scholarship's early formats being poorly preserved, on the other, with the circumstance that we lack histories of such formats. In my view, this constitutes a fundamental problem insofar as it complicates developing a deep understanding of how film historiography's shifting techniques condition film historical research, as called for by some scholars.[7] Reflecting on the Media Suite in relation to the history of hypermedia and video annotation in film studies, this chapter thus argues that tool criticism needs to become more historically oriented and should seek to establish comparative historiographical perspectives. To achieve this, my chapter makes a case for integrating media preservation perspectives and media archaeological practice with tool criticism.

4 Liliana Melgar Estrada and Marijn Koolen, "CLARIAH Expert Meeting on Video Annotation Interoperability," *The CLARIAH Media Suite* (blog), July 11, 2018, accessed March 19, 2024, https://mediasuite.clariah.nl/blog/2018/07/11/Clariah-annotation-expert-meeting.

5 Susan Aasman et al., "Tales of a Tool Encounter: Exploring Video Annotation for Doing Media History," *VIEW Journal of European Television History & Culture* 7, no. 14 (2018): 73–87.

6 Christian Gosvig Olesen, "MIMEHIST: Annotating Eye's Jean Desmet Collection," *TMG Journal for Media History* 24, nos. 1–2 (2021): 1–11.

7 Trond Lundemo, "Towards a Technological History of Historiography?" in *At the Borders of (Film) History: Temporality, Archaeology, Theories*, ed. Alberto Beltrame, Giuseppe Fidotta, and Andrea Mariani (Udine: Forum Editrice Universitaria Udinese SRL, 2015), 149–155.

Video Annotation for Film History and the Media Suite Learn Initiative

The Media Suite is the central Dutch university-level access point to digital AV collections and tools, and a hub for digital media studies scholarship in the Netherlands. It offers students and researchers access to the Netherlands Institute for Sound and Vision's entire public broadcast collections, and to film and film-related collections from Eye Filmmuseum, in particular the Jean Desmet Collection of early cinema. As an integrated environment, these collections can simultaneously be accessed and analyzed with a number of tools and data enrichments, ranging from visualization of word frequencies in collection metadata and Automatic Speech Recognition (ASR), to Optical Character Recognition (OCR) enrichments, to video annotation.[8] The video annotation tool is one of the most advanced features of the infrastructure. It allows users to create, save, and code segments of films, and make multiple thematic timelines that may be annotated and enriched with comments and links to contextual materials. Before entering the stage of annotating in the infrastructure's resource viewer, users first search for, bookmark, and create a corpus of archival materials from different collections, relying on institutional metadata, or on data enrichments that are only available in the Media Suite's environment (for instance, OCR of typewritten documents, or ASR of broadcast materials).

In the context of the Media Suite's teaching and training initiative *Learn*, I and the team I am working with have developed tutorials that introduce users to applying video annotation in film studies research and in teaching at BA and MA levels. The tutorials also serve as a general introduction to newcomers who wish to conduct research using video annotation tools in other disciplines also. The tutorial "Video Segmentation, Annotation and Structuralist Film Analysis" (2021) covers the basics of structural film analysis, based on Raymond Bellour's classic articles on film segmentation (and film segmentation in a broader sense), while the tutorials "Searching, Annotating and Linking for Film Historical Research" (2020) and "Visual Analysis and Historical Source Criticism of Desmet's Film Posters and Business Documents" (2021) focus on film historical research based on film–related sources, such as promotional materials or business documents.[9]

---

8 Nanne van Noord, Christian Olesen, Roeland Ordelman, and Julia Noordegraaf, "Automatic Annotations and Enrichments for Audiovisual Archives," in *ICAART 2021: Proceedings of the 13th International Conference on Agents and Artificial Intelligence: February 4–6, 2021 – Volume 1: ARTIDIGH 2021*, ed. A. P. Rocha, L. Steels, and J. van den Herik (Setúbal: SciTePress, 2021), 633–640.

9 The tutorials can be accessed via the Media Suite Learn platform: https://mediasuite.clariah.nl/learn, accessed March 19, 2024.

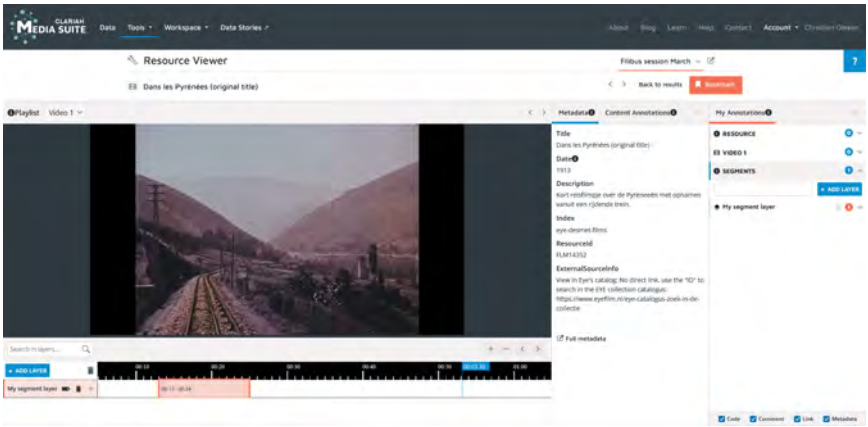


Figure 1: Video annotation in the Media Suite.

For the former, we use the tool to teach how films can be segmented in numerous different ways, starting with a classic example such as Griffith's The Lonedale Operator (1912), of which a copy is preserved in the Eye Filmmuseum's Jean Desmet Collection, by letting students annotate in groups and subsequently compare segmentations. Through this exercise, students and researchers develop a sense of how film segmentation can be a highly open-ended and subjective process, which often does not lead to the same result, yet may support an interpretation process in highly productive ways all the same. In this case, we let students and researchers browse through the materials of the Jean Desmet Collection's various subcollections, including films, posters, business papers, and promotional materials to, for instance, trace the acquisition of a film or to make a short screening and distribution history based on the materials (see Figure 1). In doing so, students and researchers can develop a sense of early cinema's commercial networks in the Netherlands and internationally, and of how commercial visual strategies of film exhibition evolved, while they learn to gather and curate a selection of visual materials on which to base their arguments.

The Media Suite and the History of Multimedia Scholarship in Film Studies

When introducing the annotation tool at workshops, it is important for me to make users aware that the video annotation tool has not emerged from a vacuum, but builds on decades of digital scholarship and publishing. I do so to instill a

sense of methodological relevance by showing that scholars have been developing such tools for decades, and to demonstrate how such tools are part of and shaped by a continuing scholarly discussion in film studies. In this respect, I highlight how the tool's design and functionalities build on a long tradition of hypermedia publishing, enriched editions, and video annotation projects in film and media scholarship from the 1990s, 2000s, and 2010s, and more recent qualitative analysis software.[10] I stress how the tool shares this tradition's ambition to give scholars the possibility to create their own pathways through digitized films and develop their own interpretations, by playing with the segmentation and temporal ordering of clips from the same or multiple items, and carefully annotating, linking, and presenting a selection of archival materials to others. At the same time, I also emphasize how such tools have become gradually more integrated with archives. This introduction embeds the tool in a history that I break down into three (roughly) periodized stages:[11]

1. Cinephile, commercial hypermedia editions from the mid-1980s to the early 1990s;
2. Academic, database-driven hypermedia publishing from the mid-1990s to the late 2000s;
3. Academic and archival video annotation software from the late 2000s until now.

The first period covers cinephile and commercial hypermedia editions such as the Voyager Company's early prestigious Criterion Collection laserdiscs containing contextual scholarly commentary.[12] Including an example of an early laser disc interface from the 1980s, showing a film clip accompanied by expert commentary, hammers home the point that video annotation in combination with film segmentation has been a desirable part of film critics' and scholars' methodological toolkit for close on forty years.

The second period covers early academic hypermedia publishing and experiments with database narratives and historical-critical edition formats, and the

10 For an example of how these examples have been used to contextualize efforts to integrate film collections in the Media Suite, see https://mediasuite.clariah.nl/blog/2017/08/31/Introducing-MIMEHIST, accessed November 28, 2023.

11 This section offers a condensed historical overview of multimedia scholarship which I develop in more detail in the framework of a broader epistemological discussion in my book *Visualizing Film History: Film Archives and Digital Scholarship* (Bloomington: Indiana University Press, 2025).

12 Robert Fischer, "The Criterion Collection: DVD Editions for Cinephiles," in *Celluloid Goes Digital: Historical-Critical Editions of Films on DVD and the Internet. Proceedings of the First International Trier Conference on Film and New Media, October 2002*, ed. Martin Loiperdinger (Trier: WVT Wissenschaftlicher Verlag Trier, 2003), 104.

gradual turn to film archives for source materials. It highlights this period's efforts to develop and integrate multimedia formats, digitized archives and databases into film studies research and teaching for purposes of citation, illustration, contextualization, and analysis of film excerpts with greater ease, playing with regimes of navigation inspired by commercial home-video releases and media art.[13] This period highlights mainly US-based, *auteurist* CD-ROM projects, centered around Hitchcock's *oeuvre*. This includes Lauren Rabinovitz' *The Rebecca Project* (1995), which analyzed and contextualized *Rebecca* (1940) through a feminist lens, or Robert E. Kapsis' Hitchcock centenary project *Multimedia Hitchcock* (1999), both groundbreaking in their presentation of film excerpts with commentary and contextual film-related sources.[14] Other projects are highlighted for their experiments combining close analysis with data-driven analysis, such as Bertrand Augst and Brian O'Connor's highly innovative CD-ROM project on THE BIRDS (2002) and Stephen Mamber's *Digital Hitchcock*. The former took Bellour's classic close analysis of THE BIRDS' Bodega Bay sequence as a starting point for automatically extracting key frames and data on various image features to enable new approaches to moving image retrieval.[15] *Digital Hitchcock*, also focusing on THE BIRDS, relied on a grid visualization containing one frame from every shot, in a web-based project and installation format allowing for navigating Hitchcock's film as a database in relation to his storyboards.[16] Subsequently, I stress how projects, such as MIT's *The Virtual Screening Room* (1992–1999) or the Labyrinth project's *Cine-Discs*, arguably gestured towards an integration of archives and annotation tools in ways prefiguring the Media Suite. The former, developed by Henry Jenkins, Ben Singer, Ellen Draper, and Janet Murray as an "interactive multimedia 'textbook'" introduction to key classic film theoretical and historical approaches, contextualized and made available 500 excerpts while offering close analyses of a handful of selected films.[17] *Cine-Discs*' releases, such as Marsha Kinder's *Blood Cinema* (1994) on 1940s and 1950s Spanish Cinema and Yuri Tsivian's *Immaterial Bodies: Cultural Anatomy of Early Russian Films* (2000), centered on clearly defined historical periods.

13 Robert Kolker, "Digital Media and the Analysis of Film," in *A Companion to Digital Humanities*, ed. Susan Schreibman, Ray Siemens, and John Unsworth (Hoboken, New Jersey: Blackwell Publishing, 2004), 384.

14 Ben Singer, "Hypermedia as a Scholarly Tool," *Cinema Journal* 34, no. 3 (1995): 86–91.

15 Betrand Augst and Brian O'Connor, "No Longer a Shot in the Dark: Engineering a Robust Environment for Film Study," *Computers and the Humanities* 33, no. 4 (1999): 352.

16 Stephen Mamber, "Space-Time Mappings as Database Browsing Tools," in *Media Computing: Computational Media Aesthetics*, ed. Chitra Dorai and Svetha Venkatesh (Boston: Springer, 2002), 44.

17 See Virtual Screening Room, "Beta Testing," accessed November 29, 2023, https://web.archive.org/web/20100612112856/http://caes.mit.edu/projects/virtual_screening_room/1beta_testing.html.

While different in scope, these projects were trailblazing in how they in similar ways deployed hyperlinked environments and video annotation for browsing, linking, contextualizing, and analyzing digitized films and related materials, while allowing users to develop non-linear, personalized pathways through materials. For instance, the Cine-discs enabled personalized pathways and facilitated annotation by allowing users to add and export commentaries in Notepad.[18] In this respect, such projects took inspiration from contemporary media art CD/DVD-ROM projects by prominent filmmakers and multimedia artists such as The Residents, Chris Marker, Michael Snow, and Pat O'Neill. The latter's CD-ROM-project *Tracing the Decay of Fiction* (2002), made as part of the Labyrinth project, stands as one of the better-known cross-overs of media art and multi-media scholarship from this period.

The third period I defined runs up to the present day. It covers academic and archival video annotation software and sketches a development towards greater emphasis on analytical and interpretative processes and the integration of tools and archives rather than publication, premised on affording students and scholars greater access to and freedom in shaping tools and analyses made with digital means. This includes the Centre Pompidou's *Lignes de Temps* software, which, based in apparatus theory, invited scholars to challenge films' spectatorial subject positioning by taking on the role of film editor, rearranging and making personal cuts of well-known films.[19] Software such as Videana, Advene, Anvil, ELAN, and VIAN, expanded the range of options for adding subjective layers of annotation with a still higher level of granularity, while more recent tools such as The Media Ecology Project's (MEP) Semantic Annotation Tool (SAT), and the associated Kinolab.org, have sought to complement archival metadata through integration with archival workflows, an ambition also shared by larger projects such as I-Media-Cities project and the VIAN tool developed by Barbara Flückiger's team.[20]

Through this historicization, briefly sketched here, users of the Media Suite gain an impression of how video annotation tools have developed towards affording greater interpretive agency, while also gradually becoming integrated with

18 Yuri Tsivian, *Immaterial Bodies: A Cultural Analysis of Early Russian Films* (Los Angeles: University of Southern California, Annenberg Center for Communication, 1999).

19 Jean-Louis Comolli, "West of the Tracks: Continuing the Journey," *Szabzian*, May 29, 2019, accessed March 19, 2024, https://sabzian.be/article/west-of-the-tracks-continuing-the-journey#footnote1_o5ejkny.

20 Bret Vukoder and Mark Williams, "The Great War at Scale. New Opportunities for Provenance in World War I Collections at the National Archives (NARA)," in *Provenance and Early Cinema*, ed. Joanne Bernardi, Paolo Cherchi Usai, Tami Williams, and Joshua Yumibe (Bloomington: Indiana University Press, 2020), 156.

digital archives and infrastructures in a manner that involves negotiating, challenging, or complementing archival metadata as a starting point for analysis.

Epistemological Differences of the Timeline

While this overview is helpful in anchoring the Media Suite's video annotation tool in a longer history of digital film history, I increasingly wonder whether it forces an evolutionary narrative of continuity upon the tools discussed, which does a disservice to our tool critical understanding of their differences and the ways in which they have conditioned film historical research at different points in time. The core problem in this regard is that this account is not tested against a hands-on engagement with different formats, which would allow one to actually evaluate differences and commonalities in practice. Epistemologically speaking, once one begins taking a closer look, one can point to multiple differences in terms of how the tools' components are deployed and the production contexts from which they emanate, which the brief overview above does not highlight. For example, throughout the past decades, the features and centrality of the timeline as an analytical device has frequently been posited rhetorically as a distinguishing feature of video annotation where interpretative agency is determined and negotiated, to the extent of suggesting, as Janet Murray has put it with a McLuhanist pun, that "the medium is the method," the medium in this case being the timeline.[21]

Yet the meanings, configurations, and epistemological implications of the timeline in different video annotation projects diverge wildly, even when starting from the same conceptual departure point. Inspired by Raymond Bellour's segmentation approach (and involving Bellour in the tool's development and launch), *Lignes de Temps* posits the timeline as a means for subverting film's ideological framings. The software carries out automated shot boundary detection, creating an audio waveform for films loaded into the software, but it may subsequently be adjusted or challenged through personal manual segmentations, alternative rough cuts, and annotations. In offering the possibility of negotiating the temporal ordering imposed by montage and automated shot segmentations, the tool contributed, according to philosopher Bernard Stiegler – then director of the Centre Pompidou's Institut de recherche et d'innovation (IRI) where the tool was developed – to creat-

21 Janet Murray, *Inventing the Medium. Principles of Interaction Design as a Cultural Practice* (Cambridge, MA: MIT Press, 2011), 16.

ing a critical community where software may be challenged, and competing interpretations of media may emerge through annotation.[22]

Conversely, in their endeavors to translate a Bellourian approach to automated film segmentation at the end of the 1990s, film scholar and information theorist Brian O'Connor, in collaboration with Bertrand Augst, made the case for going beyond the shot as the primary analytical unit, to anchor the "concept of the time line" in "patterns of changes in the relationships of pixel values" or music, "rather than extractions based on mechanical demarcations (e.g., shot or cademe or word)."[23] In this case then, a timeline relying on shot segmentation as in *Lignes de Temps* was considered altogether undesirable, imposing an irrelevant analytical unit of interest as a conceptual departure point for film analysis. Beyond such contrasting examples emerging from the conceptual foundation of Bellourian film analysis, the Indian Cinema Foundation's database-driven annotation platform *Indiancine.ma*, based on the pan.do/ra software, offers a "Timeline View" comprising a broader array of visualizations that cover both image and sound as a basis for browsing and navigating digitized Indian classics. Using this view, the temporal unfolding of each video file of a film in the database can be navigated through four different modes: "Anti-alias," "Slit-Scan," "Keyframes," and "Waveform."[24] As in O'Connor and Augst's experiment, *Indiancine.ma* defines the timeline as a device for navigation, but one which imposes a primarily linear view of the films analyzed.

Such differences also extend to the production context of video annotation tools. The 1990s' multimedia formats' experimentation with timelines, in particular the emphasis on exploratory browsing and non-linearity in CD-ROM projects, reflected an encounter between scholars and emerging modes of digital storytelling and interactive documentary and media art. In her presentation of the Labyrinth project, Kinder has, for instance, emphatically stressed how, in addition to being a research initiative and website, it was also an art collective.[25]

One could continue listing different historical configurations and definitions of the timeline in digital film historical scholarship, but the basic point here is that, as a central feature of video annotation, it has been a host for envisioning different conceptions of film analytical – and by the same token historical – interpretation of digitized films and archives. Placed within a broader genealogy of video annotation, and looking beyond the particular example of the timeline, it

22 Bernard Stiegler, "Pharmacologie de l'*épistémè* numérique," in *Digital Studies. Organologie des savoirs et technologies de la connaissance*, ed. Bernard Stiegler (Roubaix: FYP éditions, 2014), 25–26.

23 Augst and O'Connor, "No Longer," 354–355.

24 See https://indiancine.ma/timelines/year/director==Sasikumar, accessed November 29, 2023.

25 See http://www.marshakinder.com/multi/m4.html, accessed November 29, 2023.

becomes clear how, in comparison with these projects, the Media Suite in many ways embodies various broader shifts in digital scholarship that we may not yet have fully considered. These shifts entail video annotation going from being embedded in artistic and maker-based experimentation to infrastructural thinking; from the standalone configuration of disc-based projects to networked and potentially more dynamic, open-ended engagements with materials in an environment where new data are continuously added; from limited, curated subsets of collections to large-scale collections and data; from non-linear exploration of digital items to targeted searches in institutional metadata and data enrichments prior to annotation; from shot-based segmentation to bottom-up, qualitative segmentation. Digging deeper into historical definitions and configurations of video annotation projects, we thus see how they imply different forms of interactivity, participation, levels of agency, modes of searching, browsing, and segmentation. Despite only just scratching the surface of such a genealogy here, we have quickly realized how many significant shifts we can identify, so it is all the more surprising that we are not yet comparing them to each other in a tool critical perspective, and have not yet started a conversation around this particular topic, taking a historical view.

Metahistory and the Preservation Problem of Digital Film Historical Scholarship

To develop foundations for a more historically oriented tool criticism is first a matter of establishing a metahistorical perspective on film studies' digital methods, in order to facilitate a comparative historiography of them. In many ways, such an ambition may be considered akin to the ways in which scholars discussed different modes of historical writing in the 1970s, taking inspiration from contemporary theories of history and literary theory. Consider, for instance, how Edward Branigan's classic article, "Color and Cinema: Problems in the Writing of History" (1979), pondered different modes of film color historiography. Written at the height of Hayden White-inspired meta-historical discussions, it meticulously compared significant film historical publications from different periods, by classifying them into various modes of history characterized by distinct temporal orderings and emplotments. At the present moment, our discussion of digital film historical scholarship is not yet premised on the assumption that it would be useful to compare contemporary and historical formats in order to sketch a development of their different knowledge regimes. This becomes painfully visible when we consider how difficult it already is to access and use "old" digital research formats and publications, and how quickly they have fallen prey to technological obsolescence.

To make a crude comparison in media theoretical terms, imagine not being able to sit down, read, and critique a key film historical monograph published just twenty or thirty years ago. Essentially, this is the situation one is confronted with when trying to access early digital film historical scholarship produced just two or three decades ago – a situation which so far seems to be broadly accepted in film and media studies. It is striking how few digital scholarly projects in film and media studies – from the 1990s CD-ROM projects to late 2000s annotation software – have been maintained or preserved in a sustainable fashion. Currently, a lack of hardware, software, developed infrastructures, and awareness of digital scholarship's preservation challenges stand in the way of developing a comparative historiography of digital projects. Yet the intellectual work underlying analytical processes and results produced by digital film historical scholarship, as this volume testifies, is just as valuable as that of established written scholarship, and we need to be able to evaluate digital scholarship, just as with any other type of scholarship, in order to argue for its place within the academy.

If it is any consolation, more established types of digital scholarship, in particular literary studies' history of electronic publishing and database-driven scholarship, while further ahead in thinking through issues of preservation, are facing similar challenges.[26] It has been neglected at the expense of printed scholarship; Kathleen Fitzpatrick has identified several key challenges for literary studies' digital scholarship, notably its often project-based nature that seldom favors appropriate preservation measures to be taken, and its lack of research data standards and project documentation, of accessibility, and of infrastructures similar to the ones sustaining printed scholarship, all of which we can equally apply to digital film historical scholarship.

In this regard, it is a problem that the emergence of digital film historical scholarship is seldom historicized and not subject to tool criticism retrospectively. With few exceptions, we do not see digital methods for film and media studies in a longer historical perspective, and do not, in the broader digital humanities field, have historical accounts of digital film historical scholarship's emergence. Arguably, Digital Humanities has become over-saturated with historical accounts and is sidestepping the pursuit of writing them for this reason: think, for instance, of Anne Burdick, Johanna Drucker, Peter Lunenfeld, Todd Presner, and Jeffrey Schnapp's remark that the introductory publication *Digital_Humanities* (2012) is "not a [. . .] research repor[t] on the history of, or critical engagement with, the Digital Humanities. Neither is it a textbook from which to teach the discipline's foundations [. . .]

26 Kathleen Fitzpatrick, *Planned Obsolescence: Publishing, Technology, and the Future of the Academy* (New York: NYU Press, 2013), 121.

its protagonists, successes and failures, and defining moments."[27] One may see where they are coming from, insofar as origin points are always slippery and histories of founding fathers to be avoided. But for film and media studies, I contend that we do at this moment need (tool) critical histories of our own digital methods, in order to understand how shifting techniques, technologies, and development of tools have conditioned film historical research in the past.

Three Suggestions

Where can such an endeavor start? In this concluding section, I offer three suggestions, that might be considered initial key steps towards a tool critical comparative historiography.

Collecting and Preserving

As our historical understanding of audiovisual collections will increasingly be shaped by digital scholarly formats, it becomes urgent to ensure that university institutions acknowledge them as valuable pieces of scholarship, and that such work remains accessible. This is no trivial issue if we truly believe that digital film historical scholarship holds a future potential for teaching, research, and collection access. Thus, one significant step is to create collections and to dedicate resources to preserving digital film historical scholarship. Currently, preservation of digital scholarship focuses primarily on research data, but should also comprise preservation of scholarly hardware and software. For my own part, I have been building a small collection of disc-based projects – mainly CD-ROMs – and am developing an overview of such projects. In this process, I take inspiration from relevant media art conservation projects such as Sandra Fauconnier's *CD-ROM Cabinet* project (2013) – which extensively documents and develops preservation strategies for CD-ROMs.[28] Needless to say, however, such endeavors should go beyond personal initiatives and aim to become embedded in coordinated preservation efforts.[29] In this respect, tool criticism may have a lot to learn from media archaeological preservation initiatives that place an emphasis on machinery within a broader media theoretical and historical frame. The Humboldt Uni-

27 Anne Burdick, Johanna Drucker et al., *Digital_Humanities* (Cambridge, MA: MIT Press, 2012), ix.
28 See http://aaaan.net/the-cd-rom-cabinet-after-6-months/, accessed November 29, 2023.
29 Fitzpatrick, *Planned*, 59.

versity's Media Archaeological Fundus, the Media Archaeology Lab at the University of Colorado, or the Doing Experimental Media Archaeology project at the University of Luxembourg are interesting examples, insofar as they facilitate access to historical media. While conceived for engagements with the materiality of historical media they are, by the same token, sites where digital film historical scholarship may be accessed. Digital humanities labs, from which digital scholarship emanates, should be equally invested in nurturing such engagement with past decades' digital scholarship.

Thinkering

If digital scholarly initiatives can learn much from media archaeology laboratories' commitment to the preservation of hardware and software, they can equally learn from these places' conceptual foundations insofar as they invite hands-on engagements with media objects as epistemic objects. Currently, Tool Criticism and Media Archaeology are conceptually two largely separate activities but should, if we wish to nourish cross-fertilization between their perspectives, be combined by any means possible. As I have argued, comparing new and old digital film historical scholarship needs to become an integral part of how we evaluate the epistemologies of digital methods, just as we apply a longer historical perspective when comparing different periods of film history writing. In this respect, the media archaeological approach of "thinkering" – a term taken from Erkki Huhtamo's media archaeology and serving as a conceptual departure point for "playful experimentation with digital tools and technologies for historical research" in the context of the University of Luxembourg's Digital History Lab – could be seen as offering a highly relevant tool critical approach that may be productively expanded to include digital film historiography's early digital formats.[30] The value of trying to capture the experience of accessing and understanding the types of engagements with (archival) audiovisual materials that older formats afford, and how they position one as a user, is fundamental. Personally, I have long dreamed of organizing a workshop where, in combination with outlining a history of projects, participants could try out historical and contemporary formats hands-on, next to each other, and discuss differences, reflecting on the underlying epistemologies of different formats. Again, beyond the realms of media art conservation (where this comes across as a much less esoteric proposition),

30 Andreas Fickers and Tim van der Heijden, "Inside the Trading Zone: Thinkering in a Digital History Lab," *Digital Humanities Quarterly* 14, no. 3 (2020), accessed March 19, 2024, http://www.digitalhumanities.org/dhq/vol/14/3/000472/000472.html.

there are currently few (if any) places to carry out such an experiment, yet it is exactly the type of approach we should embrace as a vital part of a historically informed tool criticism.

Documenting

Concurrently with building up a small collection of disc-based media, I have created an archive of the Media Suite's development since its beginnings, comprising old video tutorials, documentation, and web pages. As far as possible, we have collected screen recordings made for tutorials in order to create a basis for comparison between the Media Suite and early and contemporary formats, and to be able to critically compare modes of navigation and interaction with audiovisual items, based on which videographic works can be made. I do so because I wish others had done the same for the multimedia projects they developed throughout the 1990s and 2000s. In several instances, one may still gain an impression of older projects by digging deep into the Internet Archive – for instance, to get an impression of what *The Virtual Screening Room*'s webpage looked like.[31] Yet, one cannot avoid being left with the feeling that such documentation should have been undertaken more systematically. Just as collecting and preserving is necessary, so is documentation of the use of older formats in action, to be able to get an impression of their workings. In this regard, one may find an inspirational example in Marina Hassapopoulou's extensive research on screen interactivity, which, alongside other things, involves looking at, documenting, and writing blog posts about CD-ROMs, such as discs released in Marsha Kinder's *Cine-Discs* series with students. The resulting Interactive Media Archive, while not invested in a tool critical perspective for evaluating different epistemological foundations of digital tools, is helpful insofar as it considers digital scholarship in the context of broader histories of media and interactivity in a teaching context, and insofar as it leaves traces that may be considered documentation of how older formats worked.[32] In addition to collecting, preserving, and thinkering with such formats, documentation should become a key activity too in both research and teaching to open a conversation on how the tools that we are building today reflect epistemological shifts compared to those that were built yesterday.

31 Virtual Screening Room, "Beta Testing."

32 See The Interactive Media Archive, accessed February 23, 2023, https://interactivemediaarchive.wordpress.com/about/.

References

Aasman, Susan, Liliana Melgar Estrada, Tom Slootweg, and Rob Wegter. "Tales of a Tool Encounter: Exploring Video Annotation for Doing Media History." *VIEW Journal of European Television History & Culture* 7, no. 14 (2018): 73–87.

Augst, Bertrand, and Brian O'Connor. "No Longer a Shot in the Dark: Engineering a Robust Environment for Film Study." *Computers and the Humanities* 33, no. 4 (1999): 345–363.

Burdick, Anne, Johanna Drucker, Peter Lunenfeld, Todd Presner, and Jeffrey Schnapp. *Digital_Humanities*. Cambridge, MA: MIT Press, 2012.

Comolli, Jean-Louis. "West of the Tracks: Continuing the Journey." *Szabzian*, May 29, 2019. Accessed March 19, 2024. https://sabzian.be/article/west-of-the-tracks-continuing-the-journey#footnote1_o5ejkny.

Estrada, Liliana Melgar, Eva Hielscher, Marijn Koolen, Christian Olesen, Julia Noordegraaf, and Jaap Blom. "Film Analysis as Annotation. Exploring Current Tools and Their Affordances." *The Moving Image* 17, no. 2 (2017): 40–70.

Estrada, Liliana Melgar, and Marijn Koolen. "CLARIAH Expert Meeting on Video Annotation Interoperability." *The CLARIAH Media Suite* (blog), July 11, 2018. Accessed March 19, 2024. https://mediasuite.clariah.nl/blog/2018/07/11/Clariah-annotation-expert-meeting.

Fickers, Andreas, and Tim van der Heijden. "Inside the Trading Zone: Thinkering in a Digital History Lab." *Digital Humanities Quarterly* 14, no. 3 (2020). Accessed March 19, 2024. http://www.digitalhumanities.org/dhq/vol/14/3/000472/000472.html.

Fischer, Robert. "The Criterion Collection: DVD Editions for Cinephiles." In *Celluloid Goes Digital: Historical-Critical Editions of Films on DVD and the Internet. Proceedings of the First International Trier Conference on Film and New Media, October 2002*, edited by Martin Loiperdinger, 99–108. Trier: WVT Wissenschaftlicher Verlag Trier, 2003.

Fitzpatrick, Kathleen. *Planned Obsolescence: Publishing, Technology, and the Future of the Academy*. New York: NYU Press, 2013.

Kolker, Robert. "Digital Media and the Analysis of Film." In *A Companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens, and John Unsworth, 383–396. Hoboken, NJ: Blackwell Publishing, 2004.

Koolen, Marijn, Jasmijn Van Gorp, and Jacco van Ossenbruggen. "Toward a Model for Digital Tool Criticism: Reflection as Integrative Practice." *Digital Scholarship in the Humanities* 34, no. 2 (2018): 368–385.

Loiperdinger, Martin, ed. *Celluloid Goes Digital: Historical-Critical Editions of Films on DVD and the Internet. Proceedings of the First International Trier Conference on Film and New Media, October 2002*. Trier: WVT Wissenschaftlicher Verlag Trier, 2003.

Lundemo, Trond. "Towards a Technological History of Historiography?" In *At the Borders of (Film) History: Temporality, Archaeology, Theories*, edited by Alberto Beltrame, Giuseppe Fidotta, and Andrea Mariani, 149–155. Udine: Forum Editrice Universitaria Udinese SRL, 2015.

Mamber, Stephen. "Space-Time Mappings as Database Browsing Tools." In *Media Computing: Computational Media Aesthetics*, edited by Chitra Dorai and Svetha Venkatesh, 39–55. Boston: Springer, 2002.

Murray, Janet. *Inventing the Medium: Principles of Interaction Design as a Cultural Practice*. Cambridge, MA: MIT Press, 2011.

Noord, Nanne van, Christian Olesen, Roeland Ordelman, and Julia Noordegraaf. "Automatic Annotations and Enrichments for Audiovisual Archives." In *ICAART 2021: Proceedings of the 13th*

International Conference on Agents and Artificial Intelligence: February 4–6, 2021 – Volume 1: ARTIDIGH 2021, edited by A. P. Rocha, L. Steels, and J. van den Herik, 633–640. Setúbal: SciTePress, 2021.

Olesen, Christian Gosvig. "MIMEHIST: Annotating Eye's Jean Desmet Collection." *TMG Journal for Media History* 24, nos. 1–2 (2021): 1–11.

Singer, Ben. "Hypermedia as a Scholarly Tool." *Cinema Journal* 34, no. 3 (1995): 86–91.

Snickars, Pelle, Andreas Fickers, and Mark J. Williams. "Editorial Special Issue Audiovisual Data in Digital Humanities." *VIEW Journal of European Television History and Culture* 7, no. 14 (2018): 1–4.

Stiegler, Bernard. "Pharmacologie de l'*épistémè* numérique." In *Digital Studies. Organologie des savoirs et technologies de la connaissance*, edited by Bernard Stiegler. Roubaix: FYP éditions, 2014.

Tsivian, Yuri. *Immaterial Bodies: A Cultural Analysis of Early Russian Films*. Los Angeles: University of Southern California, Annenberg Center for Communication, 1999.

Vukoder, Bret, and Mark Williams. "The Great War at Scale. New Opportunities for Provenance in World War I Collections at the National Archives (NARA)." In *Provenance and Early Cinema*, edited by Joanne Bernardi, Paolo Cherchi Usai, Tami Williams, and Joshua Yumibe, 155–166. Bloomington: Indiana University Press, 2020.

Tim van der Heijden, Taylor Arnold, and Lauren Tilton

Distant Viewing the Amateur Film Platform

Introduction

This chapter explores the Amateur Film Platform (2014–2023), an online platform that hosted a unique collection of more than eight thousand films and videos made by Dutch amateur filmmakers in the twentieth century.[1] The Amateur Film Platform was initiated by the Netherlands Institute for Sound & Vision (NISV) in collaboration with various regional audiovisual archives in the Netherlands, including the City Archive Rotterdam and the Groningen Audio Visual Archive (GAVA). After the platform's launch in 2014, the Frisian Film Archive, Drenthe Audio Visual Archive (DAVA), Limburgs Museum, and Eye Filmmuseum also presented parts of their amateur collections on the platform, which made the platform diverse, geographically, institutionally, and historically – featuring amateur footage from the early 1900s until the 2010s from various technological carriers. The Amateur Film Platform predominantly presented digitized film-based collections, including 16mm, 9.5mm, 8mm, and Super 8 films, but digitized electronic videos and born-digital materials were also included, such as VHS, MiniDV, and web videos.[2] Moreover, via Community Uploads, users could upload their own (digitized) films and videos directly to the platform's portal, which made the platform function as a "living archive" of both historical and contemporary amateur media productions (Figure 1).[3]

Prior to the Amateur Film Platform's discontinuation in 2023, we have explored the platform's collection by utilizing the approach of Distant Viewing, which as a

1 See https://www.amateurfilmplatform.nl/, accessed September 29, 2022. The platform was online from 2014 until 2023. For more details about the platform's discontinuation and how to retrieve some of its collections, see https://www.beeldengeluid.nl/kennis/projecten/amateurfilmplatform, accessed March 22, 2024. Captures of the platform throughout the years can be found via the Internet Archive's Wayback Machine: https://web.archive.org/web/20230315000000*/https://www.amateurfilmplatform.nl/, accessed March 22, 2024.

2 For more details on these different amateur media and technological formats, see Tim van der Heijden, "Technologies of Memory: Amateur Storage Media and Home Movie Practices in the Longue Durée," *Le Temps Des Médias* 39, no. 2 (2022): 141–159, https://doi.org/10.3917/tdm.039.0141; Leo Enticknap, *Moving Image Technology: From Zoetrope to Digital* (London, New York: Wallflower, 2005). See also: Tim van der Heijden and Valentine Kuypers, "Life in Motion: A History of Amateur Film," *Europeana*, April 12, 2023, accessed August 8, 2024, https://www.europeana.eu/en/exhibitions/life-in-motion/.

3 Cf. Susan Aasman, "Finding Traces in YouTube's Living Archive: Exploring Informal Archival Practices," *TMG Journal for Media History* 22, no. 1 (2019): 35–55, https://doi.org/10.18146/tmg.435.

https://doi.org/10.1515/9783111082486-012

methodological framework applies computer vision methods to the computational analysis of large collections of audiovisual materials.[4] Distant Viewing enables us to work with large corpora of audiovisual materials to extract, aggregate, and visualize certain features or "semantic elements of visual materials," such as color, shot length, object detection, and camera movement.[5] It makes use of the Distant Viewing Toolkit (DVT), which is designed to facilitate the computational analysis of visual culture employing the latest machine learning and computer vision techniques.[6] The software, published as an open-source library, consists of several Python packages of computer vision algorithms, which can be used to navigate audiovisual collections and analyze their features at scale.[7] What makes Distant Viewing distinct from other digital approaches to visual culture is that it makes "explicit the interpretative nature of extracting semantic metadata from images," reflecting on how the computational viewing of audiovisual materials is informed by a certain code system.[8]

By applying the Distant Viewing approach and toolkit to the Amateur Film Platform, we aim to explore how it can provide new insights into this relatively large collection of amateur films and videos from the Netherlands. More specifically, we aim to investigate whether any formal, stylistic, and aesthetic patterns or changes over time could be found in the Amateur Film Platform collection, based on the data sample that we used for the analysis and, if so, how these patterns could contribute to a broader understanding of the history of amateur film as a cultural practice.[9]

4 Taylor Arnold and Lauren Tilton, *Distant Viewing: Computational Exploration of Digital Images* (Cambridge, MA: MIT Press, 2023); Taylor Arnold and Lauren Tilton, "Distant Viewing: Analyzing Large Visual Corpora," *Digital Scholarship in the Humanities* 34, issue supplement 1 (2019): i3–i16, https://doi.org/10.1093/llc/fqz013.

5 Arnold and Tilton, "Distant Viewing," i6.

6 Distant Viewing Toolkit, https://github.com/distant-viewing/dvt, accessed March 22, 2024.

7 For the software white paper, see Taylor Arnold and Lauren Tilton, "Distant Viewing Toolkit: A Python Package for the Analysis of Visual Culture," *Journal of Open Source Software* 5, no. 45 (2020): 1–6, https://doi.org/10.21105/joss.01800. For a review of the Distant Viewing Toolkit, see Patrick Sui, "Review: Distant Viewing Toolkit," *Reviews in Digital Humanities* 2, no. 4 (2021), https://doi.org/10.21428/3e88f64f.01fedc58.

8 Arnold and Tilton, "Distant Viewing," i4.

9 Van der Heijden conducted a historical discourse analysis and archival research to study the relationship between changing media technologies (film, video, digital media) and home movie practices from a long-term historical perspective. This media historical research informed the hypotheses of the tested relationships in the present study. From a methodological perspective, we were interested in the question of whether the Distant Viewing approach could confirm or provide alternative views on those relationships. See Tim van der Heijden, "Hybrid Histories: Technologies of Memory and the Cultural Dynamics of Home Movies, 1895–2005" (PhD diss., Maastricht University, 2018).

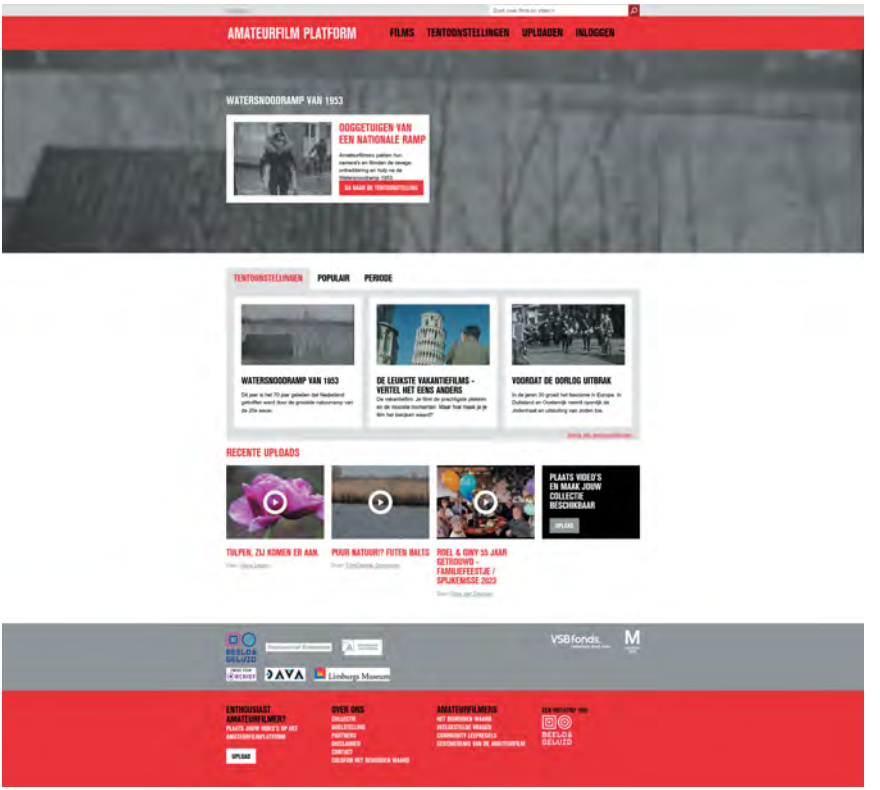


Figure 1: Homepage of the Amateur Film Platform, https://www.amateurfilmplatform.nl/ (accessed April 2, 2023).

Studying Amateur Media Collections at Scale

The Amateur Film Platform presented a large collection of digitized and born-digital amateur media content, produced by Dutch amateur filmmakers over different time periods and by means of various media technologies. The amateur films and videos were digitally searchable as part of a large dataset, which allowed for filtering based on predefined categories and other metadata. What new possibilities did this bring for studying the history of amateur film in the Netherlands? What patterns, relations, and changes over time could we analyze, for instance, in the use of certain topics, genres, narrative, or aesthetic tropes among other categories? In short, what would the potential and challenges be for exploring audiovisual collections of amateur films and videos at scale?

The platform was unique in the sense that it combined amateur collections from various audiovisual archives and presented them online for audiences to navigate and explore in various ways.[10] The user interface facilitated searching the collections by keywords in the search bar at the top, but users could also navigate the platform through various exhibitions that highlighted certain themes and films from the collection, or by means of the films themselves and their metadata. The different types of technological carriers were reflected in the metadata under the category CARRIER, which included multiple film-based formats like 16mm, 9.5mm, 8mm, and Super 8 film as well as video-based formats, such as VHS and MiniDV.[11] Furthermore, the platform allowed searching based on LOCATION (e.g., Amsterdam, Rotterdam, France), TIME PERIOD (from the 1900s until the 2010s), SUBJECT (e.g., amateur recordings, holiday, beaches, families, etc.), FILMMAKER (e.g., Piet Schendstok, Dick Laan, Roel van Deursen), and the archival COLLECTION the film or video items originally belongs to (e.g., NISV, GAVA). All these (sub)categories functioned as entities that could be clicked on to navigate the collection on the platform. To filter the results, categories could be combined in the search menu (e.g., selecting all 8mm films from the 1960s, or all films about the subject families from a specific filmmaker or collection) (Figure 2).

By foregrounding a digital approach to amateur film history, our project relates to other digital projects that involve large historical amateur film collections and archival materials, including projects from the Prelinger Archives,[12] the Bas-

10 Comparable initiatives in northern America include the Prelinger Archives (https://archive.org/details/prelinger, accessed March 22, 2024) and the Amateur Movie Database (an international platform celebrating the world of amateur cinema, which functions as a web-based tool for researching amateur films, filmmakers, and movie clubs from various international archives and collections: https://www.amateurcinema.org/, accessed March 22, 2024). For reflections on the Amateur Movie Database, see Charles Tepperman, "The Amateur Movie Database," *Screen* 61, no. 1 (2020): 124–128, https://doi.org/10.1093/screen/hjaa007; Charles Tepperman, "The Amateur Movie Database: Archives, Publics, Digital Platforms," *The Moving Image: The Journal of the Association of Moving Image Archivists* 17, no. 2 (2017): 106–110, https://doi.org/10.5749/movingimage.17.2.0106.

11 There was also a separate Video subcategory in addition to the VHS and MiniDV subcategories within the CARRIER category, which did not necessarily overlap. Within this article, we therefore use "Video" with a capital V to indicate items falling under this subcategory in the metadata, whereas we use "video" as a more descriptive term for video-based carriers in general.

12 Official website of the Prelinger Archives, an online database founded by archivist and filmmaker Rick Prelinger: http://www.panix.com/~footage/, accessed August 30, 2023. The collection includes more than 17,000 home movies, recorded on 35mm, 16mm, 8mm, Super 8, and other film formats. Currently, 8,596 films from the Prelinger Archives are accessible via the Internet Archive: https://archive.org/details/prelinger, accessed August 30, 2023.

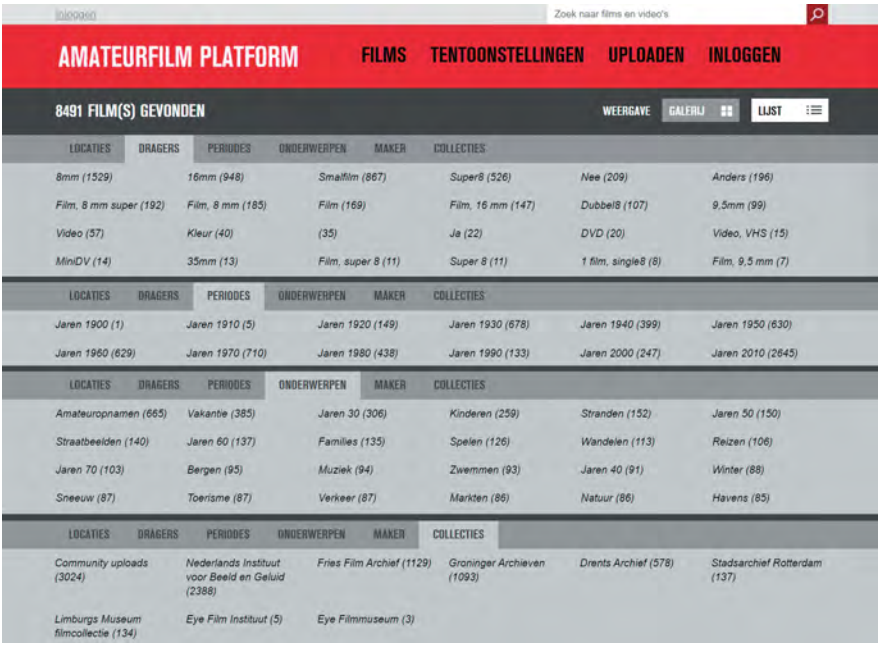


Figure 2: Search menus of the Amateur Film Platform, https://www.amateurfilmplatform.nl/ (accessed April 2, 2023).

que Films Project from the Amateur Movie Database (AMDB),[13] and the Play the City project from the Home Movies Italian Amateur Film Archive in Bologna (Archivio Nazionale del Film di Famiglia), which uses geodatabases to remediate amateur films into on-site installations and exhibitions, allowing visitors to navigate historical maps guided by historical amateur footage.[14] While such projects have

13 The Basque Films Project is a collaboration between Elias Querejeta Zine Eskola (EQZE) and the Amateur Movie Database (AMDB), which develops a detailed map of amateur filmmaking in the Basque Country: https://www.amateurcinema.org/index.php/basquefilms, accessed August 30, 2023.

14 Paolo Simoni, "The Amateur City: Digital Platforms and Tools for Research and Dissemination of Films Representing the Italian Urban Landscape," *The Moving Image: The Journal of the Association of Moving Image Archivists* 17, no. 2 (2017): 111–118, https://doi.org/10.5749/movingimage.17.2.0111. Home movies' ability to function as "time machines" was also explored by artist and digital humanities researcher Ruxandra Lupu in her "Home Movie 4.0" project. See Ruxandra Lupu, "The Home Movie 4.0: (Co)Creative Strategies for a Tacit, Embodied and Affective Reading of the Sicilian Home Movie Archive" (PhD diss., University of Leeds, 2020), accessed August 30, 2023, https://etheses.whiterose.ac.uk/27966/. See also the project website: http://homemoviesicily.com/, accessed August 30, 2023.

been successfully utilizing the affordances and "logics"[15] of databases to abstract and recontextualize amateur film collections and their national or regional histories, they do not make use of digital methods for computationally analyzing the historical films themselves and their audiovisual features or characteristics. This is in line with digital film historical scholarship and approaches in digital humanities research in general, which have predominantly been text-driven and context-oriented rather than focused on the audiovisual materials themselves.[16] Only recently have computer vision methodologies been applied to humanities scholarship, including the analysis of the content and style of audiovisual sources.[17]

While computer vision methodologies have been applied – though in a limited way so far – to archival audiovisual collections,[18] their potential for the study of historical amateur film collections has not yet been fully realized. In general, the use of digital approaches to amateur films and other substandard or non-professional media is a promising yet rather unexplored territory.[19] An explanation for this is that amateur films are generally rather diverse and complex cultural objects that,

15 Lev Manovich was among the first media theorists to write about the logics of databases and the affordances of database-driven narration and storytelling. See Lev Manovich, *The Language of New Media* (Cambridge, MA: MIT Press, 2001).

16 Cf. Charles R. Acland and Eric Hoyt, eds., *The Arclight Guidebook to Media History and the Digital Humanities* (Falmer: REFRAME Books, 2016), accessed March 22, 2024, https://projectarclight.org/book/.

17 Arnold and Tilton, *Distant Viewing*; Lev Manovich, *Cultural Analytics* (Cambridge, MA: MIT Press, 2020); Melvin Wevers and Thomas Smits, "The Visual Digital Turn: Using Neural Networks to Study Historical Images," *Digital Scholarship in the Humanities* 35, no. 1 (2020): 194–207, https://doi.org/10.1093/llc/fqy085. See also the two-part course "Computer Vision for the Humanities": Daniel van Strien et al., "Computer Vision for the Humanities: An Introduction to Deep Learning for Image Classification (Part 1)," *Programming Historian*, August 17, 2022, accessed March 22, 2024, https://programminghistorian.org/en/lessons/computer-vision-deep-learning-pt1; Daniel van Strien et al., "Computer Vision for the Humanities: An Introduction to Deep Learning for Image Classification (Part 2)," *Programming Historian*, August 17, 2022, accessed March 22, 2024, https://programminghistorian.org/en/lessons/computer-vision-deep-learning-pt2. See furthermore the Special Interest Group "AudioVisual Material in Digital Humanities" (AVinDH) from the Alliance of Digital Humanities Organizations (ADHO), accessed March 22, 2024, https://avindhsig.wordpress.com/; Taylor Arnold et al., "Introduction: Special Issue on AudioVisual Data in DH," *Digital Humanities Quarterly* 15, no. 1 (2021), accessed March 20, 2024, https://www.digitalhumanities.org/dhq/vol/15/1/000541/000541.html.

18 See, for instance, the Sensory Moving Image Archive (SEMIA) project from the University of Amsterdam: https://sensorymovingimagearchive.humanities.uva.nl/, accessed March 22, 2024.

19 Susan Aasman, "Unlocking Multiple Histories of Amateur Media: From Micro- to Macro-Histories," *Screen* 61, no. 1 (2020): 157–166, https://doi.org/10.1093/screen/hjaa011. See also Nicole Braida and Frauke Pirk's chapter in this volume, "Teaching Small-Gauge Formats with Digital Methods."

as media historians Susan Aasman and Tom Slootweg rightfully argue, are "notoriously difficult to reduce to neatly categorized units of analysis, because [they do] not necessarily adhere to formal aesthetic or narrative conventions."[20] At the same time, scholars of amateur film and home movies have been trying to unpack some of the complexities of amateur film production through different scholarly methods and disciplines, ranging from visual anthropology and cultural history to film studies and socio-pragmatics.[21] While digital humanities approaches to amateur films have been emerging and seem promising avenues,[22] computer vision methodologies and the Distant Viewing approach have not yet been explored or tested on this type of audiovisual material.

Our project comprises three interrelated objectives. First of all, we aim to provide insights into the Amateur Film Platform as a collection of historical audiovisual materials. We believe the results will be beneficial for audiovisual archives, as the Distant Viewing Toolkit has the potential to function as a digital enrichment tool, by generating metadata about the films' content and style, for

20 Susan Aasman and Tom Slootweg, "'Pure Information, Not the Real Thing': Digital Hermeneutics and Nelson Sullivan's Videographic Legacy (1983–1989)," this volume.

21 Cf. Richard Chalfen, *Snapshot Versions of Life* (Bowling Green, OH: Bowling Green State University Popular Press, 1987); Patricia R. Zimmermann, *Reel Families: A Social History of Amateur Film* (Bloomington: Indiana University Press, 1995); Roger Odin, ed., *Le film de famille: usage privé, usage public* (Paris: Méridiens Klincksieck, 1995); Susan Aasman, *Ritueel van huiselijk geluk. Een cultuurhistorische verkenning van de familiefilm* (Amsterdam: Het Spinhuis, 2004); Alexandra Schneider, *Die Stars sind wir: Heimkino als filmische Praxis* (Marburg: Schüren, 2004); Martina Roepke, *Privat-Vorstellung: Heimkino in Deutschland vor 1945* (Hildesheim: G. Olms, 2006); Ryan Shand, "Amateur Cinema: History, Theory and Genre (1930–80)" (PhD diss., University of Glasgow, 2007); Karen L. Ishizuka and Patricia R. Zimmermann, eds., *Mining the Home Movie: Excavations in Histories and Memories* (Berkeley: University of California Press, 2008); Laura Rascaroli, Gwenda Young, and Barry Monahan, eds., *Amateur Filmmaking: The Home Movie, the Archive, the Web* (New York: Bloomsbury Publishing, 2014); Charles Tepperman, *Amateur Cinema: The Rise of North American Movie Making, 1923–1960* (Oakland, California: University of California Press, 2015); Annamaria Motrescu-Mayes and Susan Aasman, *Amateur Media and Participatory Cultures: Film, Video, and Digital Media* (London: Routledge, 2019); Masha Salazkina and Enrique Fibla-Gutiérrez, eds., *Global Perspectives on Amateur Film Histories and Cultures* (Bloomington: Indiana University Press, 2021). For amateur video histories, see, among others, James M. Moran, *There's No Place like Home Video* (Minneapolis: University of Minnesota Press, 2002); Diego Cavallotti, "L'audiovisivo analogico della quotidianità. Discorsi, pratiche e testi del cinema e del video amatoriale tra gli anni Settanta e gli anni Novanta in Italia" (PhD diss., University of Udine, 2017); Tom Slootweg, "Resistance, Disruption and Belonging: Electronic Video in Three Amateur Modes" (PhD diss., University of Groningen, 2018).

22 See, for instance, the thematic dossier by Charles Tepperman, "The Complex Materiality of Amateur Cinema Research: Texts, Archives and Digital Methods – Introduction," *Screen* 61, no. 1 (2020): 119–123, https://doi.org/10.1093/screen/hjaa006.

example.[23] Secondly, the project aims to explore the potential of the Distant Viewing approach and toolkit for scholarly purposes, namely for viewing digitized collections of amateur films and videos at scale, specifically for the analysis of formal, stylistic, and aesthetic patterns or changes over time. Thirdly, by doing so, the project more generally aims to contribute to the understanding of the history of amateur film as a cultural practice in the Netherlands. It should be noted that the analysis that empirically informed this chapter was of an *exploratory* nature and based only on a small sample of data (as will be further explained below). Instead of providing a comprehensive analysis of all films and videos published on the Amateur Film Platform, it forms a necessary first step toward a larger endeavor, which requires more resources and infrastructural support.[24]

In the following section, we present our case study and describe the various (iterative) steps and processes involved in "distant viewing" the Amateur Film Platform: from the selection and collection to the analysis, visualization, and interpretation of the data. Utilizing the framework of digital hermeneutics, we critically evaluate how the selected data, tools, and algorithms have impacted the research process and results. In the conclusion, we reflect on the main insights the project brought us and present avenues for future research involving digital approaches to historical amateur film collections and audiovisual materials more broadly.

Distant Viewing the Amateur Film Platform

How could the Distant Viewing approach benefit the study of large amateur film collections? How could the Distant Viewing Toolkit be used as a means to extract, aggregate, and visualize certain semantic metadata and features, such as color use, shot length, object detection, and camera movement in relation to the type of amateur medium or technological carrier used for the recording? How could it be used to analyze any historical, aesthetic, and technological changes in the audiovisual materials from the Amateur Film Platform collection? We were specifically interested in exploring possible patterns in the relationship between different amateur media or technological carriers (e.g., film, video, digital media)

23 Cf. Nanne van Noord et al., "Automatic Annotations and Enrichments for Audiovisual Archives," in *Proceedings of the 13th International Conference on Agents and Artificial Intelligence – Volume 1: ARTIDIGH* (Vienna: SciTePress, 2021), 633–640, https://doi.org/10.5220/0010387706330640.

24 Due to limited resources, access to the Amateur Film Platform collection was constrained. Thanks to the help of Netherlands Institute for Sound & Vision conservator Valentine Kuypers, we were able to select a sample of 100 films for a first test and analysis with the Distant Viewing Toolkit.

on the one hand, and any formal, stylistic, and aesthetic changes on the other, including color versus black-and-white, silent versus sound, shot length, shot type (e.g., total shot, medium shot, close-up), camera movement, use of titles, and setting of recording (indoor versus outdoor).

In conducting our research, we took inspiration from previous projects by the Distant Viewing Lab that implemented the Distant Viewing approach and use of computer vision algorithms for the study of digital collections, such as the Photogrammar and ADDI projects.[25] We also took inspiration from various other digital humanities projects and digital film historical scholarship, including Lev Manovich's pioneering work on cultural analytics, which paved the way for the analysis of large cultural datasets through visualization techniques and data science methods. Cultural analytics allows us, Manovich argues, "to study the patterns, trends, and dynamics" of culture at scale.[26] While Manovich focuses on the analysis of *contemporary* culture, in particular social media, the cultural analytics approach offers a framework for thinking about cultural-historical data as well. Cultural historians and digital humanities scholars Melvin Wevers and Thomas Smits make a plea for a "visual digital turn" and use of computer vision and convolutional neural networks to explore both "the content (what is represented) and the style (how is it represented) of images."[27] Other inspiring projects include the project "Me and Myself: Tracing First Person in Documentary History in AV-Collections" (M&M), led by Susan Aasman, which aimed to find tropes of self-representation in historical Dutch first-person documentary films through digital video annotation,[28] and the Sensory Moving Image Archive (SEMIA) project,

25 For the Photogrammar and ADDI projects, see https://photogrammar.org/maps and https://github.com/distant-viewing/addi, accessed March 22, 2024. Other projects which utilized the Distant Viewing approach include Taylor Arnold, Lauren Tilton, and Justin Wigard, "Automatic Identification and Classification of Portraits in a Corpus of Historical Photographs," in *CEUR Workshop Proceedings* (Computational Humanities Research Conference, Antwerp, Belgium, December 2022), 25–35, accessed March 20, 2024, https://ceur-ws.org/Vol-3290/short_paper5571.pdf; Taylor Arnold, Lauren Tilton, and Annie Berke, "Visual Style in Two Network Era Sitcoms," *Journal of Cultural Analytics* 4, no. 2 (2019), https://doi.org/10.22148/16.043. For a comprehensive list, see the Distant Viewing Lab website: https://distantviewing.org/, accessed March 22, 2024.

26 Manovich, *Cultural Analytics*, 14.

27 Wevers and Smits, "The Visual Digital Turn," 195.

28 Susan Aasman et al., "Tales of a Tool Encounter: Exploring Video Annotation for Doing Media History," *VIEW Journal of European Television History and Culture* 7, no. 14 (2018): 73–87, https://doi.org/10.18146/2213-0969.2018.jethc154. For more information about the CLARIAH pilot project, see "Me and Myself: Tracing First Person in Documentary History in AV-Collections" (2017–2018), https://mediasuite.clariah.nl/learn/example-projects/me-and-myself-tracing-first-person-in-documentary-history-in-av-collections-m-and-m, accessed March 22, 2024.

which explored new ways of searching digitized audiovisual collections based on non-semantic descriptors, such as color, shape, movement, and texture.[29] An interface that similarly enabled the exploration and search of digital film collections based on color was recently implemented in the Timeline of Historical Film Colors, developed by film historian Barbara Flückiger and her team at the University of Zurich, Switzerland. In the related FilmColors project, a digital humanities approach was used to analyze historical film colors on the basis of both technological and aesthetic changes.[30]

We too are interested in the relationship between technological and aesthetic changes from a long-term historical perspective. Utilizing the Distant Viewing approach, one of our initial questions was whether computer vision algorithms could be used to detect the type of amateur medium or technological carrier (film, video, digital media) and even the specific format (e.g., 16mm, 9.5mm, 8mm film) used for recording the amateur film or video. More specifically, we wanted to explore the following aspects:

1. Whether films made on different technological carriers can be distinguished from each other based on specific formal, stylistic, or aesthetic qualities.
2. Whether films made in different time periods can be distinguished from each other based on specific formal, stylistic, or aesthetic qualities.
3. Changes in time (e.g., from black-and-white to color, from silent to sound, shot type and length, use of post-production techniques).
4. Continuities in time (e.g., topics, mise-en-scène, genre, shot type and length, use of post-production techniques).
5. Other aesthetic or narrative tropes (e.g., forms of self-representation, the appearance of amateur film and video cameras, medium-specific characteristics).

29 See https://sensorymovingimagearchive.humanities.uva.nl/, accessed March 22, 2024; Eef Masson, Christian Gosvig Olesen, Nanne van Noord, and Giovanna Fossati, "Exploring Digitised Moving Image Collections: The SEMIA Project, Visual Analysis and the Turn to Abstraction," *Digital Humanities Quarterly* 14, no. 4 (2020), accessed March 20, 2024, http://digitalhumanities.org/dhq/vol/14/4/000497/000497.html.

30 Barbara Flückiger, "A Digital Humanities Approach to Film Colors," *The Moving Image* 17, no. 2 (2017): 71–94, https://doi.org/10.5167/uzh-151113. For the Timeline of Historical Film Colors project, see https://filmcolors.org/, accessed March 22, 2024.

Step 1: Data Selection and Collection

Since the collection presented at the Amateur Film Platform was already digitized,[31] our first step was to collect all the digitized films and videos, so that the collection could be aggregated, analyzed, and visualized by means of the Distant Viewing Toolkit. For various reasons, however, it was not possible to get direct access to all the digital files. The files existed in multiple different digital formats (XML, AVI, MP4, etc.) and were stored in different locations and servers. Moreover, it would require permission from all the archival institutions represented and all the persons whose amateur collections are visible on the platform to re-use their collections for research purposes. We therefore worked with a sample of 100 films and videos, limited to items from the Netherlands Institute for Sound & Vision (NISV), which kindly permitted us to work with this selection of films and metadata from their collection. We felt that limiting the sample data to only the NISV archival collection was justified, as the NISV collection presented the majority of the archival films uploaded to the platform.[32] At the same time, like the other archival institutions represented on the platform, the NISV amateur collection is predominantly film-based. At her presentation at the "e-nedits" meetings in 2020, the NISV conservator of amateur film collections, Valentine Kuypers, mentioned a noticeable peak in film-based materials from the 1930s and late 1950s–1970s in the collection of the Amateur Film Platform (Figure 3).[33] As a result, analogue and digital video formats, like VHS, Betamax, Video2000, Hi8, and MiniDV, which used to be popular amateur recording technologies in the 1980s–1990s, were underrepresented in the collection and on the platform in general. We have tried to circumvent this bias and compensate for the temporal gap by including data from other technological carriers and time periods in the sample. This choice was motivated by our research questions and interest in the study of patterns of historical, technological, and aesthetic changes in amateur media productions.

31 The majority of the films from the NISV collection were digitized as part of the Images for the Future (2007–2014) digitization project, which aimed to digitize, preserve, and make accessible a large part of Dutch audiovisual heritage collections: https://www.beeldengeluid.nl/en/knowledge/projects/images-future, accessed March 22, 2024.

32 At the time of analysis, the NISV collection comprised 2,388 of the 8,491 total items, followed by the Frisian Film Archive (1,129 items) and Groningen Audio Visual Archive (1,093 items); the Community Uploads make for 3,024 items on the platform. A large part of the NISV collection on the Amateur Film Platform stemmed from the Small-Gauge Museum ("Smalfilmmuseum") that was initiated by the late conservator Henk Verheul in 1985 and hosted by the NISV since 2006.

33 Valentine Kuypers, "Amateurfilm Platform" (presentation at the "e-nedits" meetings, INEDITS Amateur Films/Memory of Europe, October 13, 2020), accessed March 22, 2024, http://en.inedits-europe.org/News/Meetings/2020-the-e-nedits-year.

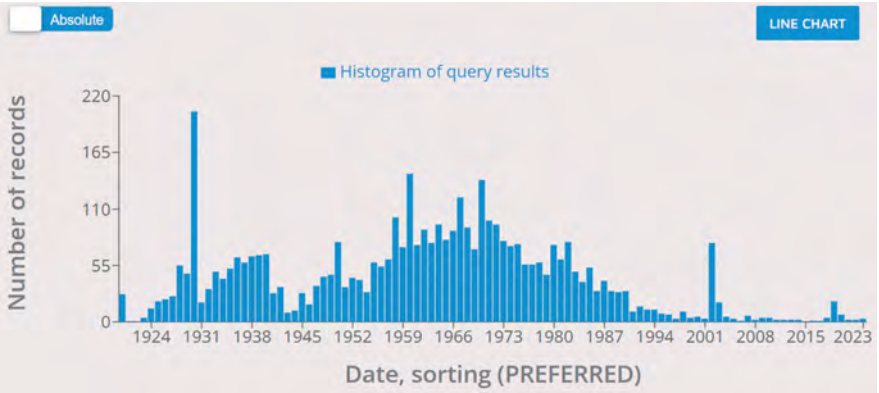


Figure 3: Visualization of the NISV amateur film collection, indicating a bias toward film-based materials from the 1930s and late 1950s–1970s. Source: CLARIAH Media Suite.

In general, we selected our sample of 100 films and videos based on two main criteria and variables: (1) *Time period*: audiovisual items from different decades were selected to enable us to study historical changes; (2) *Medium*: audiovisual items from different amateur media and technological carriers were selected to enable us to study technological changes. For both criteria, we relied on the existing metadata provided by the Amateur Film Platform's interface. As a consequence, items with no specific time period provided in the metadata, and so not appearing in the TIME PERIOD category, were excluded from the selection. The same applied to any item not appearing in the CARRIER category. On the basis of our criteria detailed above, at least two items per decade and per medium were selected. The idea was that this would result in a sample representing all technological carriers (e.g., 35mm, 16mm, 9.5mm, 8mm, Super 8, VHS, MiniDV) and time periods (1900s–2000s). The titles were randomly selected based on their appearance in the Amateur Film Platform search engine (usually alphabetically ordered by film title, sometimes in alphabetical reverse order). In a case where the first two entries were made by the same filmmaker, the following item produced by a different filmmaker was selected to create more diversity in the sample dataset. Where multiple decades were listed in an item's metadata, the item was selected on the basis of the oldest decade: thus a film whose metadata indicated that it included footage of both the 1920s and 1930s was selected as an item representing the 1920s period in our dataset.[34]

34 See, for example, the film item Familiebijeenkomst ter gelegenheid van bezoek uit de vs (Bram Sluis, 1920s-1930s), https://www.amateurfilmplatform.nl/films/sluis-familiefilms-familiebijeenkomst-ter-gelegenheid-van-bezoek-uit-de-vs, accessed October 31, 2023. Archived via: https://web.

The sample dataset was composed in the form of a spreadsheet with various metadata fields to facilitate the aggregation process. The spreadsheet included the following columns: medium (A), time period (B), title of the item (C), file name (D), link to database NISV (E), internal catalog number (F), and URL to Amateur Film Platform (G). The links to the NISV database and internal catalog number were used to facilitate communication with the archive; the URL to the Amateur Film Platform was used for data verification in the analysis phase and for mining relevant metadata, such as content description. After the 100 items had been collected, they were processed by converting the films and videos into a series of still images or frames, so that they could be analyzed by the Distant Viewing Toolkit. In addition to the medium and time period as "source" variables, we selected the following "target" variables: shot length, shot type, setting of recording, color, titles, and camera movement. For measuring these variables and analyzing their relations vis-à-vis the item's medium and time period, three different algorithms were used by the Distant Viewing Toolkit:

1. *Shot boundary detection*:[35] this algorithm helps to detect the number of cuts of an item based on changes in the frame and so helps define the average or median shot length per item (Figure 4).
2. *Image segmentation algorithm*:[36] this algorithm detects objects within the frame, but also backgrounds. Background detection helps to analyze whether a film or scene was recorded indoors or outdoors.
3. *Face detection algorithm*:[37] this algorithm detects people's faces appearing in the frame and, based on this, can be used to define the type of shot (e.g., close-up, medium, total shot).

archive.org/web/20230608061847/https://www.amateurfilmplatform.nl/films/sluis-familiefilms-familiebijeenkomst-ter-gelegenheid-van-bezoek-uit-de-vs.

35 Jakub Lokoč et al., "A Framework for Effective Known-Item Search in Video," in *Proceedings of the 27th ACM International Conference on Multimedia* (New York: Association for Computing Machinery, 2019), 1777–1785, https://doi.org/10.1145/3343031.3351046.

36 Holger Caesar, Jasper Uijlings, and Vittorio Ferrari, "COCO-Stuff: Thing and Stuff Classes in Context," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE Computer Society, 2018), 1209–1218, https://doi.org/10.1109/CVPR.2018.00132.

37 Qiong Cao et al., "VGGFace2: A Dataset for Recognising Faces across Pose and Age," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)* (IEEE Computer Society, 2018), 67–74, https://doi.org/10.1109/FG.2018.00020.

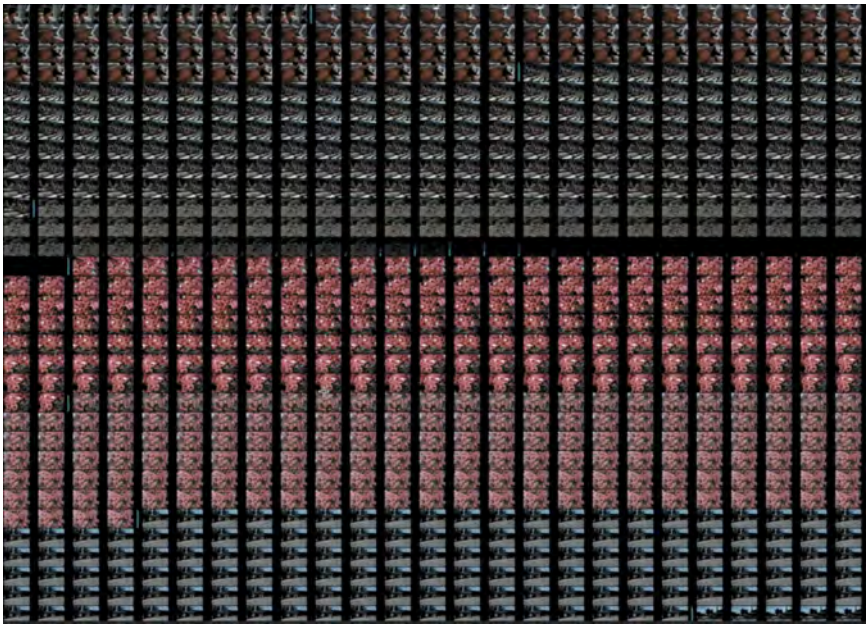


Figure 4: Visualization of the shot boundary detection algorithm being used by the Distant Viewing Toolkit to detect cuts based on frame changes in one of the items from the Amateur Film Platform collection.

Step 2: Data Analysis and Visualization

In various rounds, we analyzed the data sample and tested some of our hypotheses about possible relations between the variables that we had formulated in advance. Specifically, we researched the relationships between medium and time period; shot length in relation to time period and medium; shot type and time period; and setting of recording in relation to time period and medium.[38]

38 In future studies, we would also like to explore the relationships between color and time period; color and medium; film titles and time period; film titles and medium; camera movement and time period; camera movement and medium; topics and time period; and topics and medium.

Medium and Time Period

The first relationship we analyzed was the relation between the technological carrier or medium used and the time period. Based on changes in the emergence and use of certain amateur media technologies in the twentieth century, we hypothesized that film-based items would appear from the 1900s–1980s, electronic video-based items would appear from the 1980s–1990s, and born-digital items from the 1990s–2000s.

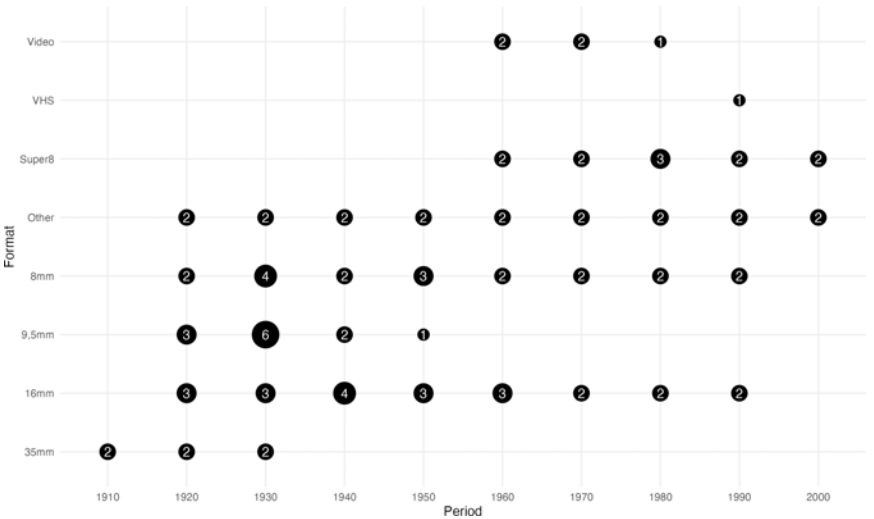


Figure 5: Visualization of the relationship between the medium and time period.

Figure 5 shows how the selected data deviate slightly from the previously mentioned idea of having all technological carriers and time periods equally represented in the sample. Most dots in the graph show the number "2," meaning two audiovisual items were selected for a particular combination of time period and medium, as intended. A variety of reasons can be given in cases where numbers are missing or deviate from this rule. Firstly, a medium has become obsolete or is no longer used by amateurs; for instance, all 35mm films that are included in the Amateur Film Platform collection were made between the 1910s and 1930s. The same was the case for 9.5mm film, which was no longer used as a medium after the 1950s in this sample. Another reason is that a medium was introduced later in time. For instance, Super 8 film was introduced as a technological carrier in 1965, which means no Super 8 films could have been made before the 1960s. The same goes for video-based carriers, like VHS and MiniDV, which were popularized in

the 1980s and 1990s, respectively. Thirdly, the collection itself is limited as it sometimes has only one item from a specific medium available for the selected time period. For instance, there is only one 9.5mm film item from the 1950s, one Video item[39] from the 1980s, and one VHS item from the 1990s. In some cases the dot outnumbers two; for instance, in the 1930s with 9.5mm film (6 items), 8mm film (4 items), and 16mm film (3 items) because, in addition to the selection criteria (time period and medium), a few extra items were selected that addressed specific themes (from exhibitions) or filmmakers that we wanted to include as well (leading to the uneven distribution of technological carriers over time within the sample).

When we look at the graph, we can see that film-based materials are overrepresented in the dataset, also because there are simply more film-based carriers than video-based carriers in the collection. So this was to be expected. More surprising is that Super 8 film items are dominant over video formats in the dataset during the 1980s and 1990s, despite the well-known popularity and widespread use of electronic video in these decades. An explanation for this is the above-mentioned gap in the collection of the Amateur Film Platform, both for electronic video as an amateur medium and for items from the 1980s and 1990s. There are even more films than videos from the 1990s and 2000s, which once again exemplifies the significant bias towards film as a medium and technological carrier in the NISV amateur collection and the collection of the Amateur Film Platform in general.

Shot Length

We used the shot boundary detection algorithm to view and analyze the relations between the shot length and the time period, and between the shot length and the medium.

Shot Length – Time Period

We hypothesized that film-based productions of the first half of the twentieth century would contain more cuts hence a shorter median shot length.[40] This hypoth-

39 See footnote 11.

40 We have chosen the *median* shot length instead of the *average* shot length, following previous research on shot lengths. See Nick Redfern, "The Log-Normal Distribution is Not an Appropriate Parametric Model for Shot Length Distributions of Hollywood Films," *Digital Scholarship in the Humanities* 30, no. 1 (2015): 137–151, https://doi.org/10.1093/llc/fqs066.

esis was informed by the fact that film was expensive, especially in the early twentieth century, so amateur filmmakers would be inclined to use film sparingly and therefore limit the duration of the recording. Apart from costs, the technology of the amateur film camera was a constraining factor: the first spring-based amateur film cameras of the 1920s and 1930s would allow for a maximum constant shooting of approximately 35–40 seconds per shot, whereas electromotor-driven cameras introduced in the 1960s could make longer duration shots.[41]

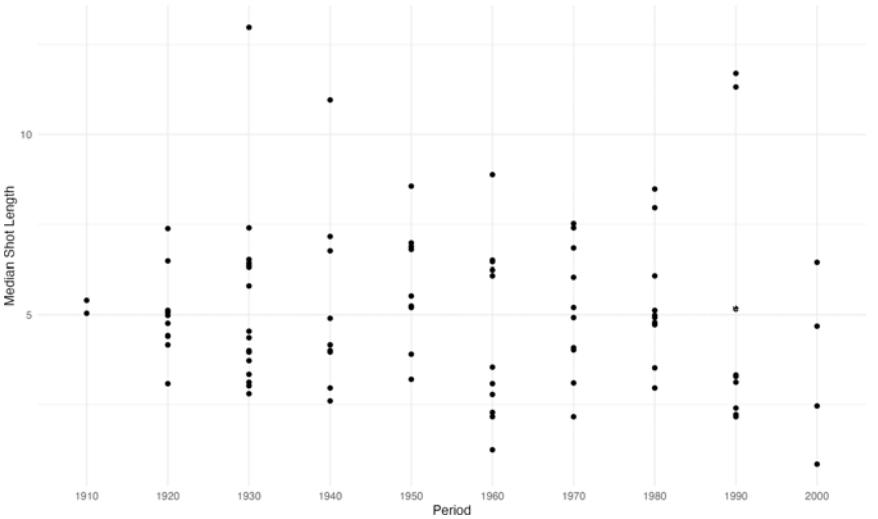


Figure 6: Visualization of the relationship between the time period (in decades) and the median shot length (in seconds).

Figure 6 shows no clear trend or changes in the relationship between the time period and median shot length. Most films and videos from the sample have a median shot length between 2.5 and 7.5 seconds. The figure shows a few outliers. The film with the longest median shot length – 12.98 seconds – was the film AUTOTOCHT NAAR ZWITSERLAND/ITALIË/OOSTENRIJK 1937, made by the amateur filmmaker

41 A spring-driven Ciné-Kodak from the 1930s, for instance, allowed for running 14 to 16 feet of film. See Eastman Kodak Company, *Instructions for Use of the Ciné-Kodak, Model K* (Rochester, New York, ca. 1930), 19. Electromotor-driven cameras no longer had to be manually wound, so could potentially record one take per film roll. Interestingly, manually driven cameras, such as the first Ciné-Kodak camera (model A) from 1923 and Pathé Baby 9.5mm film cameras from 1923–1927, similarly allowed for continuous recording, as their operation was not limited by the spring-drive mechanism.

J. M. Le Grand in 1937.[42] A close analysis of this item reveals that the long median shot length was partly the result of the use of title cards and various jump cuts not detected by the algorithm. Another outlier is the amateur video HUWELIJKSBUS: JAN WIL JE MET ME TROUWEN (Emilia van der Meer, 1993), recorded on VHS, which includes many long takes and so provides a median shot length of 11.32 seconds.[43] This example shows the possible impact of the technological affordance of video cameras to make significantly longer recordings, up to several hours rather than minutes, compared to regular amateur film cameras.[44]

Shot Length – Medium

When looking at the relationship between the shot length and the medium, however, it becomes clear how the technological affordances of the medium do not necessarily determine actual user practices. We hypothesized that because of video's ability to make multiple hours of recordings per tape, there would be fewer cuts in amateur video productions compared to film-based amateur productions. However, if we look at Figure 7, there is no clear trend that indicates this. In fact, there is one noticeable outlier from the sample – the video PROGRESSION (2000) from the amateur film- and video-maker Cor Lievendag[45] – which presents a rather low median shot length (0.49 mu).

Shot Type

We used the face detection algorithm to define the medium height of the face within the frame, which helps to define the shot type: close-up, medium, or total shot. In a close-up shot type, a high percentage of the frame will be filled with the detected face, while in a total shot type, a low percentage of the frame will be filled. Based on these definitions, we were able to view and analyze the relations between the shot type and the time period.

42 See https://www.amateurfilmplatform.nl/films/autotocht-naar-zwitserland-italie-oostenrijk-1937-1, accessed October 31, 2023. Archived via: https://web.archive.org/web/20230402080614/https://www.amateurfilmplatform.nl/films/autotocht-naar-zwitserland-italie-oostenrijk-1937-1.

43 See https://www.amateurfilmplatform.nl/films/huwelijksbus-jan-wil-je-met-me-trouwen-emilia, accessed October 31, 2023. Archived via: https://web.archive.org/web/20230322041545/https://www.amateurfilmplatform.nl/films/huwelijksbus-jan-wil-je-met-me-trouwen-emilia.

44 Moran, *There's No Place like Home Video*, 41; Slootweg, "Resistance, Disruption and Belonging," 213.

45 See https://www.amateurfilmplatform.nl/films/progression, accessed October 31, 2023.

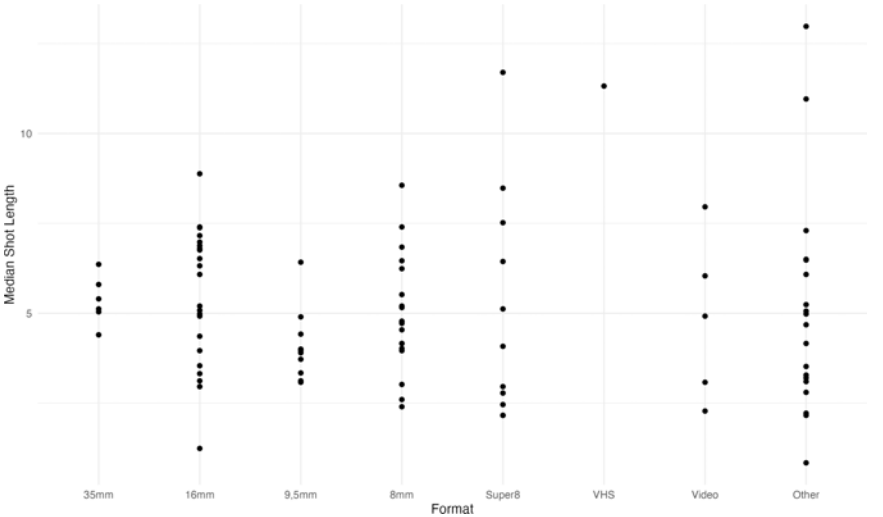


Figure 7: Visualization of the relationship between the medium (per format) and the median shot length (in seconds).

Shot Type – Time Period

Our hypothesis for the relation between the shot type and time period was that close-ups would be increasingly used over time. One of the possible reasons for this aesthetic shift is changes in the affordances of amateur recording technologies. In the 1920s and 1930s, for instance, it was technically more difficult to make close-ups without the risk of recording unsharp images. With early film cameras, one had to manually measure the distance between the lens and the subject in order to set the focus. The later emergence of autofocus functionalities, the possibility of electronic playback for video, and the live preview mode of digital cameras largely eliminated these risks. From the 1960s onwards, we also expected close-ups to become more prominent due to the introduction of the zoom lens, first on film cameras and later also on video recording technologies.

Figure 8 shows a certain trend towards a higher median size of faces based on proportions of the frame height, which could indicate an increase of close-ups as shot type. Items from the 1920s show a small percentage of close-ups with values between around 5 and 12.5, whereas items from the 2000s indicate a larger proportion of the frame being filled with the subject, with values between around 10 and 35. At the same time, however, items with values around 5 continue to be found between the 1920s and 1990s, so films and videos with many medium shots or total shots continue to be made in more recent amateur media productions.

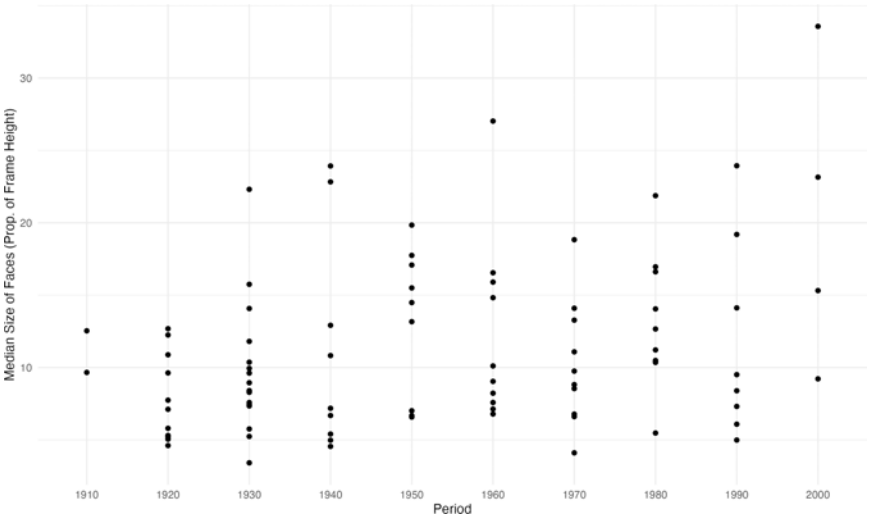


Figure 8: Visualization of the relationship between the time period (in decades) and the type of shot (in median size of faces, properties of the frame height).

Setting of Recording

We used the image segmentation algorithm to define the setting of recording (indoor versus outdoor), which helped us to view and analyze the relations between the setting and the time period, and between the setting and the medium.

Setting – Time Period

The relation between the setting and the time period could signal an aesthetic change from amateur recordings in the first half of the twentieth century to those in the second half. We assumed there would be more outdoor recordings in amateur films from the first half of the twentieth century because of the relatively low light sensitivity of early small-gauge film material. In the 1920s and 1930s, for instance, most reversal films had a speed around 10 ASA, whereas this changed to films with higher speeds ranging from 25–160 ASA and more in the post-war years.[46]

46 Alan Kattelle, *Home Movies: A History of the American Industry, 1897–1979* (Nashua, NH: Transition Pub., 2000), 215, 333–334. See also: https://www.britannica.com/technology/motion-picture-technology/Film#ref508443, accessed August 19, 2024.

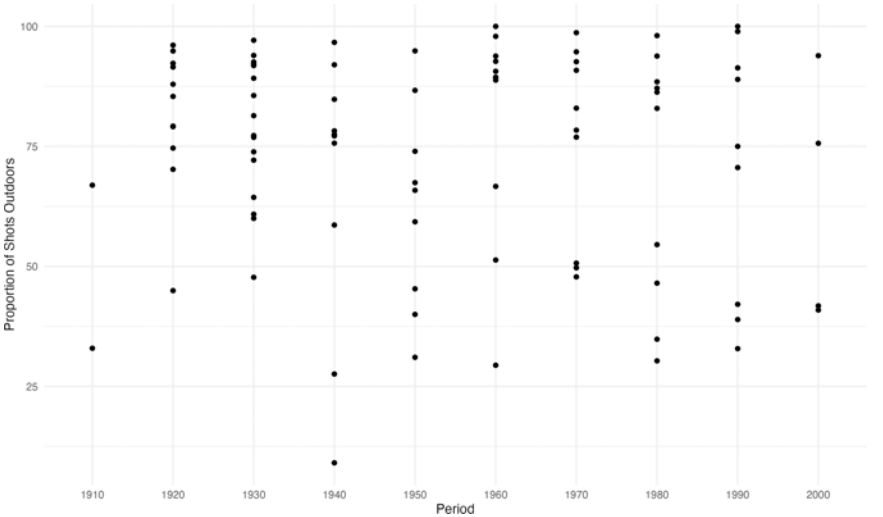


Figure 9: Visualization of the relationship between the time period (in decades) and the setting of recording (based on the proportion of outdoor shots detected by the algorithm).

Figure 9 shows no clear trend, however. The majority of films from our sample seem to have been recorded outdoors, with a few exceptions in each time period, most notably one item from the 1940s, KLEUTERKLAS, HUISHOUDSCHOOL (H. de Jong-Kleinweg de Zwaan, 1940-1941), where the subject consisted entirely of the preparation and serving of a formal meal in an indoor dining room.[47] In general, more variations in recording settings seem to have occurred from the 1940s onwards, possibly due to the above-mentioned changes in the technological affordances of the recording medium.

Setting – Medium

In relation to the medium used as a variable, we assumed there would be more indoor recordings in video-based amateur productions than in film-based amateur productions.[48] The first electronic video cameras that appeared in the 1960s and 1970s

47 See https://www.amateurfilmplatform.nl/films/kleuterklas-huishoudschool-1940-1941-2, accessed October 31, 2023. Archived via: https://web.archive.org/web/20231003132327/https://www.amateurfilmplatform.nl/films/kleuterklas-huishoudschool-1940-1941-2.
48 Moran, *There's No Place like Home Video*, 41.

were directly connected to a stationary video recorder.[49] Only later, portable video recording technologies were introduced that allowed for more flexible forms of recording, both indoors and outdoors.

Figure 10 does not indicate a significant difference between the medium and the setting of recording or between film- and video-based amateur productions specifically. In general, however, it seems that the majority of 8mm, 9.5mm and Video items were recorded outdoors (50% or higher proportions of shots outside), whereas for amateur films recorded on 35mm, 16mm, and Super 8 more variations occurred, hence these technological carriers included items with relatively more indoor shots (below 50%).

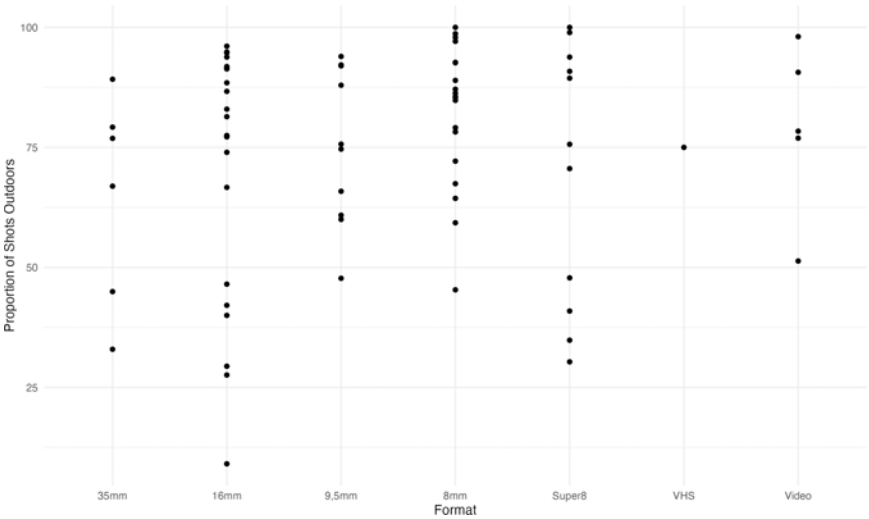


Figure 10: Visualization of the relationship between the medium (per format) and the setting of recording (based on the proportion of shots outdoors).

Step 3: Data Interpretation and Criticism

How to interpret the analyzed relationships from the previous step? What new insights do they provide? How do they reflect patterns or trends in amateur filmmaking in the Netherlands? And how can we critically reflect on these results and the ways in which they were produced by the selected algorithms? In this

49 Van der Heijden, "Hybrid Histories," 178–179.

final step, we draw on the framework of *digital hermeneutics*, defined as a "concept that enables historians to critically reflect on the various interventions of digital research infrastructures, tools, databases, and dissemination platforms in the process of thinking, doing and narrating history."[50] The digital hermeneutics framework helps us to interpret the data and reflect on how the Distant Viewing approach and toolkit have shaped our research practices and findings.

Data Bias, Patterns, and Source Criticism

For the interpretation of the results, it is important to reflect first of all on the notion of *data representation*: how representative are the results and what do they tell us? Since the collection of the Amateur Film Platform indicated a strong bias towards film-based materials, we tried to balance this out by selecting items from multiple time periods and technological carriers. Constructing a sample dataset based on these criteria helped us to study certain patterns and changes in the collection, as we have seen in the previous step. Yet, it also led to a new form of bias, which prevented the sample from being representative, either for the collection of the Amateur Film Platform or for Dutch amateur media productions in general. Based on the small sample of data we analyzed, it is therefore difficult to make valid statements about relationships, trends, and patterns in amateur media collections and the possible historical, technological, or aesthetic changes they reflect. The value of the present study lies, we would argue, on the methodological level and the various questions it raises about using digital methods for the study of historical amateur media collections.

Such questions pertain to the digital transformation of archival materials. While digital methods may create new biases, they can also help to make existing biases and their underlying "politics of digitization" in archival databases more explicit.[51] In relation to the Amateur Film Platform collection, the bias towards film-based collections largely stems from collection policies, which generally prioritized film over video as cultural heritage objects in digitization projects.[52] This also

50 Andreas Fickers, Juliane Tatarinov, and Tim van der Heijden, "Digital History and Hermeneutics – Between Theory and Practice: An Introduction," in *Digital History and Hermeneutics*, ed. Andreas Fickers and Juliane Tatarinov (Berlin: De Gruyter, 2022), 6–7, https://doi.org/10.1515/9783110723991-001.

51 Gerben Zaagsma, "Digital History and the Politics of Digitization," *Digital Scholarship in the Humanities* 38, no. 2 (2023): 830–851, https://doi.org/10.1093/llc/fqac050.

52 Frank Holthuizen, Joseph Wachelder, and Tim van der Heijden, "Amateurfilm in Limburg, Limburg in Amateurfilm," in *Publications de la Société Historique et Archéologique dans le Limburg*, vol. 156 (Maastricht: Koninklijk Limburgs Geschied- en Oudheidkundig Genootschap LGOG, 2021), 229–281.

touches upon the importance of *digital source criticism* for reflecting on provenance and digitization processes: when and how have the source materials been digitized, and how has this influenced the results? The majority of the digitized films that we used in our sample were digitized in standard definition (SD) resolution, while more recently digitized items in the collection have been scanned in high definition (HD) or even 2K image resolution.[53] The question of how differences in digitization practices and scan resolutions affect the application of the Distant Viewing Toolkit algorithms would be an interesting avenue for further exploration and comparative testing, but was beyond the scope of the present study.

In addition to questions of archival politics and practices of digitization, another point of reflection is the use of *metadata* and the way they have been structured, and hence shaped our research. As Johanna Drucker reminds us, metadata as a "system of naming, organizing, and classifying" of materials is not neutral but influences how archival items are searched and researched to a great extent.[54] In the case of the Amateur Film Platform, in fact, we have been heavily relying on the platform's metadata for selecting the items and constructing our data sample. Instead of making categorizations ourselves based on certain search criteria, we took the existing categories provided by the platform to select items from the different time periods and technological carriers. Although this enabled us to quickly select various potentially relevant items, it also entailed two large disadvantages as discussed: mislabeling and the exclusion of relevant items.

Data Conversion, Algorithms, and Tool Criticism

In addition to archival politics, digitization practices, and metadata models, our results were shaped by the affordances and constraints of the Distant Viewing approach. Applying questions of *algorithmic criticism* and *tool criticism*,[55] we can reflect on how the toolkit and its algorithms have shaped or structured our analysis and interpretation of the data. One way the toolkit restructures the data is by

53 Cf. Franziska Heller, *Update!* (Paderborn: Wilhelm Fink Verlag, 2020), https://doi.org/10.30965/9783846764602. See also Franziska Heller and Ulrich Ruedel, "Pursuing Film History with Digital Images: Towards Visual Literacy in the Age of AI and Social Media," this volume.

54 Johanna Drucker, *The Digital Humanities Coursebook: An Introduction to Digital Methods for Research and Scholarship*, 1st ed. (Abingdon, Oxon; New York: Routledge, 2021), 53.

55 Stephen Ramsay, *Reading Machines: Toward an Algorithmic Criticism* (Urbana, IL: University of Illinois Press, 2017), https://doi.org/10.5406/illinois/9780252036415.001.0001; Marijn Koolen, Jasmijn Van Gorp, and Jacco van Ossenbruggen, "Toward a Model for Digital Tool Criticism: Reflection as Integrative Practice," *Digital Scholarship in the Humanities* 34, no. 2 (2019): 368–385, https://doi.org/10.1093/llc/fqy048.

converting the audiovisual items into a series of still images or frames, so that they can be processed by the computer vision algorithms. While this step is necessary to enable us to computationally analyze audiovisual materials at scale, it also implies the transformation of the data and removal of some of the original qualities of the film and video items, including their temporal dimensions and, where applicable, their sound layer.[56] Although there are no suitable algorithms at the moment of writing that could be used by the Distant Viewing Toolkit to meaningfully analyze the sound layer in addition to the visual layer, this multimodality would be a crucial aspect to take into account in future explorations.

Other current limitations we encountered when applying the Distant Viewing Toolkit's algorithms pertain to certain aesthetic characteristics or medium-specific features. Sometimes, for instance, the toolkit's algorithms had difficulties with the interpretation of film titles. The image segmentation algorithm mistakenly detected some of those with the label "sky." Furthermore, in cases where films made use of fades (i.e., fade-in or fade-out), the shot boundary detection algorithm sometimes encountered difficulties when counting the number of cuts. Such issues can be explained by the fact that the computer vision algorithms that were used have not been trained (yet) on historical audiovisual materials, let alone amateur media productions, so we anticipated they would not work "perfectly" on digitized analogue films and videos. We were actually surprised how well the algorithms worked on even the oldest black-and-white films from the collection from the 1910s and 1920s. In fact, they worked even better on the older material compared to the electronic video items from the 1980s–1990s due to the visual noise and grainy quality of the latter. Correcting these mistakes from the computer vision algorithms prevented our work with the Distant Viewing Toolkit being a straightforward or linear process; it was rather a cyclical or iterative process, in which algorithms usually had to be "adjusted" to achieve more reliable results.

Conclusion

In this chapter, we used the Distant Viewing approach and toolkit to navigate and analyze the Amateur Film Platform collection at scale. In our study, which serves as a pilot project for a larger and more comprehensive study on the application of computer vision algorithms to amateur films and videos as audiovisual materials from archival collections, we wanted to explore possible occurrences of formal, stylistic, or aesthetic patterns in the collection based on changes in periods

56 See also Aasman and Slootweg, "Pure Information, Not the Real Thing," this volume.

of time and technological carriers. In response to our five sub-questions formulated in the beginning of this chapter, we can make the following preliminary conclusions based on the results from our sample analysis:

1. Films made on different technological carriers can, to some extent, be distinguished from each other based on specific formal, stylistic, or aesthetic qualities. There is no correlation between the amateur medium or technological carrier and the shot length. However, there seems to be some kind of relationship between a selection of technological carriers and the setting of recording: 8mm, 9.5mm and Video items were generally recorded outdoors while amateur films recorded on 35mm, 16mm, and Super 8 included indoor recordings as well.
2. Films made in different time periods can, to some extent, be distinguished from each other based on specific formal, stylistic, or aesthetic qualities. While we could not find a clear trend in the relationship between the shot length and time period, there seems to be a trend toward an increased use of medium and close-up shot types over time (higher median size of faces based on proportions of the frame height). No trend could be seen for the relation between time period and the setting of recording, although more variations in indoor and outdoor recordings seem to occur from the 1940s onwards, possibly due to the emergence of more light-sensitive amateur film material and recording equipment that enabled indoor recording practices without the use of external lamps.
3. The most noticeable changes over time were shown in the sample when it comes to changes in shot type, namely an increase of close-ups. Changes in amateur media productions from black-and-white to color, as well as from silent to sound, we also expect to be significant. Possible changes in post-production techniques, like use of title animations and montage, are probably more difficult to detect computationally.
4. The most significant continuities were found in the recording setting: most amateur films and videos from the sample indicate a high proportion of outdoor shots. In addition, the relationship between shot type and time period shows the continuation of films and videos with many medium shots and total shots.
5. No other aesthetic or narrative tropes were found in the present study. However, forms of self-representation and the appearance of amateur recording technologies like film and video cameras within the frame, to be detected via object detection algorithms, could be interesting tropes to explore in follow-up studies.

Distant and Close Viewing: Towards a Hybrid Heuristics

Clearly, Distant Viewing has considerable potential for the study of large datasets of historical amateur media collections. As demonstrated in previous projects of the Distant Viewing Lab, the data-driven approach to audiovisual data does not replace traditional film analytical methods, but rather stimulates or provides the starting point for further research and investigation. The Distant Viewing Toolkit provides a way to zoom in and out on a corpus of audiovisual data and to see possible patterns, deviations, and relations between items, in our case the sample of films and videos from the Amateur Film Platform collection. In particular, the visualization of outliers in the data can be useful as a heuristic instrument. Outliers in the data help to quickly detect those audiovisual items that deviate from a certain trend or norm. In our sample, for example, the high-speed montage video PROGRESSION (2000) stood out because of its low median shot length. Outliers can also be used effectively to find errors or biases in the algorithms and underlying code systems, when a film was mislabeled in a certain category, for instance, or when the algorithm provided an erroneous qualification. The image segmentation algorithm that we used for detecting the setting of recording (indoor versus outdoor), for example, mistakenly analyzed some of the overexposed images with the tag "snow" and hence mislabeled the scene as "outdoors." Such discrepancies illustrate once more that the results from using digital methods should never be taken for granted, but carefully evaluated and interpreted.

The question of how to "view" and "interpret" the data is a challenging one in general. This is especially the case when it comes to amateur-produced materials, which tend to become meaningful not so much through their aesthetic or narrative features but rather in relation to their contexts of production and reception.[57] For the FilmColors research project, Barbara Flückiger argued that one of the pitfalls of quantitative analysis "is its potential to disregard the meaningful context of data occurrences across the body of works studied."[58] While a data-driven approach may indeed enable new possibilities and perspectives on large collections,[59] the use of digital methods should not prevent a complementary, contextualist approach and close analysis for comprehensively understanding historical materials and their specific-

57 Odin, *Le film de famille*, 36; Aasman, *Ritueel van huiselijk geluk*, 73.

58 Flückiger, "A Digital Humanities Approach to Film Colors," 72.

59 In the FilmColors project these new possibilities included the identification of "diachronic aesthetic patterns" of film color aesthetics. See Flückiger, "A Digital Humanities Approach to Film Colors," 72.

ities. The combination of digital methods with close analysis is crucial.[60] In conclusion, we would therefore like to argue for a *hybrid heuristics*, which combines "distant" and "close" viewing, computational and hermeneutic modes, data-driven and case-study-driven forms of analysis, as well as an approach that combines the affordances of digital technologies and methods, such as computer vision, with the critical reflection on their hermeneutic implications, such as biases and data structures.

Next Steps

Potential next steps in applying the Distant Viewing approach for the study of amateur media histories includes the development of more complex analyses, involving more than two variables and algorithms. This would be useful for studying certain aspects of amateur filmmaking as a cultural practice, for instance whether a film was recorded on a tripod or by means of a hand-held camera. Such information could possibly be retrieved computationally by combining the image segmentation and face detection algorithms. Similarly, more complex queries are required for analyzing practices and forms of self-representation (e.g., detecting changes in the ways amateur filmmakers have portrayed themselves in their films and how this has been shaped by the medium used),[61] the appearance of technological objects (e.g., film/video camera), and medium specificity (e.g., recognition of technological carrier on the basis of certain aesthetic and material characteristics, such as film grain, perforation type, manufacturing codes, and other relevant information about the audiovisual material). The detection of the average number of "out of focus" shots per film or video might also be interesting to explore historically.

Another potentially interesting avenue for further exploration is to investigate how the combination of multiple variables and algorithms could shed new light on differences in types of amateur users, such as "family filmmakers" who make films to capture family memories versus "serious hobbyists" who make films as a (technical) hobby.[62] Could we confirm, for example, any patterns indicating that serious hobbyists use more film titles, have a greater variation in shot length and shot type in their films, and prefer 16mm over 8mm film as recording medium? The question of whether Distant Viewing could be used to detect the type of technological carrier

60 See also Casper Tybjerg, Jonatan Bruun Borring, and Luan Nhu Vu, "The Digitization of Silent Films and the Teaching of Film Historiography: Entanglements and Opportunities," this volume.
61 Aasman et al., "Tales of a Tool Encounter." See also Aasman and Slootweg, "Pure Information, Not the Real Thing," this volume.
62 Aasman, *Ritueel van huiselijk geluk*, 39–64; Van der Heijden, "Hybrid Histories," 89–90.

or medium used, such as whether the amateur film was recorded on 16mm, 9.5mm, 8mm, or Super 8, was one of the questions that inspired our project in the first place.[63] At this stage, however, it seems unlikely it would be possible to detect this computationally when no information on or about the perforation is included in the scan, even when doing extensive development and additional training of the algorithms. This is because there are simply too many factors that may influence the scan results and hence the quality of the digitized film, from varieties in the type of film material and circumstances of film production to how and when the film was digitized and by whom.

Nevertheless, we are convinced that there is great potential in the Distant Viewing approach and the use of computer vision methods for the study of historical amateur film collections at scale. Not only for scholarly purposes, but also for audiovisual archives, it may provide new ways to access, navigate, extract, and enrich metadata from digitized collections.[64] For understanding historical amateur film collections and the history of amateur filmmaking as a cultural practice, our project provided a first step towards a larger project with more infrastructural support. Rather than working with a data sample, we would like to fully analyze and internationally compare multiple large collections of amateur films and videos. Such a comprehensive and comparative approach would truly make use of the potential of Distant Viewing for analyzing amateur media collections at scale, thereby making visible various (hidden) patterns, relations, changes, and continuities in the history of amateur filmmaking as a cultural practice from a long-term historical perspective.

63 This question was raised after visiting Jasper Rigole's exhibition "Homeless Movies" at the House of Alijn, Museum of Everyday Life in Ghent, Belgium, in 2016–2017. Rigole is a Belgian visual artist, researcher, and collector of home movies and found footage materials. The exhibition presented Rigole's IICADOM collection, on the basis of which he constructed "an elementary taxonomy of collected memory" that allowed visitors to search and filter the collection of home movies by theme, subject, genre, action, location, object, and formal characteristics. See Jasper Rigole, *Addenda* (Ghent: AraMER, 2015), accessed August 20, 2024, https://www.jubilee-art.org/projects/addenda. For the IICADOM website, see https://iicadom.org/, accessed March 22, 2024.

64 It would also be worthwhile to explore the question of how Distant Viewing could be used for enriching metadata of archival film collections further in relation to film identification processes. Cf. Harold Brown and Camille Blot-Wellens, *Physical Characteristics of Early Films as Aids to Identification*, 2nd ed. (Brussels: FIAF, 2020).

References

Aasman, Susan. “Unlocking Multiple Histories of Amateur Media: From Micro- to Macro-Histories.” *Screen* 61, no. 1 (2020): 157–166. https://doi.org/10.1093/screen/hjaa011.

Aasman, Susan. “Finding Traces in YouTube’s Living Archive: Exploring Informal Archival Practices.” *TMG Journal for Media History* 22, no. 1 (2019): 35–55. https://doi.org/10.18146/tmg.435.

Aasman, Susan. *Ritueel van huiselijk geluk. Een cultuurhistorische verkenning van de familiefilm*. Amsterdam: Het Spinhuis, 2004.

Aasman, Susan, Liliana Melgar Estrada, Tom Slootweg, and Rob Wegter. “Tales of a Tool Encounter: Exploring Video Annotation for Doing Media History.” *VIEW Journal of European Television History and Culture* 7, no. 14 (2018): 73–87. https://doi.org/10.18146/2213-0969.2018.jethc154.

Acland, Charles R., and Eric Hoyt, eds. *The Arclight Guidebook to Media History and the Digital Humanities*. Falmer: REFRAME Books, 2016. Accessed March 22, 2024. https://projectarclight.org/book/.

Arnold, Taylor, Jasmijn Van Gorp, Stefania Scagliola, and Lauren Tilton. “Introduction: Special Issue on AudioVisual Data in DH.” *Digital Humanities Quarterly* 15, no. 1 (2021). Accessed March 20, 2024. https://www.digitalhumanities.org/dhq/vol/15/1/000541/000541.html.

Arnold, Taylor, and Lauren Tilton. *Distant Viewing: Computational Exploration of Digital Images*. Cambridge, MA: MIT Press, 2023.

Arnold, Taylor, and Lauren Tilton. “Distant Viewing Toolkit: A Python Package for the Analysis of Visual Culture.” *Journal of Open Source Software* 5, no. 45 (2020): 1–6. https://doi.org/10.21105/joss.01800.

Arnold, Taylor, and Lauren Tilton. “Distant Viewing: Analyzing Large Visual Corpora.” *Digital Scholarship in the Humanities* 34, issue supplement 1 (2019): i3–i16. https://doi.org/10.1093/llc/fqz013.

Arnold, Taylor, Lauren Tilton, and Annie Berke. “Visual Style in Two Network Era Sitcoms.” *Journal of Cultural Analytics* 4, no. 2 (2019). https://doi.org/10.22148/16.043.

Arnold, Taylor, Lauren Tilton, and Justin Wigard. “Automatic Identification and Classification of Portraits in a Corpus of Historical Photographs.” In *CEUR Workshop Proceedings*, 25–35. Computational Humanities Research Conference, Antwerp, Belgium, December 2022. Accessed March 20, 2024. https://ceur-ws.org/Vol-3290/short_paper5571.pdf.

Brown, Harold, and Camille Blot-Wellens. *Physical Characteristics of Early Films as Aids to Identification*, 2nd ed. Brussels: FIAF, 2020.

Caesar, Holger, Jasper Uijlings, and Vittorio Ferrari. “COCO-Stuff: Thing and Stuff Classes in Context.” In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1209–1218. IEEE Computer Society, 2018. https://doi.org/10.1109/CVPR.2018.00132.

Cao, Qiong, Li Shen, Weidi Xie, Omkar M. Parkhi, and Andrew Zisserman. “VGGFace2: A Dataset for Recognising Faces across Pose and Age.” In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 67–74. IEEE Computer Society, 2018. https://doi.org/10.1109/FG.2018.00020.

Cavallotti, Diego. “L’audiovisivo analogico della quotidianità. Discorsi, pratiche e testi del cinema e del video amatoriale tra gli anni Settanta e gli anni Novanta in Italia.” PhD diss., University of Udine, 2017.

Chalfen, Richard. *Snapshot Versions of Life*. Bowling Green, OH: Bowling Green State University Popular Press, 1987.

Drucker, Johanna. *The Digital Humanities Coursebook: An Introduction to Digital Methods for Research and Scholarship*, 1st ed. Abingdon, Oxon; New York: Routledge, 2021.

Eastman Kodak Company. *Instructions for Use of the Ciné-Kodak, Model K*. Rochester, New York, ca. 1930.

Enticknap, Leo. *Moving Image Technology: From Zoetrope to Digital*. London, New York: Wallflower, 2005.

Fickers, Andreas, Juliane Tatarinov, and Tim van der Heijden. "Digital History and Hermeneutics – Between Theory and Practice: An Introduction." In *Digital History and Hermeneutics*, edited by Andreas Fickers and Juliane Tatarinov, 1–20. Berlin: De Gruyter, 2022. https://doi.org/10.1515/9783110723991-001.

Flückiger, Barbara. "A Digital Humanities Approach to Film Colors." *The Moving Image* 17, no. 2 (2017): 71–94. https://doi.org/10.5167/uzh-151113.

Heller, Franziska. *Update!* Paderborn: Wilhelm Fink Verlag, 2020. https://doi.org/10.30965/9783846764602.

Holthuizen, Frank, Joseph Wachelder, and Tim van der Heijden. "Amateurfilm in Limburg, Limburg in Amateurfilm." In *Publications de la Sociéte Historique et Archéologique dans le Limburg*, vol. 156, 229–281. Maastricht: Koninklijk Limburgs Geschied- en Oudheidkundig Genootschap LGOG, 2021.

Ishizuka, Karen L., and Patricia R. Zimmermann, eds. *Mining the Home Movie: Excavations in Histories and Memories*. Berkeley: University of California Press, 2008.

Kattelle, Alan. *Home Movies: A History of the American Industry, 1897–1979*. Nashua, NH: Transition Pub., 2000.

Koolen, Marijn, Jasmijn Van Gorp, and Jacco van Ossenbruggen. "Toward a Model for Digital Tool Criticism: Reflection as Integrative Practice." *Digital Scholarship in the Humanities* 34, no. 2 (2019): 368–385. https://doi.org/10.1093/llc/fqy048.

Kuypers, Valentine. "Amateurfilm Platform." Presentation at the "e-nedits" meetings, INEDITS Amateur Films/Memory of Europe, October 13, 2020. Accessed March 22, 2024. http://en.inedits-europe.org/News/Meetings/2020-the-e-nedits-year.

Lokoč, Jakub, Gregor Kovalčik, Tomáš Souček, Jaroslav Moravec, and Přemysl Čech. "A Framework for Effective Known-Item Search in Video." In *Proceedings of the 27th ACM International Conference on Multimedia*, 1777–1785. New York: Association for Computing Machinery, 2019. https://doi.org/10.1145/3343031.3351046.

Lupu, Ruxandra. "The Home Movie 4.0: (Co)Creative Strategies for a Tacit, Embodied and Affective Reading of the Sicilian Home Movie Archive." PhD diss., University of Leeds, 2020. Accessed August 30, 2023. https://etheses.whiterose.ac.uk/27966/.

Manovich, Lev. *Cultural Analytics*. Cambridge, MA: MIT Press, 2020.

Manovich, Lev. *The Language of New Media*. Cambridge, MA: MIT Press, 2001.

Masson, Eef, Christian Gosvig Olesen, Nanne van Noord, and Giovanna Fossati. "Exploring Digitised Moving Image Collections: The SEMIA Project, Visual Analysis and the Turn to Abstraction." *Digital Humanities Quarterly* 14, no. 4 (2020). Accessed March 30, 2024. http://digitalhumanities.org/dhq/vol/14/4/000497/000497.html.

Moran, James M. *There's No Place like Home Video*. Minneapolis: University of Minnesota Press, 2002.

Motrescu-Mayes, Annamaria, and Susan Aasman. *Amateur Media and Participatory Cultures: Film, Video, and Digital Media*. London: Routledge, 2019.

Noord, Nanne van, Christian Gosvig Olesen, Roeland Ordelman, and Julia Noordegraaf. "Automatic Annotations and Enrichments for Audiovisual Archives." In *Proceedings of the 13th International Conference on Agents and Artificial Intelligence – Volume 1: ARTIDIGH*, 633–640. Vienna: SciTePress, 2021. https://doi.org/10.5220/0010387706330640.

Odin, Roger, ed. *Le film de famille: usage privé, usage public*. Paris: Méridiens Klincksieck, 1995.

Ramsay, Stephen. *Reading Machines: Toward an Algorithmic Criticism*. Urbana, IL: University of Illinois Press, 2017. https://doi.org/10.5406/illinois/9780252036415.001.0001.

Rascaroli, Laura, Gwenda Young, and Barry Monahan, eds. *Amateur Filmmaking: The Home Movie, the Archive, the Web*. New York: Bloomsbury Publishing, 2014.

Redfern, Nick. "The Log-Normal Distribution is Not an Appropriate Parametric Model for Shot Length Distributions of Hollywood Films." *Digital Scholarship in the Humanities* 30, no. 1 (2015): 137–151. https://doi.org/10.1093/llc/fqs066.

Rigole, Jasper. *Addenda*. Ghent: AraMER, 2015. Accessed August 20, 2024. https://www.jubilee-art.org/projects/addenda.

Roepke, Martina. *Privat-Vorstellung: Heimkino in Deutschland vor 1945*. Hildesheim: G. Olms, 2006.

Salazkina, Masha, and Enrique Fibla-Gutiérrez, eds. *Global Perspectives on Amateur Film Histories and Cultures*. Bloomington: Indiana University Press, 2021.

Schneider, Alexandra. *Die Stars sind wir: Heimkino als filmische Praxis*. Marburg: Schüren, 2004.

Shand, Ryan. "Amateur Cinema: History, Theory and Genre (1930–80)." PhD diss., University of Glasgow, 2007.

Simoni, Paolo. "The Amateur City: Digital Platforms and Tools for Research and Dissemination of Films Representing the Italian Urban Landscape." *The Moving Image: The Journal of the Association of Moving Image Archivists* 17, no. 2 (2017): 111–118. https://doi.org/10.5749/movingimage.17.2.0111.

Slootweg, Tom. "Resistance, Disruption and Belonging: Electronic Video in Three Amateur Modes." PhD diss., University of Groningen, 2018.

Strien, Daniel van, Kaspar Beelen, Melvin Wevers, Thomas Smits, and Katherine McDonough. "Computer Vision for the Humanities: An Introduction to Deep Learning for Image Classification (Part 1)." *Programming Historian*, August 17, 2022. Accessed March 22, 2024. https://programminghistorian.org/en/lessons/computer-vision-deep-learning-pt1.

Strien, Daniel van, Kaspar Beelen, Melvin Wevers, Thomas Smits, and Katherine McDonough. "Computer Vision for the Humanities: An Introduction to Deep Learning for Image Classification (Part 2)." *Programming Historian*, August 17, 2022. Accessed March 22, 2024. https://programminghistorian.org/en/lessons/computer-vision-deep-learning-pt2.

Sui, Patrick. "Review: Distant Viewing Toolkit." *Reviews in Digital Humanities* 2, no. 4 (2021). https://doi.org/10.21428/3e88f64f.01fedc58.

Tepperman, Charles. "The Amateur Movie Database." *Screen* 61, no. 1 (2020): 124–128. https://doi.org/10.1093/screen/hjaa007.

Tepperman, Charles. "The Complex Materiality of Amateur Cinema Research: Texts, Archives and Digital Methods – Introduction." *Screen* 61, no. 1 (2020): 119–123. https://doi.org/10.1093/screen/hjaa006.

Tepperman, Charles. "The Amateur Movie Database: Archives, Publics, Digital Platforms." *The Moving Image: The Journal of the Association of Moving Image Archivists* 17, no. 2 (2017): 106–110. https://doi.org/10.5749/movingimage.17.2.0106.

Tepperman, Charles. *Amateur Cinema: The Rise of North American Movie Making, 1923–1960*. Oakland, California: University of California Press, 2015.

Van der Heijden, Tim. "Technologies of Memory: Amateur Storage Media and Home Movie Practices in the Longue Durée." *Le Temps Des Médias* 39, no. 2 (2022): 141–159. https://doi.org/10.3917/tdm.039.0141.

Van der Heijden, Tim. "Hybrid Histories: Technologies of Memory and the Cultural Dynamics of Home Movies, 1895–2005." PhD diss., Maastricht University, 2018.

Van der Heijden, Tim, and Valentine Kuypers. "Life in Motion: A History of Amateur Film." *Europeana*, April 12, 2023. Accessed August 8, 2024. https://www.europeana.eu/en/exhibitions/life-in-motion/.

Wevers, Melvin, and Thomas Smits. "The Visual Digital Turn: Using Neural Networks to Study Historical Images." *Digital Scholarship in the Humanities* 35, no. 1 (2020): 194–207. https://doi.org/10.1093/llc/fqy085.

Zaagsma, Gerben. "Digital History and the Politics of Digitization." *Digital Scholarship in the Humanities* 38, no. 2 (2023): 830–851. https://doi.org/10.1093/llc/fqac050.

Zimmermann, Patricia R. *Reel Families: A Social History of Amateur Film*. Bloomington: Indiana University Press, 1995.

Franziska Heller and Ulrich Ruedel

Pursuing Film History with Digital Images: Towards Visual Literacy in the Age of AI and Social Media

Divergent Perspectives on Digitized Film Heritage

To propose that film restorers need to be aware of YouTube, and film scholars require a sufficient understanding of digital film restoration may be greeted by surprise from either group. However, these conclusions from the following deliberations are inspired by observations from our research practice, and by a critical admission to begin with: is it not quite common for a media scholar to occasionally (and conveniently) consult or even present exemplary film clips or archival film excerpts through the YouTube platform?

Have you ever wondered how much advertising thus inadvertently enters the academic classroom – it being, after all, an intrinsic part of the platform logic? Can you ascertain what kind of images you are watching or showing? Which version of the film is a certain clip from? Is the aspect ratio correct? What speed is a silent film clip transferred at? Is there any cropping or, worse, stretching of the source material? How about the colors? More fundamentally speaking, do you consider it relevant what visual (and auditive) source material you are working with in this (somehow) digitized form? Considering that such imagery has become the norm on platforms and within the algorithmic and indeed social network known as "the internet," how does this issue shape the specific discourse about archival images framed by such forms of communication? Does this media environment influence or even shape (or re-shape) the audience's attitude towards historic film images?

One might also turn around the perspective to reflect on the very target group – the next generation of students and scholars: are these not, for the most part, "digital natives," socialized within a digital, hence entropic media environment?[1] An MA student, when recently asked if she had seen a specific film from the 1960s, responded: "I wanted to watch it, but it was *nowhere* to be found, it

1 Cf. Malte Hagener, "Cinephilia and Film Culture in the Age of Digital Networks," in *The State of Post-Cinema: Tracing the Moving Image in the Age of Digital Dissemination*, ed. Malte Hagener, Vinzenz Hediger, and Alena Strohmaier (London: Palgrave Macmillan, 2016), 191.

https://doi.org/10.1515/9783111082486-013

was not available."[2] What the student actually implied, of course, was that the film was not freely accessible online. Indeed, the idea that it could still be distributed and be very much accessible on a digital carrier such as DVD or Blu-ray, a physical access medium, probably did not even cross her mind, since she may very well belong to a generation that has completely and exclusively adapted to streaming media content rather than physical media. Ironically though, such a physical release may oftentimes and illegitimately be the very source of the even further compressed video she clearly expected to find online.

Becoming increasingly aware of such issues relevant for academic researchers, teachers, and students, we decided to engage in a transdisciplinary dialogue informed by the realization that we both teach specific approaches to film history, hence we both develop specific forms of historiography employing digital images. As professors in distinctly different study programs, we pursue very divergent study goals (more details in the following sections of this chapter). Yet both programs practice approaches that ultimately lead to specific *views* on what film history is, how film history is understood, researched, preserved, and shared, and thus transferred and communicated to future generations.

We emphasize the term *view* on film history because it is important to note that we redirect the focus on "doing film history" by working with actual images – which, in most cases, reach us and the classroom in the aforementioned digital forms and versions. In the following, we try to situate and systematize the individual disciplinary perspectives related to the broad spectrum of images digitized from our photochemical audiovisual heritage. We thus aspire to establish a broader synergistic methodology that leads to a more comprehensive and nuanced understanding of what visual literacy of historical images should comprise in today's media culture landscape.

Theoretical Concepts and Specific Approaches

From the outset, our dialogue commenced under two specific and mutually agreed theoretical premises. The term "digital" is complex and in need of further deliberation and definition, as is also evident in the variety of implicit or explicit understandings within the different chapters of this volume. Every approach to the new directions that "digital" might bring to film historiography has to be informed by its understanding of the term, depending on the definition of where

2 Emphasis added.

the specific medial qualities, potentials, and characteristics of digital tools, data, or media are situated.[3]

In our approach the notion of "digital" is linked to the discourse of image theory. We understand digital data as the underlying code that is rendered by the program in order to become a visible phenomenon. On the one hand, the visible appearance is defined by the mathematical specifications of the program, such as color space, contrasts, gamma, etc. The binary code and its modular quality also entail the so-called malleability or flexibility of visual appearance, widely discussed since the 1990s, and contributing to the narrative that digital images are so very prone and open to manipulation.[4] In several of her publications, Barbara Flückiger in particular[5] has elaborated from a "technobol" perspective[6] on how code properties of digital ecosystems influence the visual experience – especially that of originally analog, subsequently digitized material.

On the other hand, as discussed in detail in Franziska Heller's book *Update!*,[7] image digitization also needs to be understood as a process of cultural interpretation. There are many sociocultural parameters that implicate how the code takes a specific visual and aesthetic shape in the process of becoming an image.[8] Furthermore, the fact that the visual qualities of digital images can differ considerably (that is, depending on the environment and particularly the display device used to watch the images) has also frequently been addressed.[9] These questions

3 Cf. Franziska Heller, *Update! Film- und Mediengeschichte im Zeitalter der digitalen Reproduzierbarkeit* (Paderborn, Munich: Brill, Wilhelm Fink, 2020), https://doi.org/10.30965/9783846764602. Chapter 2 discusses the various approaches to the term "digital" and the consequences for different fields of Media Studies research.

4 Cf. William J. Mitchell, *The Reconfigured Eye: Visual Truth in the Post-Photographic Era* (Cambridge, MA: MIT Press, 1992); Lev Manovich, *The Language of New Media* (Cambridge, MA: MIT Press, 2001); Barbara Flückiger, *Visual Effects: Filmbilder aus dem Computer* (Marburg: Schüren Verlag, 2008).

5 See, among others, Barbara Flückiger, "Material Properties of Historical Film in the Digital Age," *NECSUS. European Journal of Media Studies* 1, no. 2 (2012): 135–153, https://doi.org/10.5117/NECSUS2012.2.FLUE.

6 Flückiger, *Visual Effects*, 16.

7 Heller, *Update!*, 29–84 (chapter 2).

8 Cf. Barbara Flückiger et al., "'Digital Desmet': Translating Early Applied Colors," *The Moving Image* 16, no. 1 (2016): 106–124. The authors develop an approach called "archival pragmatics," where processes of film digitization are critically analyzed and systematized. This also includes (not necessarily comprehensively) selection categories of films for digitization and aspects of funding, but also technological considerations.

9 Flückiger et al., "'Digital Desmet'"; Claudy Op den Kamp, Barbara Flückiger, and David Pfluger, "A Material-Based Approach to the Digitization of Early Applied Colors," in *The Colour Fantastic: Chromatic Worlds of Silent Cinema*, ed. Giovanna Fossati et al. (Amsterdam: Amsterdam University Press, 2018), 237–259.

could also be considered as a configuration of the apparatus or – in a broader, more Foucauldian sense – of the dispositive.[10]

When using the term "digital" in this paper, we emphatically and deliberately include all the levels mentioned above. We consider the technological data basis as well as the perceptual and sensual properties of images.[11] We also include the forms of communication of current media culture, i.e., how the user experience is shaped within a certain type of dispositive and in terms of the psychological and sociocultural aspects of image perception.

Referring to the German cultural historian Hartmut Böhme,[12] Heller has previously highlighted that such user experience is intertwined with the so-called "technological imaginary."[13] Such an imaginary is shaped by recurrent narratives surrounding digital media and digital images, and influences forms of usage and expectations of the perceived aesthetics of images circulating within digital media culture. Böhme's concept of the imagology of technology proves very productive when applied to the systematic analysis of digitized moving images within the internet and social media. He emphasizes the recurring metaphors and rhetorics of the fantastic: expressions referencing "magic" and "wizardry" are established tropes.[14] The fetishistic reference to magic as a specific form of the imaginary applies even more to digital media where the structural and functional foundation of the code remains opaque; therefore, the aesthetic design of the mode of access and reception gains importance. An obvious example from today's popular digital image culture is Instagram's *mise-en-scène* of how the user can interact with and modify digital images: "Tap the 'magic wand' on the upper right-hand corner of

10 Heller, *Update!*, 98–194.

11 In this regard our perspective differs from other approaches discussed in this volume. In our reading, these often use the term data for meta information of sociocultural phenomena expressed in numbers to be (re-)interpreted, read, and harvested in new and different ways by means of suitable digital tools.

12 Hartmut Böhme, *Kulturgeschichte der Technik*, 2000, accessed June 24, 2023, https://www.hartmutboehme.de/static/archiv/volltexte/texte/kgdt.html.

13 Heller, *Update!*, 75; furthermore: esp. 263–317 (chapter 6).

14 Barbara Klinger has already pointed out similar rhetorics in the debate revolving around the digitization of film heritage. She describes the recurring references in promotional taglines to the idea of "digital wizardry": Barbara Klinger, *Beyond the Multiplex: Cinema, New Technologies, and the Home* (Berkeley, Los Angeles: University of California Press, 2006), 122. The phenomenon has been discussed extensively within the popular contextualization of film restorations in DVD-editions, focusing on the audiovisual paratexts, such as the so-called film-historical documentaries or restoration documentaries. Special attention has been given to the historiographic narratives, as well as the particular aesthetics. Cf. Heller, *Update!*, esp. 263–317 (chapter 6).

your screen." Or, as the German Instagram Business Team prompts on their website promoting the app, "try your own magic tricks and defy the laws of physics!"[15]

Returning to our initial observations, we propose that the technological imaginary as an element of discourse has to be acknowledged when digitized archival images circulate within platforms such as YouTube, Instagram, or Facebook. The modes of discourse are formed by the affordances,[16] the logics of attention, the aesthetics, and the narratives established on and by these platforms: images circulate, they are re-used, re-worked, appropriated, shared, and commented upon.[17] Thus, they build and inform public taste and attitude towards images. In that sense, we link the usage of digital images to the digital media culture where specific modes of dissemination and communication are ritualized, habitualized, and thus normalized.

In the case of imaging practices within the digital culture, such as the application of filters[18] or digital coloring, tropes and questions emerge that were already discussed in earlier historical cases, such as the colorization debate in the 1980s. That past and in-depth debate has indeed led to academic statements that remain more valid than ever whenever it comes to situating popular (media) culture in relation to a notion of film history. Discussing the historical case of electronic film colorization and other modifications of audiovisual images within the apparatus of television and video cassette, Stuart Klawans, as early as 1990, raised the crucial question: "Where does film history really reside?"[19] Charles Acland fur-

15 Instagram Business Team, https://www.simplygram.com/the-best-instagram-filters-and-other-awesome-tricks-you-should-be-trying-now, accessed March 30, 2021. See also https://business.instagram.com/blog/introducing-face-filters?locale=de_DE, accessed March 30, 2021.

16 The term "digital affordances" has been used within the critical analysis of the forms of interaction that digital media and platforms offer. See Janet Murray, *Inventing the Medium: Principles of Interaction Design as a Cultural Practice* (Cambridge, MA: MIT Press, 2011); Catherine Grant, "The Shudder of a Cinephiliac Idea? Videographic Film Studies Practice as Material Thinking," *Aniki* 1, no. 1 (2014): 49–62, https://doi.org/10.1515/9789048543922-012. Among others, Bucher and Helmond discuss the origin of the term in ecological psychology; see Taina Bucher and Anne Helmond, "The Affordances of Social Media Platforms," in *The SAGE Handbook of Social Media*, ed. Jean Burgess, Alice E. Marwick, and Thomas Poell (London: Sage Publications, 2018), 233–253. The study of the relationship between possible interactions for the individual and the physical environment based on concepts of cognitive psychology have proven useful for design studies. This approach has allowed us to reflect on design patterns and on structures shaped by technologies, as well as on their usability.

17 Henry Jenkins, *Fans, Bloggers, and Gamers: Exploring Participatory Culture* (New York: New York University Press, 2006); Pelle Snickars and Patrick Vonderau, eds., *The YouTube Reader* (Stockholm: National Library of Sweden, 2009).

18 Katja Gunkel, *Der Instagram-Effekt* (Bielefeld: Transcript, 2018).

19 Stuart Klawans, "Colorization: Rose-Tinted Spectacles," in *Seeing Through Movies*, ed. Mark Crispin Miller (New York: Pantheon, 1990), 166.

ther relates the discussion about the notion of film history to the dynamics of popular culture.[20] Referring to Bennett and Woollacott, Acland defines the essential quality of popular media culture: an "'incessantly mobile reordering of the relations' between texts is an essential quality of popular culture: its movement, and, by the same token, its resistance to the kind of stability signified by museums, archives, catalogues, artists, and originals."[21]

Once we include contemporary popular digital culture and its forms of communication in our approach to archival images, a much more dynamic and performative understanding of moving image history emerges beyond the classical institutions and gateways for dissemination. When we take into account current medial forms of reception, exchange, sharing, and commenting, statements by French film and media historian Pierre Sorlin appear in a fresh context. Sorlin states that history does not exist outside the discourse that is enunciating that history. History can thus appear in different forms, depending on the means of its realization and expression.[22]

On a more methodological level, our approach could thus be described as an adaptation of the idea of digital hermeneutics. We propose that an awareness and critical understanding of the link between analog preconditions of moving images and their transition into the digital realm – including its communicative modes of dissemination, appropriation, framing, and attention logic – is mandatory. We thus aim to link our approach to the concept of digital source criticism,[23] with the goal of applying it to digital images. We shall illustrate our approach to visual source criticism with the case studies presented in the fourth section of this chapter.

Thus, in the following, we propose what we believe is an urgently needed, specifically visual literacy, and in a two-fold, transdisciplinary manner at that. What film studies can learn from film restorers is the lesson that, while the im-

20 Charles R. Acland, "Tampering with the Inventory: Colorization and Popular Histories," *Wide Angle* 12, no. 2 (1990): 12–20.

21 Acland, "Tampering with the Inventory," 16.

22 Pierre Sorlin, "Ist es möglich, eine Geschichte des Kinos zu Schreiben?" *Montage AV* 5, no. 1 (2009): 23–37.

23 Hence the methodology we suggest here could be understood as complementing existing approaches within the field of digital source criticism. See, among others, Julia Noordegraaf, Kathleen Lotze, and Jaap Boter, "Writing Cinema Histories with Digital Databases: The Case of Cinema Context," *TMG Journal for Media History* 21, no. 2 (2018): 106–126, https://doi.org/10.18146/2213-7653.2018.369; Julia Noordegraaf, "Zooming Out: Towards Scalable Digital Film Studies" (presentation and discussion at the workshop Teaching Digital Methods for Film Historiography, Johannes Gutenberg University Mainz, Germany, within the DFG-Network New Directions in Film Historiography, February 18, 2021).

ages we work with are indeed digital, the material basis of film images has to be taken into account to understand their historiographic impact – although it may indeed be difficult to conclusively determine what the "original" material artefact is, even for the analog medium that is film (due to its "reproducibility" as per Walter Benjamin[24]). Conversely and consequently, film restorers cannot ignore the media environment for which they try to restore images in the first place, nor where they may end up for better or worse – a digital discourse and consumption space where these images are then recontextualized, framed, and most often perceived. In other words, the forms of communication are bound to (re-)shape public attitude towards archival images through their exponentially proliferating, playful modification. This incremental yet tectonic shift offers methodological challenges regarding our notion of historiography.

Divergent Methodologies: From Interdisciplinary Dialogue to Transdisciplinary Pragmatics

The transdisciplinary dialogue between Media Cultural Studies, Film Studies, and moving image heritage practices outlined in the introduction revealed two, sometimes instructively conflicting, conceptual frameworks.

(1) One main objective of the field of moving image preservation is to adhere to guidelines and ethics regarding current digital restoration practices. Practitioners are expected to act as responsible film restorers that respect the historical artefact and its historic, technical, and material roots. Especially with regard to digital imaging technologies, restorers have to learn to evaluate their choices while literally working *on and within* the images. The restorer must be able to distinguish between subsequent damage or wear versus traces of the historical original production process, and analog versus digital artifacts, such as those accompanying the digitization and digital restoration processes. Although there is no lack of academic discourse on restoration ethics, such as Wallmüller's texts,[25] implementing them in actual restora-

24 Walter Benjamin, *Das Kunstwerk im Zeitalter seiner technischen Reproduzierbarkeit* [1936] (Frankfurt am Main: Reclam, 1977).

25 Julia Wallmüller, "Criteria for the Use of Digital Technology in Moving Image Restoration," *The Moving Image* 7, no. 1 (2007): 78–91; Julia Wallmüller, "A Classical Approach to Authentically Restoring Film," *Aesthetic Investigations* 2, no. 2 (2019): 144–162.

tion practice (remember G. Fossati's famous application of the SPIDER-MAN [Sam Raimi, 2002] quote for film restoration: "With great power comes great responsibility")[26] is still riddled with grey areas – a void FIAF's recent Digital Statement III tried to address by offering more specific guidelines and proposing firm "red lines."[27] As evidenced in numerous film restoration case studies,[28] it is obvious that many restorations pose different specific questions for digital imaging and that all digital handling needs to be related to the specific historic context of the source material. This includes researching and respecting the visual characteristics and limitations of, say, a historic color system or special effect technique, acknowledging specific circumstances in media history, investigating the genealogy of the analog materials and channels of circulation, and oftentimes the elusive concept of "artistic intent."[29] Thus, restorers need guidelines on how to evaluate the options of their actions when they execute the actual restoration work. The result will be visible for the contemporary public and build aesthetic taste for the future. Recurrent key questions revolve around the definition of "authenticity" and the "proper" choice of reference regarding what constitutes the "original."

(2) In turn, for the field of digital culture studies as practiced in the department for Media and Communication Studies at Martin-Luther-University (MLU), and specifically by the chair for Digital Cultures, the focus of interest is on practices shaping the modes of digital expression and communication that appropriate archival moving images. This can be observed on platforms such as Facebook, Instagram, Twitter, and especially in relation to so-called meme culture. This critical

26 Giovanna Fossati, *From Grain to Pixel: The Archival Life of Film in Transition* [2009], 3rd rev. ed. (Amsterdam: Amsterdam University Press, 2018), 98.

27 Cf. Robert Byrne et al., "The Digital Statement III. Image Restoration, Manipulation, Treatment, and Ethics," *Journal of Film Preservation* 104 (2021): 25–37.

28 See, for instance, Scott MacQueen, "Technicolor and the True Believer," *The Perfect Vision* 3, no. 12 (1991–1992): 23–38; Scott MacQueen, "Mystery of the Wax Museum: Restoring Two-Color Technicolor with Digital Tools," *Journal of Film Preservation* 103 (2020): 104–114; and for reflections on digital restoration ethics, Wallmüller, "Criteria for the Use of Digital Technology" and "A Classical Approach to Authentically Restoring Film."

29 For the classical restoration approach, consider Nicholas Stanley Price, M. Kirby Talley Jr., and Alessandra Melucco Vaccaro, eds., *Historical and Philosophical Issues in the Conservation of Cultural Heritage* (Los Angeles: Getty Conservation Institute, 1996), especially the texts by Albert Albano, John Richardson, and Ernst van de Wetering in "Part II: The Original Intent of the Artist" (161–199), and the texts by E. H. Gombrich, John Ruskin, and Bernard Berenson in "Part I: The Eye's Caress: Looking Appreciation, and Connoisseurship" (110–138). See also Michael Kaminski, "Saving Star Wars: The Special Edition Restoration Process and Its Changing Physicality," *The Secret History of Star Wars*, 2009, accessed June 29, 2023, http://fd.noneinc.com/secrethistoryofstarwarscom/secrethistoryofstarwars.com/savingstarwars.html.

perspective aims at the identification of socio-cultural frameworks that foster such practices. How are specific forms of usage and forms of visual communication naturalized and normalized as carriers of affect and emotions? What consequences do these practices of visual discourse entail for our understanding of film history? What historiographic narratives are implicitly promoted in the repurposing and modification of images?

Inspection of the discursive framing of such practices reveals a precarious ambiguity of terms such as "restoration" and "remastering," and thus demands caution. This critical perspective includes phenomena where archival images within the digital remix culture are turned into an aesthetic and communicative commodity (cf. memes based on imagery from classical films). Most importantly, the film studies approach contributes a more discursive understanding of historic "authenticity." With reference to the theory of documentary film and semio-pragmatics, inspired by the original concept by Roger Odin, (historic) "authenticity" can be modelled as a medial effect, which is induced not only by certain internal aesthetic strategies of the images, but also by external rhetoric and by the pragmatic, communicative context.[30]

Obviously an understanding of authenticity as an effect of medial configurations can be (incorrectly, we argue) perceived as a conflict with the need of restorers to have a clear definition of what the actual reference is: to put it in simplified form, the restorer considers historic authenticity as faithfulness to history as conveyed by and seen through its sources, while a film scholar might focus – among many other questions – on the specific sociocultural preconditions and medial framings which create the impression of historic authenticity. This might even lead to the analysis of why and how a general public may seek emotional immersion as the preferred nostalgic mode of engagement with history.

This brings us back to the impetus for our interdisciplinary dialogue: we do not understand these positions to be mutually exclusive. Rather, we ask ourselves what can be productively taken from the two perspectives in order to develop a transdisciplinary framework to lay the foundation for a critical visual literacy, which might lead to new pragmatics in doing film history. This is even more true in the light of new digital technologies like artificial intelligence, and digital platforms like social media that function as popular discourse spaces. We will illustrate our endeavor in the following section by discussing case studies on the colorization, creation, and enhancement of star images of Louise Brooks, AI generated imagery of comic icon Louis de Funès, and AI colorized and digitally allegedly "enhanced"

30 Roger Odin, "Dokumentarischer Film – Dokumentarisierende Lektüre [1984]," in *Sprung im Spiegel*, ed. Christa Blümlinger (Vienna: Sonderzahl, 1990), 125–146; Heller, *Update!*, 109–113.

Lumière films and other early film imagery. In doing so, we discuss what we consider the importance of attention logics, image and color fetishism, the impact of the online communication sphere, the impending consequences of the AI imaginary, and the cultural narrative of technological progress. In conclusion, towards the realization of a truly transdisciplinary approach, we then propose specific fields of further research aimed at a specific visual literacy for historical images and their circulating digital derivatives.

Case Studies: Film Restorers and the Internet Discourse. Attention Economy, Affective Communication and the Visible Spectacle of AI

"Let the hate begin – LOL" – Colorization Debates Then and Now

As early as in texts dating back to the 1980s, followed by the publication of a book in 1991, Frederic Jameson offered a cultural diagnosis for the globalized "postmodern, late capitalist" social structures. His conclusions turn out to be equally helpful for describing today's phenomena, such as the proliferation of digital colorizations and the use of filters in communication across digital platforms. They are even more valuable within the context of the perception and assessment of historicity. Jameson argues that aesthetic production has increasingly become an economically exploitable commodity,[31] and discusses in particular the imaginary and transcendental dimension attributed to digital technological developments in such a social environment. Jameson may not yet have considered an almost entirely digitized world of communication. But his further diagnoses, associating mass media phenomena in this sociocultural environment with certain aesthetics marked by recycling and quotation as principles, are entirely applicable to the digitized realm, especially when considering the oft-quoted description of digital culture as essentially a remix culture.

As already mentioned, the 1980s offer a historical precedent in which technological development, economic interests, new media environments, and aesthetic forms came into tension in the then-emerging practice of so-called colorization or

31 Fredric Jameson, *Postmodernism, or, the Cultural Logic of Late Capitalism* (Durham, NC; London: Duke University Press, 1991), 4.

color conversion:[32] colorization for the American television and video market was a historical practice where audiovisual heritage was electronically and aesthetically altered and "updated" as an economic model, that is, to maintain or increase value for the market.[33]

In addition to colorization, it is useful to consider other changes that films and moving images can undergo as they migrate from photochemical film and the movie theatre to other media and media dispositives. This was most obviously the case with television transforming films through such practices as pan and scan, cropping of the frames, reediting, etc. But these and other modifications also occurred in both obvious or subtle forms when films were distributed in "digital dispositives," such as DVD or Blu-ray.[34]

The problems and questions already provoked by distributional practices in the history of media are still pivotal today in view of digital interventions on moving images, especially in the globalized communication sphere of the internet and its myriad platforms. However, it is important to look closely at specific economies of attention and thus market mechanisms that shape the re–inventing and dissemination of images. With a basic assessment of what constitutes the fascination in the reworking of existing film images and why the practice appears economically attractive when in circulation, Acland made an important specification: it is not necessarily the manifest leap in quality that is decisive, but the very *spectacle* of (the possibility of) some promoted "improvement," of update, of renewal: "Contrary to popular opinion, color per se does not attract audiences. It is the colorised – the spectacle of the refinished product, a creation of technological wizardry – that succeeds in doing so."[35]

This observation still holds true for the digital realm, through the discursive framing in the digital communication sphere (with its own logic of spectacle and attention), as evidenced in the following example. Emphatically proclaiming "Let the hate begin – LOL," on January 26, 2021, a user posted a colorized photographic portrait of silent film icon Louise Brooks on Facebook (Figure 1).

The picture originated from a blog called "Colors by Klimbim," which stated its mission in an early web page version as follows: "This blog is meant to share the images found on the Internet and colorized by me with other bloggers and

32 Klawans, "Colorization: Rose-Tinted Spectacles," 159.

33 Cf. Acland, "Tampering with the Inventory," 12.

34 Cf. Jan Distelmeyer, *Das flexible Kino. Ästhetik und Dispositiv der DVD & Blu-ray* (Berlin: Bertz und Fischer, 2012); Heller, *Update!*.

35 Acland, "Tampering with the Inventory," 15. Emphasis added.

Figure 1: Colorized photographic portrait of Louise Brooks shared on Facebook with the declared purpose to trigger emotional responses. Screenshot by the authors. Source: https://www.facebook.com/groups/2204605501/permalink/10158961416665502 (accessed April 15, 2024).

colouring enthusiasts."[36] The mentioning of "colouring enthusiasts" points to a larger community of similar online users constantly producing, sharing, and engaging with a myriad of modified and remodified archival images as a ritualized, habitualized, and normalized imaging practice. However, the Facebook post referred to above tries to stoke online reaction beyond the like-minded and approving communicative sphere of "DIY colorists": the post, at the time of writing, was already commented on 118 times. It blends the "spectacle" of the colored image with the logic of the attention economy of social media in a textbook fashion, with its directly emotionalizing and perceivably provocative address to friends and followers from a different communicative "bubble," that of film historians, since the image was posted in *The Silent Film Group*, among others. "Let the hate begin – LOL" is, after all, an ironic provocation in announcing what would be an

36 The quote continues "and [it] is not meant to deliberately infringe on the rights of the image owners." Cf. Klimbim, accessed June 21, 2023, https://klimbim2014.wordpress.com/about/. We do not further consider here that the colorization debate is always accompanied by a question of image rights and ownership claims. Cf. Claudy Op den Kamp, *The Greatest Films Never Seen: The Film Archive and the Copyright Smokescreen* (Amsterdam: Amsterdam University Press, 2018).

absolute "no-go" from a historian's perspective. It is intended to generate lively follow-up communication and emotional reactions on the platform: such crossing of boundaries only works in a discourse in which the reference to a historically fixed original is still assumed, an original that presumably everybody within the discursive user bubble still has in mind in monochrome (black and white), since this is how most of the still image material of the silent era had been produced. The author has deliberately chosen an emotionally inciting wording to appeal to a discourse in which the digital coloring of the photo is perceived as historically inauthentic sacrilege. One may even assume that the object in the photo – movie icon Louise Brooks – adds to the emotionalization, since a kind of eroticized nostalgic fetishism surrounds female silent film stars (and especially Louise Brooks as one of the biggest). It is a fetishism in which the historical patina is part of the fascination.[37]

We can thus clearly observe how voyeuristic star fetishism of the silent era blends with the practices of the platform, where the spectacle of colorization is used as an affective provocation to initiate follow-up communication according to the logic of engagement in social media. Such colorized images now circulate and are in turn framed by means of digital visual discourse (comments, emojis, etc.[38]) within the digital ecosystem – with consequences for the reference status. Depending on the pragmatic context, the images can even become an encyclopedic reference – as the next example shows.

Archival Images and the AI Imaginary on the Internet: Paradoxes of DIY Restorations and Algorithmic 'Craftsmanship'

In April 2023, a Google search in German for famous French comedian Louis de Funès directed internet users to the German Wikipedia. The second accompanying photograph, in black and white, right next to the basic entry in the paragraph titled "Life," demands closer inspection, showing the likeness of the famous comic in portrait format, as he gives a thumbs up (or something resembling that gesture). The

37 Arguably, such colorizing devalues the uniqueness of select full color photographs that do exist, such as the George Zoller Autochromes preserved at George Eastman Museum depicting Charlie Chaplin in his iconic tramp costume, e.g., https://collections.eastman.org/objects/189079/charlie-chaplin-as-the-tramp, accessed June 23, 2023.

38 Cf. Crispin Thurlow, Christa Dürscheid, and Federica Diémoz, eds., *Visualizing Digital Discourse: Interactional, Institutional and Ideological Perspectives* (Berlin, Boston: De Gruyter Mouton, 2020), https://doi.org/10.1515/9781501510113.

picture caption first gives the name of the actor – as obvious as it is important, given that the basic biographic context of the comedian's life is established here. Yet, next to this, the image caption specifies: "AI-based generated portrait" (Figure 2).

Figure 2: AI generated portrait of Louis de Funès prominently displayed on top of the comedian's Wikipedia.de entry in March 2023. Source: Wikipedia, "Louis de Funès," https://de.wikipedia.org/wiki/Louis_de_Fun%C3%A8s (accessed March 25, 2023).

Careful inspection of the image reveals that the eyes of the person in the picture are of two different shades, and the eyes appear to be of different sizes. The lighting conditions are extremely confusing and far from natural, resulting in a rather uncanny valley-type depiction of the famous comic's flexible likeness. Leaving the technical shortcomings of the image aside, issues remain to be discursively analyzed. Despite all the scientific reservations about the internet encyclopedia and its status as user-driven, it is remarkable that such an image could be publicly placed within this specific context. The image conveys the discursive dimension of (1) being archival (because of a nearly patina-like monochrome nature); but, at the same time, (2) the image implies that someone used automated, self-learning "intelligent" technology to create an impression of the actor's physiognomy. The subtext is quite paradoxical: on the one hand, there is the narrative, evoked by the perceivably historic monochrome image tone, that the image proves the his-

toric and iconic status of the actor. On the other hand, someone appropriated the image and thus made it his/her own work which – another paradox – excels in its technological quality by being "created" by an artificial intelligence. This specific characteristic in turn implies that there was no human creator needed, with the image being purely data driven. The latter concept leads to the connotation of a technological "objective" quality for the picture. A new framing of the idea of historic authenticity is thus created within popular discourse. Within the historical, chronical context of a popular encyclopedic medium such as Wikipedia, an AI-generated image is normalized in relation to an image with historic reference character. The effect is amplified by the fact that the image has remained on the website for at least several weeks at the time of writing.

The example shows the necessity of a visual literacy that demands a critical view on the discursive framing of images, and on the pragmatic context and resulting referential character ascribed in the context of digital technologies. Most recently, such image discourse has been intimately linked to the challenges and opportunities of AI. It is necessary to have a closer look at the technological imaginary that is evoked: the example of the German Wikipedia article already points at the paradoxical quality of cultural narratives that come into play and that often stem from much older concepts from the analog era. Furthermore, a critical media studies perspective can productively adapt a vital question raised by the restoration practice: where does AI actually source its information in producing an image and how does it assemble it?

At this point, the dynamics of digitized images circulating on the internet becomes interesting. A postgraduate student's paper, inspired by our interdisciplinary dialogue, used a thought-provoking example to address the manifold discourses revolving around the conflation of film restoration, AI, and digital culture. Berenike Beigang, MA, examined the YouTube channel of notable "digital artist" Denis Shiryaev as the leading example for current image practices.[39] YouTube channels like *glamourdaze*[40] and *Nineteenth century videos. Back to life*,[41] as well as Shiryaev's channel, have proved extremely popular. Shiryaev manipulates material circulating on the web and re-arranges it, adding new levels of discourse – often in the context of AI algorithms: apparent "defects" are "corrected," the material is recolored, music and sound effects are added. Seemingly playfully and artistically, he

39 Berenike Beigang, "Erlaubt ist, was begeistert? Medien- und Technikgeschichte(n) am Beispiel des Filmes *The Flying Train* aus dem Jahre 1902 sowie seiner Bearbeitung durch Denis Shiryaev im Jahr 2020" (unpublished student paper, Martin Luther University Halle-Wittenberg, 2022).

40 See https://www.youtube.com/channel/UCywuq7AxUM4uYrVmSFm3Ezw, accessed June 24, 2023.

41 See https://www.youtube.com/c/nineteenthcenturyvideosbacktolife, accessed June 24, 2023.

showcases what is technically possible. Since 2020, Shiryaev has been the CEO of the company *neural.love*, which, according to its own statements, tries to explore the limits of artificial intelligence to "improve" images and videos, as well as "to make machine learning magic available to everyone."[42] He thus promotes the programming language Meta Language on his YouTube channel.[43]

For his project based on *Lumière* films,[44] Shiryaev sourced the images from "an amazing YouTube channel with some Lumière films uploaded with great source quality"[45] – a rip, it is reasonably safe to assume, from the official Blu-ray of the restored films![46] Thus, rather than starting from the best possible source material in terms of image generation and quality, as would be the standard for virtually any proper restoration or remaster project (the historic film elements, or an uncompressed, high-resolution scan thereof), Shiryaev acts pragmatically by restricting himself to what is available in the very online ecosystem where he aims to re-upload his version in modified form. Then, in a particularly ironic twist, the genuine restoration/digitization's original silent film speed correction by frame duplication, repeating individual (historic) frames for correct playback speed at the higher, modern 24 frames/second, needs to be undone by Shiryaev for his more invasive digital tools to do their task. After removing these duplicate frames, AI, based on its learned algorithms, then creates new intermittent, interpolated, artificial (i.e., ahistoric) frames to generate what is considered perceptually smooth motion, yielding a 60p (60 frames per second) result. One could

42 See https://neural.love/about, accessed February 14, 2022.

43 Cf. Beigang, "Erlaubt ist, was begeistert?" 8.

44 Denis Shiryaev, "The Lumière Brothers' collection, 1895–1902," accessed June 24, 2023, https://www.youtube.com/watch?v=YZuP41ALx_Q, published January 3, 2021.

45 Shiryaev himself describes the genesis of the project in the voice over in the opening of the video as follows: "A few months ago I found an amazing YouTube channel with some Lumière films uploaded with great source quality. It was hard to resist not doing some Machine Learning magic with them. And here is the result of a few months of processing and sound editing. In the next few minutes, I'll describe all the steps I took for this enhancement." But then Shiryaev himself cautions: "Before we begin, however, I want to raise awareness regarding the enhancements that machine learning enthusiasts – like me – are doing with the original source material. I kindly ask you to share this knowledge with a broader audience if you are a machine learning enthusiast like me. Some people mistakenly think that colors in this video are the original source color or that the source material had audio, or that these enhanced faces are real. All these enhancements were done to achieve a 'time travel experience' for artistic or entertainment purposes, and should not be considered as historically accurate data. Neural Network can only guess colors, new pixels when we're doing upscale, and faces when working with some facial restoration, all of them working with the source data, and then those algorithms generate new data based on their training set." Shiryaev, January 3, 2021.

46 France Télévisions Distribution, "Le Cinématographe 1895–1900 Lumière," Blu-ray ed., 2016.

demagogically argue that – if it is assumed, say, that the films were shot and transferred at an original frame rate of 18 frames per second and now interpolated up to a speed of 60 frames per second – the result would consist of 70% modern computer-generated imagery. And this is even before the software assigns somewhat hypothetical detail and color to the image. Arguably, with 70% artificial imagery/motion and 100% artificial color and modern added detail, an entirely new, if derivative, moving image work or remix is created.

Shiryaev might not even disagree with this conclusion, since he himself cautions and asks: "I kindly ask you to share this knowledge [. . .]. Some people mistakenly think that colors in this video are the original source color [. . .] or that these enhanced faces are real. All these enhancements were done [. . .] for artistic or entertainment purposes, and should not be considered as historically accurate."[47]

While Shiryaev celebrates and promotes the achievements of his software and programming language, he is seemingly oblivious to the fact that the online video he sources is apparently a further compressed online conversion of the restored Lumières' film commercial Blu-ray release. He thus follows the internet logic of acknowledging only those moving images already present in the online realm (and subject to further modification through the online community according to their liking).

A comparative case study presented by user Andy Myers in 2020 on YouTube,[48] thus within the same digital communication realm, addresses this very issue. For his video, [Actual 4K Scan] *The Arrival of a Train at La Ciotat Station* – Lumiére Brothers – 1896, Myers quips: "This is the real deal, no artificial preservatives or fillers. [. . .] No cropping out parts of the frame. No interpolating extra fake frames to boost the frame rate. No using a neural net to guess how to fill in the details missing from a muddy ultra-compressed 720p source." This video, Myers elaborates, is a "Video from 1080p Blu-ray source, gently upscaled back to 4k to boost the allowed bitrate on YouTube." Thus, he draws his video from the 4K source restoration, as published in higher compression and lower resolution (HD-1080p Blu-ray), and re-upscales this data-set to 4K in order to minimize the quality imposed by the YouTube-platform's compression, Myers does not share the "actual 4K scan." However, he does strive to source and maintain the best and most authentic quality in comparison with Shiryaev's workflow, while working within the constraints of the physical media marketplace and digital ecosphere.

It is instructive indeed to compare the result with Shiryaev's: zooming in to see the little girl on the train platform, we ironically see more detail in the "origi-

47 See n. 45 for the entire quote by Shiryaev.

48 Andy Myers, "[Actual 4K Scan] *The Arrival of a Train at La Ciotat Station* – Lumière Brothers – 1896," accessed June 24, 2023, https://www.youtube.com/watch?v=1FAj9fJQRZA, published February 5, 2020.

nal" monochrome image (along with the pattern of the photographic grain that has "analogically encoded" it) than in Shiryaev's digitally colored and smooth-patchy version (Figure 3). While Shiryaev and Myers have entirely different ideas of desirable "digital image quality," both demonstrate the need for source and workflow transparency and criticism within and beyond the online platform discourse.

Figure 3: Comparison compiled by the authors of image detail from Myers' (left) and Shiryaev's (right) respective YouTube versions of the Lumière Brothers 1896 THE ARRIVAL OF A TRAIN AT LA CIOTAT STATION, illustrating the need for digital source criticism in the context of moving images. Screenshots by the authors.

Yet at the time of writing this, Shiryaev's video has been watched an approximate 2.7 million times, while Myers' has attracted only 133,000 viewers. What is the digital landscape, what are its dissemination, communication, and discussion mechanisms that help explain this discrepancy? To elucidate this, we consider the case of the film THE FLYING TRAIN, originally from 1902. This film itself is already embedded in a narrative of progress, in which historical context, thematic motifs, and discourses of technology in the context of industrialization intertwine. The suspension monorail in Wuppertal, the most famous suspension railway in Europe, celebrated its opening in March 1901 as a means of transport – and at the same time as a symbolic landmark of one of Europe's oldest industrial and economic regions in the *Bergische Land* region in West Germany. The film shows a classic phantom ride across the river Wupper, through the steel struts of the support structure and the historic industrial center.

In August 2020, the Museum of Modern Art, New York, in the wake of the COVID-19 pandemic, streamed selected "film treasures" as part of the so-called Film Vault Summer Camp. Already in the first week, THE FLYING TRAIN was put online and is now freely available on MoMA's YouTube channel.[49] This 2-minute video film has also been appropriated by Denis Shiryaev. His version[50] of the archival film is presented in a specific way. The first opening text panel directly combines historical information about the place of filming and the year of production with the technical information about what the programmer did with the film. He once again motion-interpolated for speed correction, "added" resolution, and enhanced image stability: "60 fps, upscale to 4k, stabilization." In the following text panel, the reference to the person who performed the "upscale" appears: Denis Shiryaev is credited. Additionally, the link to Shiryaev's company is shown. On the right side of the text panel, Shiryaev refers to the origin of the source material processed by him from MoMA ("first link in the description"). Throughout the opening credits, the soundtrack "Deluge" by Cellophane Sam, available free of charge from the freemusicarchive, can be heard. Then the source material from MoMA is shown in black and white as a small film frame in the middle of a black area filling the screen. After a few seconds, the speed of the film is noticeably slowed down. White lettering above the image explains: "Speed correction and 60fps boosting." On the right below the film clip, there is also always the request to subscribe to Shiryaev's channel (Figure 4).

The aesthetic procedure of conveying Shiryaev's algorithmic work on the film footage is reminiscent of similar aesthetic practices in Peter Jackson's THEY SHALL NOT GROW OLD.[51] With the text overlay, the small film frame slowly grows and finally takes up almost the entire window. By upscaling the material, the image material now appears to be richer in detail. The inscriptions that were previously difficult to decipher, such as house lettering or billboards, can now be read more easily. Finally, in one smooth turn of the monorail, the image is dynamically colored. At the next slight turn, ambient sound is added. In this way, aesthetic characteristics of the archival moving images – and, in addition, symbolic images of industrialization – are used to convey the qualities of digital processing in a sensual and immersive way (Figures 4 and 5).

49 Beigang, "Erlaubt ist, was begeistert?" 3–4.

50 Shiryaev, Denis, "[60 fps] *The Flying Train*, Germany, 1902," August 9, 2020, accessed June 24, 2023, https://www.youtube.com/watch?v=EQs5VxNPhzk.

51 Franziska Heller and Ulrich Ruedel, "Das Menschliche Gesicht des Krieges? Archivbilder im digitalen Wandel," *Kinofenster.de. Filmpädagogisches Online-Portal der Bundeszentrale für politische Bildung (bpb)*, June 27, 2019, accessed April 10, 2024, https://www.kinofenster.de/filme/ archiv-film-des-monats/kf1907/kf1907-they-shall-not-grow-old-hg1-archivbilder-im-digitalen-wandel/.

Figure 4: Frame from the beginning of Shiryaev's YouTube video [60 FPS] *THE FLYING TRAIN*, GERMANY, 1902, introducing and highlighting the digital image manipulation showcased. Screenshot by the authors.

Figure 5: Symbolic images of industrialization, the spectacle of colorization, and aesthetic qualities of the moving images merge in a sensual and immersive way in Shiryaev's digitally manipulated THE FLYING TRAIN. Screenshot by the authors.

The spectacle of the algorithmic upgrade is omnipresent, paradoxically even more visible in small glitches. A glitch can be seen on the right video edge as a digital artifact, which Shiryaev himself concedes in a comment: "please excuse me for a glitch on the right side of the video, it is not a problem of neural networks, but my mistake in the final render."

As of May 2023, the video THE FLYING TRAIN by Denis Shiryaev has been viewed around 2.6 million times and commented on over 7,040 times. The majority of the comments in response to the YouTube video are extremely positive and testify to the online community's great enthusiasm.[52] "Less than a minute in and my brain can't handle this. This video looks like something from the future [sic!]. So artistic and incredibly beautiful that I struggle to believe it's real and over 110 years old" (comment by user Gregory Ashton); or "Just wow, I have never had a video move me more than this one" (comment by user Jeremiah Terry). The association with (magical) time travel is also a recurring theme: the immersive quality results in the clip being experienced as a nostalgic gateway to the past.

However, some users also perceive digital imaging as a way of preserving and disseminating history. This could lead to the belief that such appropriations make the historical material more accessible to a broader target group – especially digital natives. This hypothesis at first seems quite logical, particularly against the background of the much lower call-up figures of the "original material" published by MoMA. As of May 2023, it has been viewed, at the time of writing, around 855,000 times and commented on 453 times.

The discussion of the specialized press as well as international scientists could hardly have been more different, however. While the depth of detail of the work was effusively praised in a press release from the city of Wuppertal, and countless online magazines such as *gamestar*, *Wired UK*, and *PetaPixel* suddenly reported on the previously largely unknown digital artist, criticism of the editing of historical material also grew louder.[53] Such digital imaging, it was said, in no way led to increased authenticity, but rather to an alienation that caused the actual historical document to be replaced or at least forgotten.

It needs to be noted that the digital artist Shiryaev does not see his revision as a historical document, but rather locates his work "somewhere between art and entertainment."[54] In doing so, he does not want to question the authenticity or the value of the work, but to create awareness of the existence of the content through

52 Cf. Beigang, "Erlaubt ist, was begeistert?" 11–13.

53 Beigang, "Erlaubt ist, was begeistert?" 13.

54 Thomas Brandstetter, "Historische Aufnahmen. KI bringt die Vergangenheit auf Hochglanz," *Spektrum.de, Spektrum Kompakt, Algorithmen Im Alltag*, May 13, 2021, accessed April 21, 2023, https://www.spektrum.de/news/wie-kuenstliche-intelligenz-historische- filme-koloriert/.

his interpretation. Here, it might be called for to apply a term previously proposed for methods employed in history TV shows, specifically the practice of (electronic) colorization such as in APOCALYPSE in France. This could aptly be called – following approaches such as that of Bohn and Janz with reference to the German TV series WELTENBRAND[55] – a form of media histotainment (as it is aptly phrased, at least in German or French).[56] Could we not accept such re-interpretations as proof of a constantly shifting and dynamic popular image culture?

We maintain that there are serious implications when such a visually creative, algorithmically imprinted media *histotainment* happens within the digital communication culture. Shiryaev must be credited for addressing the nature of the appropriation of the film transparently, indeed making it clear that what we see in terms of colors or faces is not historic, both in his opening credits and in the video description.[57]

But what happens if Shiryaev's version develops its own life out of context within the realm of social media? It will no longer necessarily be recognizable as an interpretation in the form of a new remix. There are already examples where similar things have happened, as vivid debates on Facebook prove – for example, colorized archival images originating from the German TV series WELTENBRAND that were celebrated and circulated as "new" archival material. (Indeed, even the AI portrait of Louis de Funès, referred to above, taken down after a few weeks for license reasons, has already been duplicated on at least two websites since, without context.) Thus, Shiryaev's version may eventually become harvested as source material itself for yet another permutation of these ever-improving, ever-changing moving images.

It is furthermore important to note that the immersive qualities of Shiryaev's clip also contribute to the specific effects. By blending the original qualities of the film with the smooth transitions in the quality jumps and digital "upscaling" (e.g., combined with a route curve within the train ride), the potentials of the AI are sensually (re)experienced. In this way, a utopian, sensual-emotional idea of AI and the potential of algorithms is conveyed through the moving images.

A rhetorical question in place of a definitive conclusion of our discussion is in order – with a grain of salt for historiography. Who would have known about the film THE FLYING TRAIN unless they followed the Museum of Modern Art's YouTube channel? It was Shiryaev who made the material more widely known through more effectively serving the platform's affective attention logic – as seen in the

55 Anna Bohn and Oliver Janz, "Kann denn Farbe Sünde sein?" *Der Tagesspiegel*, September 20, 2012, accessed January 31, 2013, http://www.tagesspiegel.de/medien/alte-bilder-neues-sehen-kann-denn-farbe-suende-sein/7155174.html. Cf. Heller, *Update!*, 77–80.

56 Cf. Beigang, "Erlaubt ist, was begeistert?" 16–17.

57 Cf. complete quote by Shiryaev in n. 45.

great difference in the access numbers for the respective videos. Shiryaev, whether deliberately or unconsciously, creates awareness in a section of the internet public that might not consult original archive collections or might have less of an emotional or even intellectual access to historical film material. But what is the focus of the response and the awareness? Is it indeed, as audience members state and appear to believe, the excitement of "lifelike" images of historical periods and settings, or is it the very spectacle of knowing that the latest digital tools made these images possible? The snowballing of sharing them normalizes the modified images as an acceptable practice of engaging with "historic" images. The effective replacement of source images by their modified versions within digital circulation by the sheer power of numbers may actually render these modifications close to irreversible as the dominating expectation of visual historicity.

There are many similar examples on the internet. Through the proliferation of cheap image modification "enhancement" and "restoration" tools, the average internet *prosumer* can become a DIY "restorer." In releasing the results of their work into the digital social sphere, they consciously or unconsciously speak to the logics and economy of attention of digital social platforms where the technological imaginary of the spectacle of AI bears at least as much relevance as the images it has produced and, in doing so, renders the original source of the images more and more irrelevant. As if polemically echoing the statement of director John Milus that "everybody is a filmmaker today" through newly available technical means, it appears that this very availability of digital tools compels everyone to think of themselves as film restorers. Using these with little if any thought, merely conveniently relying on automated decision processes and free source material at hand, online prosumers modify historic imagery without any of the solid ethical, theoretical, material, and cultural training or historical and visual sensibility of a professionally trained film restorer. Instead of the critical historical consideration and pondering required in source criticism and visual evaluation, decision making is delegated to codecs. Critical study and choice of source material is replaced by the convenience of sources already circulating online – the images become a commodity as an algorithmic showcase for data potency, reinforcing a narrative of technological supremacy. In consequence, by sheer force of number, the proliferating, attention-grabbing images become a mere playground to showcase digital imaging aptness and coding, Darwinistically displacing and replacing the original historic imagery. Here we observe "digitization" as a cultural narrative in all its imaginary and real facets. This proceeds at the expense of the hermeneutic interpretation and reflection of traces and documents of the past. The belief in codes and algorithms threatens to shape our visual history.

Preliminary Conclusion: Towards a Visual Literacy in the Digital Sphere – Proposing Key Competences

In a culture of media that is changing through sharing, digital media scholars need to develop an understanding of what they are seeing with any given piece of material. Much beyond known questions of remastering, restoring, re-restoring, re-grading, the democratization of moving image manipulation and sharing has led to an avalanche of versions that threaten to basically bury the most "authentic"/"original" surviving one by the sheer force of numbers. Of course, such changes are not necessarily a bad thing per se – the very point of preserving and sharing access to moving image heritage is to allow engagement with it, and that may well include modifying it, re-using it, re-mixing it – or creating a meme out of it.

To propose a methodological answer, we take inspiration from Andreas Fickers' concept of "digital hermeneutics" as a methodological update in investigating "Ontological changes from 'source' to 'document' to 'data'."[58] In this, Fickers focuses on the specific hybridity of a situation comprising the availability of both analog sources and digitized texts and media in which current historians operate. We propose that in the exceedingly popular medium of moving images, however, we also have to consider everyday digital communication culture. Increasingly, the digital sphere not only becomes communication context, but also the moving image culture's very surrogate, with seductively easy, digitally accessible (and malleable) source material. Thus, it is crucial to also consider this phenomenon's effect on the material at hand and the discourse about and around it. As history is increasingly (re-)written in the virtual digital realm, it is all the more important to understand the material roots of our cultural-visual memory, regardless of the reproducible nature of moving images from the introduction of the negative-positive process in the mid-nineteenth century to modern multiple digital duplication.

We strongly advocate, along the lines proposed by Fickers, that the same kind of critical approach established towards analog sources needs to be applied to their digital derivatives, the analog-to-digital transition, and subsequent digital modification. As we have outlined, however, the latter processes are subject to technological, sociocultural, and political selection processes, which manifest themselves predominantly and frequently in further deliberate modification and/or generational degradation in the digital (communication) realm. This is inherent in the platform operation and mechanisms both in technical logistics and the

58 Andreas Fickers, "Update für die Hermeneutik. Geschichtswissenschaft auf dem Weg zur digitalen Forensik?" *Zeithistorische Forschungen/Studies in Contemporary History* 17, no. 1 (2020): 157–168, https://doi.org/10.14765/ZZF.DOK-1765.

logic of attention economy. Thus, we propose turning the interdisciplinary dialogue between film restoration and digital culture media studies into a transdisciplinary methodology, not only as part of our academic practice, but beyond.

Towards the realization of the truly transdisciplinary approach initiated with the above deliberations and case studies, we now propose some specific fields of further research towards a specific visual literacy for historical images and their digital derivatives. These could or should include the following:[59]

- The critical evaluation of *image texture and detail*: recognizing "natural" photographic/film grain (which is "natural" for analog photographic images, however not natural for the photographic subject) versus digital smoothness, detecting digital image artifacts, such as those caused by compression or quantization, blockiness, and issues such as shadow crushing and highlight clipping.
- Developing awareness for *perceiving and understanding motion appearance* both inherent in analog historic moving images and/or imposed upon by digital manipulation: "filmic" limited frame rate motion (like grain, another "limitation" of analog techniques that has turned into an embraced characteristic) versus artificial higher frame rate and/or interpolated motion.
- Awareness, visual literacy and arguably connoisseurship in evaluating and distinguishing *color appearance* between and in between different historical systems of photographically capturing and reproducing colors, and historic or indeed modern digital techniques artificially adding color to monochrome images (and the difference of all of the above to the naturally perceived appearance of objects and scenes themselves).
- A thorough understanding of the *image source/genealogy* (including the digital data and derivative versions in the online realm!), which is not only relevant for the technical and visual sense but also in terms of pragmatic communicative framing and the genealogy the latter imposes on the former.
- Critical consideration of the *discursive framing rhetorics* in comment sections, etc. – often intertwined with visual fetishism and the building of a utopian technological imaginary towards an idea of "the digital," or recently AI and machine learning in particular – regarding the spectacle of "enhanced image quality" or showcasing of colorization processes. This also needs to be evaluated in view of the *modes of communication* that might focus on affective

59 This list bears some resemblance to the titles of William J. Mitchell's chapter "Intention and Artifice" in his seminal book *The Reconfigured Eye*, 23–57. It is interesting that Mitchell's book, which deals with digitization on photography as early as the 1990s, can still be adapted to today's discussion.

(iconic) forms of expressions (such as emojis) or other responses and emotionalizing content.
– Questioning the *historiographic impact* of sensation logics turning historical sources into technological showcases aimed at fostering a new participatory (yet mostly affirmative!) and emotional engagement with the images.

References

Acland, Charles R. "Tampering with the Inventory: Colorization and Popular Histories." *Wide Angle* 12, no. 2 (1990): 12–20.

Beigang, Berenike. "Erlaubt ist, was begeistert? Medien- und Technikgeschichte(n) am Beispiel des Filmes *The Flying Train* aus dem Jahre 1902 sowie seiner Bearbeitung durch Denis Shiryaev im Jahr 2020." Unpublished student paper, Martin Luther University Halle-Wittenberg, 2022.

Benjamin, Walter. *Das Kunstwerk im Zeitalter seiner technischen Reproduzierbarkeit* [1936]. Frankfurt am Main: Reclam, 1977.

Böhme, Hartmut. *Kulturgeschichte der Technik*, 2000. Accessed June 24, 2023. https://www.hartmut boehme.de/static/archiv/volltexte/texte/kgdt.html.

Bohn, Anna, and Oliver Janz. "Kann denn Farbe Sünde sein?" *Der Tagesspiegel*, September 20, 2012. Accessed January 31, 2013. http://www.tagesspiegel.de/medien/alte-bilder-neues-sehen-kann-denn-farbe-suende-sein/7155174.html.

Brandstetter, Thomas. "Historische Aufnahmen. KI bringt die Vergangenheit auf Hochglanz." *Spektrum.de. Spektrum Kompakt, Algorithmen im Alltag*, May 13, 2021. Accessed April 21, 2023. https://www.spektrum.de/news/wie-kuenstliche-intelligenz-historische-filme-koloriert/.

Bucher, Taina, and Anne Helmond. "The Affordances of Social Media Platforms." In *The SAGE Handbook of Social Media*, edited by Jean Burgess, Alice E. Marwick, and Thomas Poell, 233–253. London: Sage Publications, 2018.

Byrne, Robert, Caroline Fournier, Anne Gant, and Ulrich Ruedel. "The Digital Statement III. Image Restoration, Manipulation, Treatment, and Ethics." *Journal of Film Preservation* 104 (2021): 25–37.

Distelmeyer, Jan. *Das flexible Kino. Ästhetik und Dispositiv der DVD & Blu-ray*. Berlin: Bertz und Fischer, 2012.

Fickers, Andreas. "Update für die Hermeneutik. Geschichtswissenschaft auf dem Weg zur digitalen Forensik?" *Zeithistorische Forschungen/Studies in Contemporary History* 17, no. 1 (2020): 157–168. https://doi.org/10.14765/ZZF.DOK-1765.

Flückiger, Barbara. "Material Properties of Historical Film in the Digital Age." *NECSUS. European Journal of Media Studies* 1, no. 2 (2012): 135–153. https://doi.org/10.5117/NECSUS2012.2.FLUE.

Flückiger, Barbara. *Visual Effects: Filmbilder aus dem Computer*. Marburg: Schüren Verlag, 2008.

Flückiger, Barbara, Franziska Heller, Claudy Op den Kamp, and David Pfluger. "'Digital Desmet': Translating Early Applied Colors." *The Moving Image* 16, no. 1 (2016): 106–124.

Fossati, Giovanna. *From Grain to Pixel: The Archival Life of Film in Transition* [2009], 3rd rev. ed. Amsterdam: Amsterdam University Press, 2018.

Grant, Catherine. "The Shudder of a Cinephiliac Idea? Videographic Film Studies Practice as Material Thinking." *Aniki* 1, no. 1 (2014): 49–62. https://doi.org/10.1515/9789048543922-012.

Gunkel, Katja. *Der Instagram-Effekt*. Bielefeld: Transcript, 2018.

Hagener, Malte. “Cinephilia and Film Culture in the Age of Digital Networks.” In *The State of Post-Cinema: Tracing the Moving Image in the Age of Digital Dissemination*, edited by Malte Hagener, Vinzenz Hediger, and Alena Strohmaier, 181–194. London: Palgrave Macmillan, 2016.

Heller, Franziska. *Update! Film- und Mediengeschichte im Zeitalter der digitalen Reproduzierbarkeit*. Paderborn, Munich: Brill, Wilhelm Fink, 2020. https://doi.org/10.30965/9783846764602.

Heller, Franziska, and Ulrich Ruedel. “Das Menschliche Gesicht des Krieges? Archivbilder Im digitalen Wandel.” *Kinofenster.de. Filmpädagogisches Online-Portal Der Bundeszentrale Für Politische Bildung (Bpb)*, June 27, 2019. Accessed April 10, 2024. https://www.kinofenster.de/filme/archiv-film-des-monats/kf1907/kf1907-they-shall-not-grow-old-hg1-archivbilder-im-digitalen-wandel/.

Jameson, Fredric. *Postmodernism, or, the Cultural Logic of Late Capitalism*. Durham, NC; London: Duke University Press, 1991.

Jenkins, Henry. *Fans, Bloggers, and Gamers: Exploring Participatory Culture*. New York: New York University Press, 2006.

Kaminski, Michael. “Saving Star Wars: The Special Edition Restoration Process and Its Changing Physicality.” *The Secret History of Star Wars*, 2009. Accessed June 29, 2023. http://fd.noneinc.com/secrethistoryofstarwarscom/secrethistoryofstarwars.com/savingstarwars.html.

Klawans, Stuart. “Colorization: Rose-Tinted Spectacles.” In *Seeing Through Movies*, edited by Mark Crispin Miller, 150–185. New York: Pantheon, 1990.

Klinger, Barbara. *Beyond the Multiplex: Cinema, New Technologies, and the Home*. Berkeley, Los Angeles: University of California Press, 2006.

MacQueen, Scott. “Mystery of the Wax Museum: Restoring Two-Color Technicolor with Digital Tools.” *Journal of Film Preservation* 103 (2020): 104–114.

MacQueen, Scott. “Technicolor and the True Believer.” *The Perfect Vision* 3, no. 12 (1991–1992): 23–38.

Manovich, Lev. *The Language of New Media*. Cambridge, MA: MIT Press, 2001.

Mitchell, William J. *The Reconfigured Eye: Visual Truth in the Post-Photographic Era*. Cambridge, MA: MIT Press, 1992.

Murray, Janet. *Inventing the Medium: Principles of Interaction Design as a Cultural Practice*. Cambridge, MA: MIT Press, 2011.

Noordegraaf, Julia. “Zooming Out: Towards Scalable Digital Film Studies.” Presentation and discussion at the workshop Teaching Digital Methods for Film Historiography, Johannes Gutenberg University Mainz, Germany, within the DFG-Network New Directions in Film Historiography, February 18, 2021.

Noordegraaf, Julia, Kathleen Lotze, and Jaap Boter. “Writing Cinema Histories with Digital Databases: The Case of Cinema Context.” *TMG Journal for Media History* 21, no. 2 (2018): 106–126. https://doi.org/10.18146/2213-7653.2018.369.

Odin, Roger. “Dokumentarischer Film – Dokumentarisierende Lektüre [1984].” In *Sprung im Spiegel*, edited by Christa Blümlinger, 125–146. Vienna: Sonderzahl, 1990.

Op den Kamp, Claudy. *The Greatest Films Never Seen: The Film Archive and the Copyright Smokescreen*. Amsterdam: Amsterdam University Press, 2018.

Op den Kamp, Claudy, Barbara Flückiger, and David Pfluger. “A Material-Based Approach to the Digitization of Early Applied Colors.” In *The Colour Fantastic: Chromatic Worlds of Silent Cinema*, edited by Giovanna Fossati, Victoria Jackson, Bregt Lameris, Elif Rongen-Kaynakçi, Sarah Street, and Joshua Yumibe, 237–259. Amsterdam: Amsterdam University Press, 2018.

Price, Nicholas Stanley, M. Kirby, Talley Jr., and Alessandra Melucco Vaccaro, eds. *Historical and Philosophical Issues in the Conservation of Cultural Heritage*. Los Angeles: Getty Conservation Institute, 1996.

Snickars, Pelle, and Patrick Vonderau, eds. *The YouTube Reader*. Stockholm: National Library of Sweden, 2009.

Sorlin, Pierre. "Ist es möglich, eine Geschichte des Kinos zu schreiben?" *Montage AV* 5, no. 1 (2009): 23–37.

Thurlow, Crispin, Christa Dürscheid, and Federica Diémoz, eds. *Visualizing Digital Discourse: Interactional, Institutional and Ideological Perspectives*. Berlin, Boston: De Gruyter Mouton, 2020. https://doi.org/10.1515/9781501510113.

Wallmüller, Julia. "A Classical Approach to Authentically Restoring Film." *Aesthetic Investigations* 2, no. 2 (2019): 144–162.

Wallmüller, Julia. "Criteria for the Use of Digital Technology in Moving Image Restoration." *The Moving Image* 7, no. 1 (2007): 78–91.

Zoller, Charles. C. "Charlie Chaplin as The Tramp." *Eastman Museum*, ca. 1917–1918. Accessed June 24, 2023. https://collections.eastman.org/objects/189079/charlie-chaplin-as-the-tramp.

Filmography

[60 FPS] *THE FLYING TRAIN*, GERMANY, 1902. Denis Shiryaev. Published August 9, 2020. Accessed June 24, 2023. https://www.youtube.com/watch?v=EQs5VxNPhzk.

[ACTUAL 4K SCAN] *THE ARRIVAL OF A TRAIN AT LA CIOTAT STATION* – LUMIÈRE BROTHERS – 1896. Andy Meyers. Published February 5, 2020. Accessed June 24, 2023. https://www.youtube.com/watch?v=1FAj9fJQRZA.

LE CINÉMATOGRAPHE 1895–1900 LUMIÈRE. France Télévisions Distribution. Blu-ray ed. France, 2016.

THE LUMIÈRE BROTHERS' COLLECTION, 1895–1902. Denis Shiryaev. Published January 3, 2021. Accessed June 24, 2023. https://www.youtube.com/watch?v=YZuP41ALx_Q.

THEY SHALL NOT GROW OLD, dir. Peter Jackson. Warner Bros. Blu-ray ed. United Kingdom, 2018.

Josephine Diecke and Malte Hagener

Managing Tools and Expectations: Dos and Don'ts of Teaching Digital Methods for Film Analysis and Film Historiography

Introduction

Within the realm of film studies, the assimilation of digital methods into pedagogical practices has, until recently, charted a relatively unexplored territory, both in its theoretical and practical dimensions. However, a distinct scarcity exists when it comes to published accounts of specific experiences within this context.[1] In contrast to disciplines like literary studies, linguistics, history, and art history, where digital humanities have gained traction within university curricula,[2] the incorporation of digital tools and methodologies into film historiography instruction has been characterized by slow progress. Beyond the lack of availability, there is also a paucity of discourse among educators sharing their experiences in this domain. This chapter aims to bridge this evident void by presenting a comprehensive case study: the teaching of Digital Methods in Film and Media Studies, a course carefully designed to introduce Master's students in Media Studies at the University of Marburg to the intricacies of digital film analysis and historiography.

1 Among the few written discussions are selective accounts of experiences in blog post format, such as Jan Teurlings and Markus Stauff's report on teaching with the tool Perusall (see Jan Teurlings and Markus Stauff, "Teaching with Perusall," *Open Media Studies Blog*, 2022, accessed August 8, 2023, https://mediastudies.hypotheses.org/3086). See also Nicole Braida and Frauke Pirk's chapter in this volume, "Teaching Small-Gauge Formats with Digital Methods."

2 See, for instance, some of the most recent publications in these fields, such as Andreas Fickers and Juliane Tatarinov, eds., *Digital History and Hermeneutics: Between Theory and Practice* (Berlin: De Gruyter Oldenbourg, 2022), https://doi.org/10.1515/9783110723991; Karoline Dominika Döring et al., *Digital History: Konzepte, Methoden und Kritiken Digitaler Geschichtswissenschaft* (Berlin: De Gruyter Oldenbourg, 2022); Lina Franken, *Digitale Methoden für qualitative Forschung. Computationelle Daten und Verfahren* (Stuttgart: UTB, 2022); Johanna Drucker, *The Digital Humanities Coursebook: An Introduction to Digital Methods for Research and Scholarship*, 1st ed. (Abingdon, Oxon; New York: Routledge/Taylor & Francis, 2021); Kathryn Brown, *The Routledge Companion to Digital Humanities and Art History* (New York: Routledge, 2020); Christopher J. Young et al., *Quick Hits for Teaching with Digital Humanities: Successful Strategies from Award-Winning Teachers* (Bloomington: Indiana University Press, 2020).

https://doi.org/10.1515/9783111082486-014

Within the evolving terrain of digital film historiography, a paradigm emerges that is both "experimental" and "exploratory." These descriptors encapsulate the inherent openness and dynamic processuality that underpin the utilization of digital tools and methodologies in addressing questions and datasets within the humanities, including those within film studies. As summed up in Sönke Ahrens' propositions in *Experiment und Exploration* (2010), technology not only unveils novel avenues for exploration but also presents a realm of exploration and playful experimentation in its own right. Ahrens contends that the true potential of technology is perennially unexplored because each exploration leads to new experimental forms, which in turn enlighten further exploration.[3] It functions as both a catalyst, propelling the exploratory to its universal boundaries, and a well of continuous possibilities that fosters exploration, whether internally or as a mediator.[4] This dichotomy, in turn, informs our approach as we navigate the dynamic terrain of digital methodology integration, drawing on Ahrens' conceptual framework as a guiding light. Ahrens' perspective, which recognizes education as a platform for experimentation and knowledge generation, harmonizes with our mission to cultivate an interactive and dynamic learning environment.

Another lens through which to view the dynamic essence of employing digital tools and methodologies for humanities inquiry is to consider the processes as "flexible" and "iterative."[5] These descriptors shift the focus of investigation from linear, step-by-step analyses of questions and datasets to a multi-step approach that involves recurrently customized search queries and approximations during data preprocessing.[6] In our endeavor to offer a seminar on digital methods in film research, we aimed to integrate these time-consuming and labor-intensive steps into the research process itself, rather than treating them as mere peripheral components. Thus, our course sought to cultivate a sense of iteration, experimentation, and exploration for the students, empowering them to venture into unexplored territories in the domain of digital film research.

3 See Sönke Ahrens, *Experiment und Exploration: Bildung als experimentelle Form der Welterschließung* (Bielefeld: Transcript, 2010), 287.

4 Ahrens, *Experiment und Exploration*, 287–288.

5 See Charles R. Acland, Eric Hoyt, and Kit Hughes, "A Guide to the Arclight Guidebook," in *The Arclight Guidebook to Media History and the Digital Humanities*, ed. Charles R. Acland and Eric Hoyt (Falmer: REFRAME Books, 2016), 16; Andreas Fickers, Juliane Tatarinov, and Tim van der Heijden, "Digital History and Hermeneutics. Between Theory and Practice: An Introduction," in *Digital History and Hermeneutics: Between Theory and Practice*, ed. Andreas Fickers and Juliane Tatarinov (Berlin: De Gruyter Oldenbourg, 2022), 8.

6 See Acland, Hoyt, and Hughes, "A Guide to the Arclight Guidebook," 16.

This chapter delves into the course's structure, preliminary considerations, and noteworthy insights that have left a lasting imprint.[7] Through these explorations, we aim to underscore the significance and timeliness of digital methods' application and usage as both tools and subjects of inquiry in film studies. Teaching digital film historiography presents a distinct set of challenges that, in turn, refine and reframe broader questions within this emergent field. Our reflections on teaching transcend mere didactic considerations, delving into the very essence of the subject. In this context, the process of teaching digital methods parallels the research journey itself, mirroring the research process to a certain extent – particularly when one does not rely solely on established facts or meta-debates. Aligning with Ahrens' concept of education as an experimental avenue for exploring the world, the classroom transforms into both a laboratory and experimental domain.[8] Here, a new generation of practitioners is initiated, while it also serves as a testing ground for novel approaches and methodologies. Our core objective was to allow students to wield digital tools and methods as epistemological instruments "in the wild." This approach aimed to leverage students' existing knowledge of film studies, while also assessing their prowess in utilizing digital tools to access information and collaborate – skills associated with the "digital natives" generation. However, as our journey unfolded, it became evident that the latter assumption held only partial truth.

The Class – Introduction

In the summer semester of 2021, when the COVID-19 pandemic made in-person teaching very difficult still, we taught an online seminar entitled "Digital Methods in Film Research: Perspectives and Applications." The course format can be described as a project seminar with a workload of 12 credit points. It was attended by students in the second semester of the Master's program *Medien und kulturelle Praxis* (Media and Cultural Practice). In total, we worked with a group of 12 students who explored selected approaches and tools in the field of Digital Humanities between mid-April and mid-July 2021. Following the teaching period and over

7 This seminar was taught in the framework of the Volkswagen Foundation project "Digital Cinema-Hub (DiCi-Hub): A Research Hub for Digital Film Studies" (2021–2026), with the aim to develop and test new conceptual frameworks and tools for film-related research and teaching. For more on the project, see https://www.uni-marburg.de/de/fb09/medienwissenschaft/forschung/forschungsprojekte/dici-hub, accessed March 15, 2024.

8 See Ahrens, *Experiment und Exploration*, 301.

the course of the summer, the students had time to prepare portfolios which formed the basis for grading.

The class started with a general introduction to Digital Humanities and Film Studies built around a small selection of texts. The choice of readings was motivated by the goal of giving the students an idea of the potential of digital methods, while also making them aware of the limitations that these methods might entail. The texts that were discussed in more detail in class were Franco Moretti's "Planet Hollywood,"[9] Julian Sittel's "Digital Humanities in der Filmwissenschaft"[10] (a recent overview of approaches and tools in German), and the introductory essay from *The Arclight Guidebook to Media History and Digital Humanities*, written by Eric Hoyt, Kit Hughes, and Charles Acland.[11] The focus of the discussions was on the epistemological potential of digital methods and for which questions such methods and tools can be reasonably used. Thus, our teaching was not concerned with circumscribing or defining Digital Humanities as a field; this kind of meta-discussion rarely makes much sense in teaching, at least not for students at the Bachelor or Master level.[12]

Furthermore, we presented and explored a small group of digital tools and platforms such as Project Arclight,[13] Voyant,[14] VIAN,[15] the Timeline of Historical Film Colors,[16] and the German-language media repository media/rep/.[17] These tools can all be used without any coding skills, specific training, or prior experience. Since the idea was a hands-on course for students who did not necessarily have experience in the concrete implementation of digital methods, this was intended to give them a quick overview and easy access to the field. It also provided the groundwork for discussing critically such concepts as the quality of data and the accuracy of the methods and tools, as well as the meaningfulness of the results. These aspects became important later, when the groups prepared and executed their own projects.

9 Franco Moretti, "Planet Hollywood," *New Left Review* 9 (2001), accessed December 11, 2023, https://newleftreview.org/issues/ii9/articles/franco-moretti-planet-hollywood.

10 Julian Sittel, "Digital Humanities in der Filmwissenschaft," *MEDIENwissenschaft: Rezensionen / Reviews* 4 (2017): 472–489.

11 Acland, Hoyt, and Hughes, "A Guide to the Arclight Guidebook."

12 Ryan Cordell, "How not to Teach Digital Humanities," in *Debates in the Digital Humanities*, ed. Matthew K. Gold and Lauren F. Klein (Minneapolis, MN: University of Minnesota Press, 2016): 460–461.

13 See https://search.projectarclight.org, accessed March 27, 2023.

14 See https://voyant-tools.org/, accessed March 27, 2023.

15 See https://www.vian.app, accessed October 21, 2024.

16 Barbara Flückiger, "Timeline of Historical Film Colors," accessed March 27, 2023, https://filmcolors.org.

17 See https://mediarep.org, accessed March 27, 2023.

The second and biggest part of the seminar was dedicated to the design, conception, and execution of the three group projects. The twelve students split up into three groups with separate research objects and questions. In fact, finding suitable projects turned out to be quite difficult and required substantial time and energy, but it was intended to be part of the exercise. Just as in "real" research, many steps were necessary, and finding the right question, dataset, and tools were an integral part of the research. After a longer period of collecting and discussing ideas, the groups settled on three topics.

The first group ("Weimar cinema") concentrated on five quintessential "horror films" from the German cinema of the 1920s, which belong to the canon of expressionist classics, such as Das Cabinet des Dr. Caligari (Robert Wiene, GER 1919/1920), Schloß Vogelöd (Friedrich Wilhelm Murnau, GER 1921), Nosferatu – eine Symphonie des Grauens (Friedrich Wilhelm Murnau, GER 1921/1922), Orlac's Hände (Robert Wiene, AUT 1924), and Faust – eine deutsche Volkssage (Friedrich Wilhelm Murnau, GER 1925/1926).

The second group ("Dudow") investigated the three-color films Frauenschicksale (GDR 1951/1952), Der Hauptmann von Köln (GDR 1956), and Verwirrung der Liebe (GDR 1958/1959) from the Bulgarian-born director Slatan Dudow, which he made in the 1950s for the East German state-owned film production company Deutsche Film AG (DEFA).

The third group ("Tatort") analyzed three episodes of the German crime television series Tatort: Für immer und dich (Julia von Heinz, GER 2019), Der letzte Schrey (Mira Thiel, GER 2020), and Der Herr des Waldes (Christian Theede, GER 2021).

When thinking about doing digital film history, one can distinguish (in a somewhat schematic or simplistic manner) between three basic approaches: one can start off with a tool and then look for an implementation; one can start with a circumscribed dataset, which one then explores and examines; or one can start with a specific question that then requires digital methods to be answered. We had made no prior specification as to whether the project was to be tool-driven, data-driven, or question-driven, which might have contributed to the difficulty of finding topics. If all three parameters are open, the field is so vast that it might overburden students. Yet again, what kind of closures do we want to set as preconditions? If we give them a certain dataset that is well-prepared, we might give a wrong impression of how data-driven research works in our field. In our case, the solution to this challenging conundrum was to make the pros and cons of each approach a decisive part of the weekly group discussions as well as the submitted portfolios. For future courses, however, we would like to address these considerations earlier and include them at the introductory session level. One could then discuss the difficulties of obtaining a dataset, of making a tool work,

or of building a data model for falsifying a thesis at the very beginning, instead of confronting the students with these issues when they are just about to begin their research. Nevertheless, learning by doing (i.e., "the hard way") how complicated and unwieldy such undertakings are provided an important lesson to learn.

The Class – Process

Since the students were (initially) mostly interested in working on the images of the film itself, not on meta-data of any kind, the digital tool that we started working with for the group projects was the video annotation software VIAN.[18] For this purpose, we invited Jonathan Schaber, part of the DevOps-team at the University of Zurich, to give an introduction to the tool's features.[19] However, working with a powerful tool that is still in development had its pitfalls. Not everyone was able to run the software on their personal computers, some of which were already a couple of years old. Even with extensive trouble shooting, only half of the students could properly install and run the software because their computers were not suitable for the technical requirements of VIAN. Even though the students all had smart phones and many of them also had tablets, they often owned no traditional computer or had only older laptops that they had inherited from friends and family – and a lack of computing power and non-compatible operating systems did not allow for a proper installation. As a result, the members of each group were eventually divided between working with VIAN and working with Voyant. Therefore, each group ended up combining video analysis with textual analysis, which we considered a reasonable procedure because mixed methods approaches are after all quite typical of Digital Humanities. The research for historical material (e.g., film reviews and other texts about films) relied mostly on the collections of the Media History Digital Library and on articles kindly provided by the DEFA Foundation upon request.

Whereas we are convinced that mixed-methods approaches are the most fruitful way to engage with the digital, it does make the research design more complicated. The question of how to relate textual analysis to visual analysis re-

18 For more information about VIAN, see Barbara Flückiger and Gaudenz Halter, "Methods and Advanced Tools for the Analysis of Film Colors in Digital Humanities," *Digital Humanities Quarterly* 14, no. 4 (2020), accessed March 27, 2023, http://www.digitalhumanities.org/dhq/vol/14/4/000500/000500.html.

19 This is one of the big advantages of online teaching – the ease with which one can integrate colleagues from elsewhere into a seminar.

mained a stubborn point of discussion and resistance throughout because of the combination of multiple methods and data formats (annotation data, unstructured written language, images, etc.). Yet again, since this is also a key question in "real" research, it is well worth having these discussions, even when they are taking time away from other tasks on the project.

The features of the individual tools included (1) digital film analysis with VIAN and the VIAN WebApp with tools for video annotation, classification, and visualization; (2) text analysis and visualization with Voyant, which detects and visualizes, for instance, word frequencies, trends, and clusters; and (3) the full text search in the collections of the North-American repository Media History Digital Library in combination with Arclight's search and graph visualization.

In addition to applying the digital tools and methods, we gave the students the task of discussing and documenting their progress within the groups. For this purpose, they used file sharing and collaborative writing platforms provided by the university (Hessenbox, OnlyOffice), as well as by commercial providers (Google Drive and Dropbox). Texts and film files were also made available via Hessenbox and the University of Marburg's own teaching and learning platform Ilias. For the purpose of regular communication and our weekly meetings, we worked with the video conference system Big Blue Button, sometimes with all students in the same virtual room and sometimes with smaller groups in breakout rooms. In fact, the digital ecosystem that had been implemented and tested by that time allowed for a welcome degree of flexibility and openness – we did not need to announce rooms, work assignments, and tasks at the beginning of the semester for each session, and we were able to react quickly to problems and developments. Therefore, each week looked somewhat different and we could use the seminar slot either to update everyone on the progress of the project or to have group sessions with divided tasks. A final presentation by all groups to everyone in the course concluded the seminar phase in the last week of teaching. For the final assignment, each person submitted a portfolio of 60 to 80 pages documenting their work, two thirds to be produced by the entire group and one third individually, including a reflection on the whole process. In the end, the group work and the individual work were graded in equal parts. An integral component of their task involved engaging in a critical self-assessment of the time allocated to data (pre)processing and the subsequent adjustments they made to the parameters of their group projects.

Regarding the actual seminar sessions, the focus was on the presentation, description, and critical discussion of the digital tools and methods used, as well as on the presentation of the research question and the source materials. Reflections of the different roadmaps and timelines of the three groups could be seen in the portfolios, as could individual challenges and epistemological trajectories. Last

but not least, selected results and approaches with regard to the specific topics were also included in the final portfolios.

The "Dudow" group used Voyant for an examination of the most common terms that appeared in contemporary reviews of the three-color films. The "Weimar Cinema" group analyzed pre-existing visualizations from the VIAN WebApp for the color scheme of their individual corpus films, such as the digitally restored tinted and toned version of Das Cabinet des Dr. Caligari. The "Tatort" group segmented their selected three episodes according to recurring narrative patterns, and subsequently visualized their occurrence over time in the Mac-tool Numbers. These examples show how the functionality of tools can be combined with existing data pools in order to answer research questions.

Especially when comparing the individual portfolios, similar and divergent applications of the tools became visible, underlining the scientific and creative impulses of the individual group participants. Nevertheless, in the final reflections, the assessments of all twelve course participants coincided to a large extent. They stated that they had approached the digital tools and methods with high expectations but were only able to implement them to a small extent. In particular, they had expected more from the quantitative and automated approaches in VIAN and Voyant but were not able to give them sufficient attention because of the long and laborious process of research design and data preparation. This resulted, at least for some, in a fundamental critique of the technical challenges that especially VIAN presented. As mentioned, only half of the students were even able to get the program to run continuously, so one student brought up the question of a potential structural discrimination inherent in such tools. Moreover, some problems, such as dealing with installation problems, were difficult to assess via distance learning with Big Blue Button. Here, the flexibility of the digital environment became a hindrance because we were not able to gather around a laptop and work together on the troubleshooting. These circumstances contributed to the fact that the contents of the portfolios mainly referred to process descriptions and experience reports, and only to a smaller extent to results and further research questions. For example, most portfolios concluded that further work with a larger corpus was needed in order to confirm or falsify the preliminary results.

Lessons Learned and Results

In their approach to "Digital Hermeneutics," Andreas Fickers and Tim van der Heijden refer to a special form of exploratory research and teaching as *thinkering*.[20] Combining "thinking" and "tinkering," their methodological approach to digital tools is to be understood as "the tinkering with technology combined with the critical reflection on the practice of doing digital history."[21] It is precisely the aspect of playfulness that would stimulate creative thinking and lure students and teachers out of their "comfort zone," as Fickers explains.[22] Part of this process is the necessary acquisition of new theoretical and technical competencies, often referred to as data and tool literacy, that also involve collaboration and reflection for a "shared project,"[23] as well as one's own experimentation. Since such projects are active processes of negotiation, the interdisciplinary collaboration between computational sciences and the humanities is also described as a "trading zone" between both parties, in which (in the best case) analog and digital processes continue to coexist in a hybrid relationship.[24]

Following this call for *thinkering* with tools and data, we as instructors also left our comfort zones, and want to share five observations and lessons learned from teaching digital methods for students in film and media programs. We assume that the vast majority of such students have a basic understanding of computers and their functioning, but no particular skills in coding or software applications designed for specific purposes.[25]

First of all, if you start a course without knowing all the steps to come, but you want to give creative leeway to students nevertheless, the content and results will usually turn out to be more process-oriented and less result-oriented, especially if the students have no previous experience with the methods and tools used. Our advice would be to either adjust your own and your student's expectations from the beginning or to change the course program according to the com-

20 Andreas Fickers and Tim van der Heijden, "Inside the Trading Zone: Thinkering in a Digital History Lab," *Digital Humanities Quarterly* 13, no. 3 (2020), accessed March 27, 2023, http://www.digitalhumanities.org/dhq/vol/14/3/000472/000472.html.

21 Fickers and Van der Heijden, "Inside the Trading Zone."

22 Andreas Fickers, "Forum: 'Mehr Mut zum Experimentieren' – Digitale Lehre – Interview mit Andreas Fickers (Universität Luxembourg)," *H-Soz-Kult*, April 10, 2020, accessed March 27, 2023, http://www.hsozkult.de/debate/id/diskussionen-4964.

23 Acland, Hoyt, and Hughes, "A Guide to the Arclight Guidebook," 3.

24 Fickers, Tatarinov, and Van der Heijden, "Digital History and Hermeneutics," 6–7.

25 This caveat is meant as a reminder that we are discussing here a very specific situated case that should not be too easily generalized.

petences of your class. If the dataset(s) and/or the tool(s) are predetermined, one saves a lot of time and energy that is otherwise spent on exploratory steps.

Secondly, it is even more important to define the technical and methodological requirements as precisely as possible before the course begins, for example, by preparing the appropriate ramifications such as datasets and software. This involves investing time in handling technical foundations and difficulties, or looking for technical support if you do not have these skills yourself. Also, make sure that dealing with storage space and ways of file sharing do not occupy a lot of precious work time. Working with high resolution video files is not just a challenge for the process of video annotation itself, but also for people who are used to outsourcing data to cloud services which – this was at least our impression – many students are.

Thirdly, encourage experimentation and playfulness, but do not underestimate the importance of defining boundaries for text, data and video analysis. As a result of our initial reluctance to clearly delineate the research question, method, and tools, our students encountered challenges not only in their individual application but also in their integration and subsequent analysis. In a best case scenario, this approach will foster an environment of newly found excitement for traditional questions and approaches of film studies. But there is a high risk that students will get caught up in marginal tasks or remote problems, thereby losing sight of the bigger picture.

This being said, combine – this is our fourth point – qualitative and quantitative methods, respectively methods of close reading and distant reading, so that the students experience how empirical statements can support or dismantle small-scale hypothesis. Illustrating this concept is the amalgamation of individual video annotations for a select group of films, such as CALIGARI, METROPOLIS, and NOSFERAT, with preexisting curated annotation data and visualizations encompassing a larger body of films spanning the Weimar Cinema era. To achieve this, teachers and students alike need to leave their comfort zones in which they are experts in analyzing and interpreting films and film-related materials based on their mostly qualitatively informed expertise and training. Moreover, it is also only by bringing together approaches such as manual and automated video annotation that the pros and cons of both ways of working become visible.

Last but not least, one should feed and swallow one bite at a time, by accompanying hands-on approaches with theoretical reflections and readings. This aim is to use the momentum of students' personal explorations to address critical topics and issues instead of feeding them too much at the beginning of the course. Since a variety of perspectives and applications of digital methods in film studies does exist, one cannot and should not introduce students to "the" Digital Humanities in one class.

In the future, we would rather focus on specific aspects of digital methods that have been underrepresented in film studies research and teaching, in the field of video and text analysis, for example, or data modeling and data visualization. A stronger specialization would focus expectations and streamline at least a certain part of the course structure without completely taking away the lessons learned from the uncertainties and openness. As always, the more you understand about a particular topic and the more you are able to deal with the specific tools, the better you are able explain to your students the details and ramifications. Especially in digital methods, it is important not to lose sight of the bigger picture and keep the perspective open, while also staying domain-specific and close to the subject at hand.

Conclusion

Utilizing the classroom as a platform for experimentation and exploration, where both instructors and students engage with digital methods, tools, and data, yielded a valuable experience that provided a mix of anticipated and unexpected insights for both groups. Our initial idea of combining students' expertise in film studies with their familiarity with digital tools in daily use turned out to be only partially accurate. While certain individuals demonstrated prior experience and innovative approaches, these skills were more reflective of individual capabilities rather than generational traits, prompting consideration for future course design.

A crucial question that we should ask ourselves regularly is – what is it that we teach, what are the supposed outcomes and results of a class, both on the level of the individual student and of the group? Do we teach a specific knowledge of the domain that we are working in (in this case: film history) or do we teach specific skills, i.e., how to use a specific software or how to apply a specific method? In a broader sense, we hope that we are going both ways and ultimately do something more, that we are teaching ways of problem solving rather than just knowing particular facts or being able to operate a specific software. Since we are not educating specialists in particular fields that need to perform highly specialized tasks such as operating on human bodies or managing a complex machine, but we are rather training generalists that have a talent for problem solving and thinking multi-directionally, our courses have to be designed for giving the students these skills. Only by pursuing these goals can we meet the pedagogical mission of education (in German *Bildung*), which, according to David M. Berry

"teaches knowledge acquisition as a process rather than the acquisition of knowledge as a product."[26] And indeed, digital methods are a welcome additional process to the text-based seminars that will remain the cornerstone of humanities education.

Digital methods represent a specific challenge because the domain-based expertise has to be complemented by digital skills. As Andreas Fickers has argued: "The combination of the hermeneutical tradition of 'close reading' with the machine-based technique of exploratory 'distant reading' brings the humanities in a new dialogue with data and computer science which widens the horizon for both sides."[27] A potential strategy here is co-teaching with experts from the informatics domain, thereby transferring the *trading zone* into the classroom, or declaring the classroom to be another trading zone for digital film historiography. At the Marburg Center for Digital Culture and Infrastructure we are currently testing similar approaches, for instance, in the framework of the Master's program Cultural Data Studies, which offers innovative teaching courses such as a Cultural Data Management Lab.[28] These exploratory teaching opportunities are based on a rather cumbersome labor of small steps and might not announce themselves as the shining discovery of the new – digital tools, data and DH do not change humanistic subjects overnight. Yet again, these forays promise to bring into existence interesting new ideas and approaches with students, highlighting the proximity to and mutual interaction with the process of research.

References

Acland, Charles R., Eric Hoyt, and Kit Hughes. "A Guide to the Arclight Guidebook." In *The Arclight Guidebook to Media History and the Digital Humanities*, edited by Charles R. Acland and Eric Hoyt, 1–29. Falmer: REFRAME Books, 2016.

Ahrens, Sönke. *Experiment und Exploration: Bildung als experimentelle Form der Welterschließung*. Bielefeld: Transcript, 2010.

Berry, David M. "Introduction: Understanding the Digital Humanities." In *Understanding Digital Humanities*, edited by David M. Berry, 1–20. London: Palgrave Macmillan, 2012.

Brown, Kathryn. *The Routledge Companion to Digital Humanities and Art History*. New York: Routledge, 2020.

26 David M. Berry, "Introduction: Understanding the Digital Humanities," in *Understanding Digital Humanities*, ed. David M. Berry (London: Palgrave Macmillan, 2012), 7.

27 Fickers, "Forum: 'Mehr Mut zum Experimentieren'" (translation by the authors).

28 Marburg Center for Digital Culture and Infrastructure, "Cultural Data Management Lab," accessed March 27, 2023, https://www.uni-marburg.de/de/mcdci/studium/ma-cds/modulhandbuch/cultural-data-management-lab.

Cordell, Ryan. "How not to Teach Digital Humanities." In *Debates in the Digital Humanities*, edited by Matthew K. Gold and Lauren F. Klein, 459–474. Minneapolis, MN: University of Minnesota Press, 2016.

Döring, Karoline Dominika, Stefan Haas, Mareike König, and Jörg Wettlaufer. *Digital History: Konzepte, Methoden und Kritiken Digitaler Geschichtswissenschaft*. Berlin: De Gruyter Oldenbourg, 2022.

Drucker, Johanna. *The Digital Humanities Coursebook: An Introduction to Digital Methods for Research and Scholarship*, 1st ed. Abingdon, Oxon; New York: Routledge/Taylor & Francis, 2021.

Fickers, Andreas. "Forum: 'Mehr Mut zum Experimentieren' – Digitale Lehre – Interview mit Andreas Fickers (Universität Luxembourg)." *H-Soz-Kult*, April 10, 2020. Accessed March 27, 2023. https://www.hsozkult.de/debate/id/fddebate-132393.

Fickers, Andreas, and Juliane Tatarinov, eds. *Digital History and Hermeneutics*. Berlin: De Gruyter Oldenbourg, 2022. https://doi.org/10.1515/9783110723991.

Fickers, Andreas, Juliane Tatarinov, and Tim van der Heijden. "Digital History and Hermeneutics. Between Theory and Practice: An Introduction." In *Digital History and Hermeneutics: Between Theory and Practice*, edited by Andreas Fickers and Juliane Tatarinov, 1–19. Berlin, Boston: De Gruyter Oldenbourg, 2022.

Fickers, Andreas, and Tim van der Heijden. "Inside the Trading Zone. Thinkering in a Digital History Lab." *Digital Humanities Quarterly* 13, no. 3 (2020). Accessed March 27, 2023. http://www.digitalhumanities.org/dhq/vol/14/3/000472/000472.html.

Flückiger, Barbara. "Timeline of Historical Film Colors." Accessed March 27, 2023. https://filmcolors.org.

Flückiger, Barbara, and Gaudenz Halter. "Methods and Advanced Tools for the Analysis of Film Colors in Digital Humanities." *Digital Humanities Quarterly* 14, no. 4 (2020). Accessed March 27, 2023. http://www.digitalhumanities.org/dhq/vol/14/4/000500/000500.html.

Franken, Lina. *Digitale Methoden für qualitative Forschung. Computationelle Daten und Verfahren*. Stuttgart: UTB, 2022.

Marburg Center for Digital Culture and Infrastructure. "Cultural Data Management Lab." Accessed March 27, 2023. https://www.uni-marburg.de/de/mcdci/studium/ma-cds/modulhandbuch/cultural-data-management-lab.

Moretti, Franco. "Planet Hollywood." *New Left Review* 9 (2001): 90–101.

Sittel, Julian. "Digital Humanities in der Filmwissenschaft." *MEDIENwissenschaft: Rezensionen / Reviews* 4 (2017): 472–489.

Young, Christopher J., Michael C. Morrone, Thomas C. Wilson, and Emma Annette Wilson. *Quick Hits for Teaching with Digital Humanities: Successful Strategies from Award-Winning Teachers*. Bloomington: Indiana University Press, 2020.

Filmography

Das Cabinet des Dr. Caligari, dir. Robert Wiene. Germany, 1919/1920.

Faust – eine deutsche Volkssage, dir. Friedrich Wilhelm Murnau. Germany, 1925/1926.

Frauenschicksale, dir. Slatan Dudow. East Germany, 1951/1952.

Der Hauptmann von Köln, dir. Slatan Dudow. East Germany, 1956.

Nosferatu – eine Symphonie des Grauens, dir. Friedrich Wilhelm Murnau. Germany, 1921/1922.

Orlac's Hände, dir. Robert Wiene. Austria, 1924.
Schloß Vogelöd, dir. Friedrich Wilhelm Murnau. Germany, 1921.
Tatort: Für immer und dich, dir. Julia von Heinz. Germany, 2019.
Tatort: Der letzte Schrey, dir. Mira Thiel. Germany, 2020.
Tatort: Der Herr des Waldes, dir. Christian Theede. Germany, 2021.
Verwirrung der Liebe, dir. Slatan Dudow. East Germany, 1958/1959.

IV Visualization

Marcus Burkhardt and Skadi Loist

Visualization In/As Digital Media Studies

Introduction

In this chapter, we are concerned with visualization as a method of digital study that we are experimenting with in our research practice.[1] Our aim is to refrain from the media theoretical impulse to look at and reflect upon visualization from a distance as something others have created, do, or use. This necessarily includes a reflexivity that is sensitive to media theoretical insights, but is in effect practical. We find inspiration for such an approach in the works of Johanna Drucker,[2] and Deb Verhoeven,[3] for example, as well as others who combine media theoretical reflection with innovative practices of visualization.[4] At the same time we have to acknowledge and admit that we ourselves are dilettantes in the field of visual design, who at best cautiously and experimentally adopt practices of research with visual forms. In our eyes, this opens valuable perspectives on the potential of visualization as a mode of digital media studies.

Visual Epistemology: Promises and Challenges

Data abounds in digital cultures. Visualizations have become increasingly important when dealing with vast amounts of data. In the words of Ben Fry, they "help us 'see' things not previously understood in abstract data."[5] As such, visualizations provide a visual interface to data which is comprehensible by and for humans. The capacity to

1 This work was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG), project number 262513311 (CRC 1187: "Media of Cooperation") and the German Federal Ministry of Education and Research (BMBF), project number 01UL1710X ("Film Circulation on the International Film Festival Network and the Impact on Global Film Culture").

2 Johanna Drucker, *Visualization and Interpretation: Humanistic Approaches to Display* (Cambridge, MA: MIT Press, 2020).

3 Deb Verhoeven, "Visualising Data in Digital Cinema Studies: More Than Just Going Through the Motions?" *Alphaville: Journal of Film and Screen Media* 11 (2016): 92–104, https://doi.org/10.33178/alpha.11.06.

4 We are further inspired by the works of many pioneers, practitioners, and evangelists of information and data visualization such as John Tukey, Edward Tufte, Mike Bostock, Nadieh Bremer, and Shirley Wu, to name just a few.

5 Ben Fry, "Computational Information Design" (PhD diss., Massachusetts Institute of Technology, 2004), 33.

https://doi.org/10.1515/9783111082486-015

render unintelligible data intelligible and to convey complex matters in ways that are relatively easy to understand proves to be both a promise and a peril of visualizations. Graphs, maps, and diagrams offer seemingly immediate insight and overview, and communicate clear, unambiguous messages. At the same time, they can be suggestive, difficult, misleading, and even deceiving.

Visualizations share this problematic ambivalence with images in general, which are said to be worth a thousand words. The origins of this saying can be traced back to the early twentieth century when both journalists and marketers were pondering the transformation of their professions and argued for the use of images. Frederick Barnard played an important part in the popularization of the phrase, when, in 1927, he published an ad in the trade journal *Printers' Ink* that stated: "One Picture Worth Ten Thousand Words." Barnard claimed that the saying was of Chinese origin, a myth that still lingers on.[6] The axiom articulates a promise that something can be shown in pictures that would be difficult to say in words. The increasing popularization of visualizations in digital cultures in general, as well as digital research in particular, is also based on this premise. Yet, while advertising is often suspected of seducing customers with misleading images, visualizations are typically associated with a certain claim to truth and correctness. However, the two are not mutually exclusive: visual expressions unfold a different logic from verbal language. Images escape the propositional logic of language, as Ernst Gombrich argues in *Art and Illusion*:

> Logicians tell us [. . .] that the terms "true" and "false" can only be applied to statements, propositions. And whatever may be the usage of critical parlance, a picture is never a statement in that sense of the term. It can no more be true or false than a statement can be blue or green.[7]

Images in themselves cannot lie according to Gombrich. However, they can be entangled in lies when, for example, they are incorrectly labeled as depicting something they are not, or are implicitly taken as such. Only when related to an explicit claim – articulated in language – or implicit assumptions – based on experiences and practices – can an image be judged to be truthful or not. The subject of such an assessment is the information conveyed by the image in relation to the claimed or assumed statement. The insight that images are non-propositional does not devalue

6 The East Asia Institute at Ludwigshafen University of Applied Sciences maintains a German language website with Chinese proverbs, that contains an entry for the saying just for the sake of pointing out that this is a misconception; Ostasieninstitut, "Ein Bild sagt mehr als tausend Worte. 一画胜千言。," accessed October 14, 2023, https://ostasieninstitut.com/bibliothek/sprichwoerter-ostasiens/ein-bild-sagt-mehr-als-tausend-worte-一画胜千言。-yi-hua-sheng-qian-yan/.

7 Ernst H. Gombrich, *Art and Illusion: A Study in the Psychology of Pictorial Representation* (London: Phaidon, 1984), 56.

them. On the contrary, Gombrich attributes a genuine capacity to images. They do not convey truth, but information. Much of what Gombrich writes about in *Art and Illusion*, as well as the essays collected in *The Image and the Eye*, is a detailed analysis of what information different kinds of images can or cannot convey.

What is important here for our discussion of visualization as a method in digital research is that images always show too much and say too little. Even more, as media philosopher Dieter Mersch argues, images do "not show the modalities of their showing in the process – they elude the visualization of their function where it concerns the creation of visibility."[8] This is especially challenging when conducting research with visualizations whose purpose is to provide insights into what is not yet known. Johanna Drucker puts this problematic as follows: "Most information visualizations are acts of interpretation masquerading as presentation. In other words, they are images that act as if they are just showing us what is, but in actuality, they are arguments made in graphical form."[9]

In her 2020 book *Visualization and Interpretation*, Drucker explores the non-representational dimension of graphical expression. While visualizations are often taken as mere representations, she urges us to acknowledge that visualizations also have a generative dimension: they add some level of interpretation. Therefore, visualizations are not neutral, objective, or universal (re)presentations. Rather they offer perspectives that always remain "partial, situated, and historically/culturally specific."[10] This typically remains hidden in the visualization of data and information, which is why she explores new forms of visual expressions that are (better) suited to model interpretations, rather than presenting them as seemingly objective truth. In this respect Drucker's project of visual epistemology, which she conceptualizes as *graphesis* in distinction to *mathesis*, is closely aligned to the "critical epistemological approach of the humanities"[11] in general.

Envisioning graphical forms that expose their contingency and situatedness rather than hiding them is not an easy undertaking. It finds its limits in Mersch's thesis, quoted earlier, that, in what images do show, something always remains unshown. There remains much room for improvement, as Drucker shows in the appendix of *Visualization and Interpretation*, which proposes five principles of

8 Dieter Mersch, "Gaze and Withdrawal: On the 'Logic' of Iconic Structures," in *Technology and Desire: The Transgressive Art of Moving Images*, ed. Rania Gaafar and Martin Schulz (Bristol: Intellect, 2014), 348.

9 Johanna Drucker, *Graphesis: Visual Forms of Knowledge Production* (Cambridge, MA: Harvard University Press, 2014), 10.

10 Drucker, *Visualization and Interpretation*, 111.

11 Drucker, *Visualization and Interpretation*, 5.

making "visualization rooted in interpretation":[12] "graphic argument, generative (affective) metrics, inflection, point of view systems, and direct input."[13] These principles reflect practices in which the exploration of visual forms takes primacy when, for example, changes to their form effect changes to underlying data structures (direct input) or employ visual inflection that "nuances an otherwise blunt or reductive visual form."[14] An example of an inflection is the use of brightness, glow, or other visual cues to indicate additional relational attributes of graphs, such as the eigenvector centrality or betweenness centrality of network diagrams, that are otherwise invisible.

The design principles outlined by Drucker invite the iterative exploration of new forms of graphical expression that go beyond generic visualization techniques afforded by contemporary visualization platforms.[15] Drucker's critique of existing software tools as well as forms of graphical expression stems from her decade-long engagement in projects that push the boundaries of visual epistemology. While we are sympathetic to this critique and the ambitious goals Drucker pursues, we feel a certain level of unease therewith as in our own research, we rely heavily on applications such as RAWGraphs, Gephi, or Tableau. Instead of following Drucker in her attempt to move beyond existing visualization platforms and applications, we want to ask how to make positive use of them in digital media studies despite their shortcomings.

We acknowledge that existing visualization tools are limited. At the same time, however, we contend that these limitations are what allows us, as non-experts in visual design, to explore the possibilities of visualization techniques as a method to engage with our research subjects in new ways. Yet, in this regard, we recognize that notions of visualization platforms as quasi-magical tools that lead to desired results and insights at the touch of a button are just as problematic as viewing visualizations as immediately comprehensible forms of expression. To some extent, these misconceptions are fueled by software developers and visualization experts alike when they advertise the ease of use of their applications and praise the communicative qualities of visual representations. John Tukey might serve as an example of the latter claim here. In 1977, in his seminal book *Exploratory Data Analysis* he declared:

12 Drucker, *Visualization and Interpretation*, 140.
13 Drucker, *Visualization and Interpretation*, 141.
14 Drucker, *Visualization and Interpretation*, 147.
15 Drucker, *Visualization and Interpretation*, 140.

> pictures based on exploration of data should force their messages upon us. Pictures that emphasize what we already know – "security blankets" to reassure us – are frequently not worth the space they take. [. . .] The greatest value of a picture is when it forces us to notice what we never expected to see.[16]

While Tukey underlines that visualizations can and should be more than secondary reassurances of what is already known, the rhetoric that they can *force a message upon* humans could be interpreted as implying that understanding visualizations is immediate and universal. We do not claim that Tukey actually held this opinion: perhaps he only wanted to distinguish better types of graphical expression from poorer ones with his pointed formulation. Yet, this relates to a rhetoric of immediacy that is highly problematic, because reading visualizations needs to be learned before one can reason about them.

The semiotician Charles Sanders Peirce called this form of reasoning "diagrammatic."[17] For Peirce, diagrams are a certain type of iconic sign, different from pictures and metaphors. Diagrams do not display qualities of objects as pictures do or establish similarities between objects like metaphors, but rather show relations, structures, functional logics, and sequences of events. At the same time, diagrams open up certain possibilities for reconfiguring the relationships depicted. Pie charts, for example, allow for the speculative redistribution of proportions in relation to an abstract whole. In newscasts following elections, this aspect of pie charts and similar diagrams is frequently used to present potential majorities that could be formed by coalitions. Diagrams used in such a manner are media of visual thinking. This mode of thinking with diagrams through their possible reconfiguration is deductive in nature, according to Peirce, while their creation is abductive, and their validation is inductive. However, the rules of deductive variation of diagrams are not universal, as brain teasers like the nine dots puzzle show. The task of connecting nine dots arranged in a square with four straight lines without lifting your pencil can only be solved when the lines are drawn beyond the boundaries of the square. Finding this solution is challenging because many assume that the lines must remain inside the virtual box outlined by the dots. (Literally) thinking outside this box is necessary to solve the puzzle. When using visualization as a method of research, however, it is imperative to explore, explicate, and reflect upon the rules intrinsic to different forms of diagrammatic expression rather than transgressing

16 John W. Tukey, *Exploratory Data Analysis* (Reading: Addison-Wesley, 1977), vi.

17 For extensive reconstructions of Peircian diagrammatics, see Frederik Stjernfelt, *Diagrammatology: An Investigation on the Borderlines of Phenomenology, Ontology, and Semiotics* (Dordrecht: Springer, 2007); Matthias Bauer and Christoph Ernst, *Diagrammatik: Einführung in ein Kultur- und medienwissenschaftliches Forschungsfeld* (Bielefeld: Transcript, 2010).

them. This provides the foundation for the skillful and reflexive engagement with data through forms of visual expression. Such an engagement is necessarily iterative, exploratory, and full of trials and errors.

Exploratory Research Through Visual Network Analysis

The exploration of datasets is comparable to "detective work."[18] It requires specific detailed knowledge of the subject area of analysis, but there are a number of general techniques that can guide analysis. In what follows, we will describe some strategies, approaches, and practices of exploring data in the process of visualization. For Tukey, as a mathematical statistician, the focus was on exploring quantitative relationships. He therefore proposed forms of diagrammatic expression that expose the numerical distribution of datasets such as stem-and-leaf, box-and-whisker, or scatter plots.[19] By contrast, we are interested in gaining insights into the structures, genealogies, and interrelationships of our mediatized world.

For this, we use forms of Visual Network Analysis (VNA). Network analysis has become an increasingly common method in a range of vastly different disciplines to explore relational datasets and illustrate findings. VNA is not without its problems, however. Tommaso Venturini, Mathieu Jacomy, and Pablo Jensen have pointed them out, but also identified the qualities that make VNA especially insightful for exploring the richness of relational datasets and medium-sized networks:

> Unlike a bar chart or a scatter plot, a points-and-lines chart [i.e. a network graph] is not straightforwardly shaped by its rules of construction. Instead, its form depends on the relationships between its elements in ways that cannot be easily recognized, outside trivially simple networks such as trees, stars, or grids. Graphs are multidimensional mathematical objects and visualization squeezes them in a two-dimensional space, flattening their complexity. No wonder that scientists are wary of graph charts. [. . .] The same ambiguity that makes network charts unfit for hypothesis confirmation, we contend, makes them invaluable for exploratory data analysis.[20]

18 Tukey, *Exploratory Data Analysis*, 1.

19 Tukey, *Exploratory Data Analysis*, 1ff., 39ff., and 205ff.

20 Tommaso Venturini, Mathieu Jacomy, and Pablo Jensen, "What Do We See When We Look at Networks: Visual Network Analysis, Relational Ambiguity, and Force-Directed Layouts," *Big Data & Society* 8, no. 1 (2021): 2, https://doi.org/10.1177/20539517211018488.

In their defense of VNA, Venturini et al. point to the "evocative power of network visualizations," when this topological type of visualization turns relational structures into visual patterns.[21] Network analysis has evolved over time. As early network science worked with small datasets, and with a few dozen nodes and edges, researchers were able to read networks as functional diagrams, such as flowcharts or trees. Venturini et al. point out that this diagrammatic approach has become untenable for the medium and large networks that recent network analysis of social or cultural big data is working with.[22] They point to the usefulness of utilizing a topological perspective to explore large datasets and look for relational structures such as clustering, centrality, density, and, as a result, detect patterns.[23]

Conceptualization and Iterative Analysis

The examples that follow draw on projects we pursued separately with different collaborators. The experiences we had in these different contexts formed the basis of a conversation about deploying visualization as a method in digital media studies without being experts in design or programming.

Our first case study is part of the project "Film Circulation on the International Film Festival Network and the Impact on Global Film Culture," which set out to study the complex temporal and spatial circulation patterns of films traveling in the film festival sector.[24] The project aimed to employ approaches from the digital humanities and geo-visualization in order to develop new analytical methods for film studies and film festival studies in particular. The project operated on two levels: first, to study how films circulate and, second, how festivals are connected in the film festival ecosystem through the circulation of films. It worked from the theorization of the festival circuit as a network and drawing inspiration from pioneer-

21 Venturini, Jacomy, and Jensen, "What Do We See," 13.

22 Venturini, Jacomy, and Jensen, "What Do We See," 9.

23 Venturini, Jacomy, and Jensen, "What Do We See," 10.

24 Skadi Loist, "Studying Film Circulation: Moving Film Festival Research to an Evidence-Based, Global Perspective," in *Shaping Film Festivals in a Changing World: Practice and Methods*, ed. Dorota Ostrowska and Tamara Falicov (Amsterdam: Amsterdam University Press, 2025), 211–227; Skadi Loist and Zhenya Samoilova, "How to Capture the Festival Network: Reflections on the Film Circulation Dataset," *NECSUS: European Journal of Media Studies* 12, no. 1 (2023): 363–390, https://doi.org/10.25969/mediarep/19615.

ing cultural data analytics projects, such as the Kinomatics group,[25] which utilize a variety of digital methods and data visualization techniques. This led to a collaboration between Skadi Loist, and data and network specialists Zhenya Samoilova, Katharina Burgdorf, and Martha E. Ehrich. Together they began to develop a network analysis approach for film festival research, which uses VNA to approach festivals from a data perspective.

The second case study was concerned with the evolution of Facebook as a platform. Tatjana Seitz, Anne Helmond, Fernando van der Vlist, Angeles Briones, and Marcus Burkhardt collaborated in unpacking the complex history of the Facebook platform through the lens of its Application Programming Interface (API) as core infrastructure.[26] The methodological aim of the project was to explore how much can be learned about the emergence, stabilization, commodification, and transformation of the platform through engagement with publicly available documentation aimed at developers. They were encouraged to create applications that enhance the companies' social network(s) and services – mainly Facebook, but also Instagram and WhatsApp – or to integrate their networking capabilities into other applications.

The Film Circulation project was an exploratory endeavor from the outset. As a project that was designed to develop digital methods and tools to study the festival ecosystem and film circulation from a quantitative perspective, a large part of the work entailed translating concepts and methods and thus enabling interdisciplinary collaboration.[27] In the beginning, the relatively young field of film festival studies was dominated by historical case studies and theorizations in film studies. New Cinema History and Media Industries Studies had developed in parallel, and had pointed the way to data-driven research as well as utilizing industry knowledge. Both influenced the project in developing operationalization of the complex mechanisms behind festival operations. As a field that had not readily been datafied, the data had to be identified and transformed into a dataset suitable for digital research. The first data visualizations were diagrams that accompanied statistical analysis about the field, and the trajectory of films – the time they spent traveling

25 Verhoeven, "Visualizing Data in Digital Cinema Studies"; Deb Verhoeven et al., "Disciplinary Itineraries and Digital Methods: Examining the Kinomatics Collaboration Networks," *NECSUS: European Journal of Media Studies* 9, no. 2 (2020): 273–298, https://doi.org/10.25969/mediarep/15320.

26 Fernando N. van der Vlist et al., "API Governance: The Case of Facebook's Evolution," *Social Media + Society* 8, no. 2 (2022), https://doi.org/10.1177/20563051221086228.

27 Verhoeven et al., "Disciplinary Itineraries."

on the festival circuit, the amount of screenings, or revenue gathered.[28] For a relational analysis between films and festivals, as well as between festivals, VNA has proved to be a complex and productive tool to examine not only the movement of films, but the interconnectedness and positioning of festivals within the network.

Following the theorization of the festival sector as a network with reference to Actor-Network-Theory,[29] we conceptualized the festival sector as a network of festivals connected through shared films.[30] Thus, if a film was screened at festivals A and B, we understood these two festivals as sharing a tie and being connected. In a first step, we conceptualized the festival sector as an undirected network, in which festivals constitute nodes and shared film screenings constitute ties. This brought the interconnection of festivals into focus. We used the network visualization software Gephi and the commonly applied ForceAtlas2 algorithm to visualize and identify patterns in the data. Figure 1 shows a network visualization of our festival network based on the exchange of films. By calculating network measures, we could analyze the position of festivals within the network. Our analysis built on network visualizations that reduce complexity and illustrate relationships between festivals on a large scale. In this way, the relational character of the festival ecosystem is highlighted, which offers insights into the core and periphery of the festival sector. This method attempted a visual translation of festival research knowledge and further utilization of gained insights or questions.

We knew from industry knowledge and festival theorization that important industry festivals, such as Cannes, Berlin, and Toronto (TIFF) are at the core of the network. The algorithm differentiates the core from periphery in such a way that festivals grouped in the core are the most interconnected through screening the same film, while festivals on the periphery have fewer connections. In addition, the distance between the nodes also represents how connected (via shared films) the festivals are. In Figure 1 we can see that, while Cannes, Berlin, and TIFF are grouped together in the core of the network, the International Documentary

28 Skadi Loist and Zhenya Samoilova, "Digitale Methoden und Open Media Studies: Zur Erforschung von Filmfestivalruns," *Open Media Studies Blog* (May 8, 2019), accessed April 30, 2023, https://doi.org/10.5281/zenodo.7934483.

29 Thomas Elsaesser, "Film Festival Networks: The New Topographies of Cinema in Europe," in *European Cinema: Face to Face with Hollywood* (Amsterdam: Amsterdam University Press, 2005), 82–107; Marijke de Valck, *Film Festivals: From European Geopolitics to Global Cinephilia* (Amsterdam: Amsterdam University Press, 2007).

30 Skadi Loist, Evgenia Samoilova, Martha Emilie Ehrich, and Katharina Burgdorf, "New Approaches to Studying Film Circulation: Network Analysis in the Festival Ecosystem," paper presented at Festivals et dynamiques cinématographiques trans(nationales): formes de production, de circulation et de représentation, Toulouse, March 28–31, 2022, *Zenodo*, https://doi.org/10.5281/zenodo.10004780.

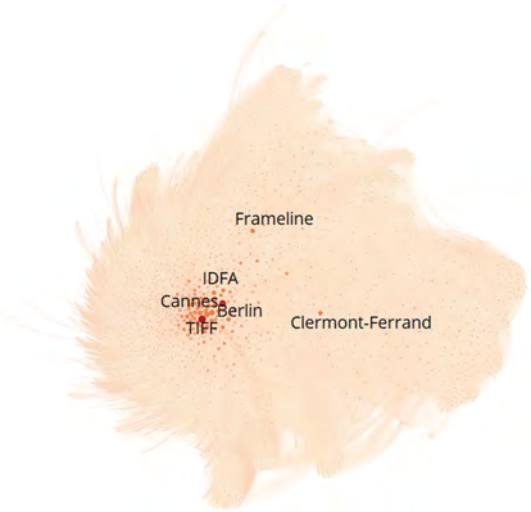


Figure 1: Core and periphery of the festival network. Visualization of degree centrality in an undirected network of sampled film festivals in the International Movie Database (IMDb), where color and size of the nodes (festivals) is scaled by the degree centrality in the entire dataset where at least one film is shared (3,167 festivals and 183,758 connections). Visualization by Zhenya Samoilova using the software Gephi and its ForceAtlas2 algorithm, https://doi.org/10.6084/m9.figshare.25585602.v1.

Film Festival Amsterdam (IDFA) is a bit further away, but still quite close to the core. On the other hand, Clermont-Ferrand and Frameline seem to be quite far distant from the core.

When using this visualization, also beyond film studies and film festival studies, it is important to explain the context, creation and logic of network visualization without having to rely on fine-grained expertise. As the festival network, for example, draws on temporal and spatial data, which certainly is connected to geopolitical issues, it is relevant to explain that the algorithmic logics of VNA used to create the visualization above (Figure 1), do not depict any geographic relations of the data. Thus, as discussed in the beginning of this chapter, when utilizing the method of visualization within film festival studies – and not only there – the process of its creation and context need to be explained. For this there are several translational processes at play. There is a translation from the research context (e.g., film festival research) into data operationalization and conceptualization of the method and the mode of inquiry, in this case VNA. The visualizations need to be interpreted in line with the employed network measures, which are by no means self-explanatory or self-evident to a lay and sometimes not even expert audience. However, this also makes VNA a useful tool for exploration. The process

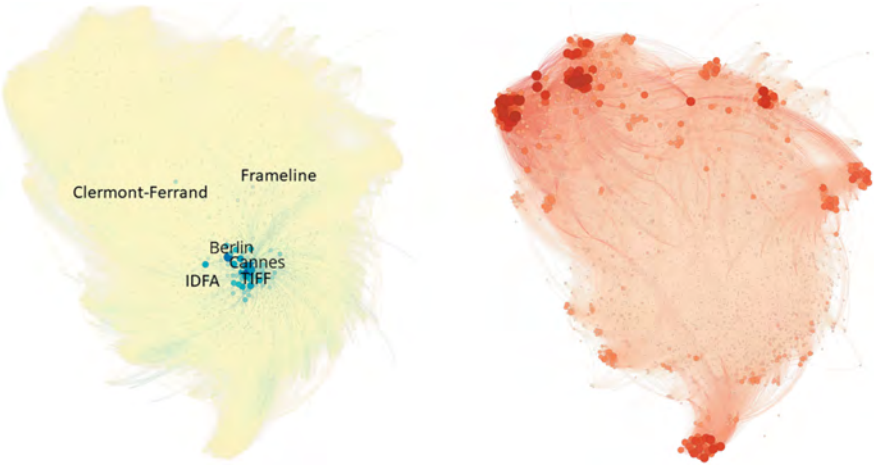


Figure 2: Network visualization of sending festivals (left) and receiving festivals (right). Left: Visualization of outdegree centrality calculated in the directed network of film festivals sampled in the IMDb, where color and size of the nodes (festivals) is scaled by the outdegree centrality (3,167 festivals and 183,758 connections). Right: Visualization of a film's festival run in an undirected network, size and color of nodes correspond to the average position of the festival in the festival run (3,167 festivals and 183,758 connections). Visualizations by Zhenya Samoilova using the software Gephi and its ForceAtlas2 algorithm, https://doi.org/10.6084/m9.figshare.25585641.v1.

of grasping the meaning of a visualized network needs to include going back to the original dataset and theoretical context and possibly changing the algorithm and measures used to visualize the network relations. In fact, we went back as a research team to the dataset and previous models to try to explain the structures of the festival ecosystem and to consider both the potential and the limitations of the available dataset, which enriched the process of interpretation.

Comparative Analysis

In Figure 1, we focused on the core and periphery positions of festivals, capturing how festivals are interconnected through shared film screenings. Since we were interested in the festival runs of films, which entails a *temporal* component, where the sequence of screenings carries important meaning. To do so, we went back to the dataset and the underlying data structure for visualization. We considered the temporal flow of films, i.e., the films' festival runs, within the festival network. In Figure 2, we comparatively visualize how festivals are connected through *undirected* (left) and *directed* connections (right). We used different tools

to make the direction of film flow and the resulting positioning of the festivals in the festival run visible.

In Figure 2 left, we stick to the previous conceptualization of the festival sector as a so-called undirected network, in which a shared film screening constitutes a tie between festivals A and B. Yet, we added an external metric to the dataset, namely the average position of a festival within the festival run. To do so, we ordered all the festivals from 1 to n for each film, then calculated the average position for a given festival. This calculation was used to inform the visualization Figure 2 left, where the color and size of the nodes corresponds to the position of the festival run sequence. The bigger and darker the node is, the further the festival is from the start of the festival run, thus highlighting the receiving festivals at the end of the festival run.

In Figure 2 right, we used a slightly different approach. Here we conceptualize the sector as a directed network. When the same film is presented first at festival A and then at festival B, festival A is connected to festival B through a directed tie. In other words, festival A sends a tie – a film in our case – to festival B, and festival B receives a tie from festival A. For our purposes we visualized information about the festival runs and the position of festivals within them to assess which festivals were "sending festivals" (Figure 2, left) and which festivals were "receiving festivals" (Figure 2, right) by calculating their indegree centrality (the number of films a festival receives by other festivals).

Comparatively, both approaches, that of an undirected and a directed network, enable us to discuss core-periphery dynamics and thereby help make deductions about the hierarchical structure of the festival sector. For festival studies, VNA techniques are useful to analyze which festivals are "premiere" festivals that send films off down the path of a festival run, and which are the festivals that are "the end of the line," and that provide the exhibition spaces for films.[31]

But, as mentioned above, the complexity of the visualized network always demands contextualization of the images and an explanation of how the data was prepared to create the networks at hand. This meant that, as part of the research process, we went back to the dataset and available measures, as well as further sources, to develop explanations and delve deeper into smaller groups of festivals, and in addition changed our mode of analysis when looking at different scales.

31 Skadi Loist, "Zirkulation im Netzwerk: Eine Betrachtung zur Zirkulationskraft von Filmfestivals," *Zeitschrift für Medienwissenschaft* 12, no. 2 (2020): 55–63, http://dx.doi.org/10.25969/mediarep/14833.

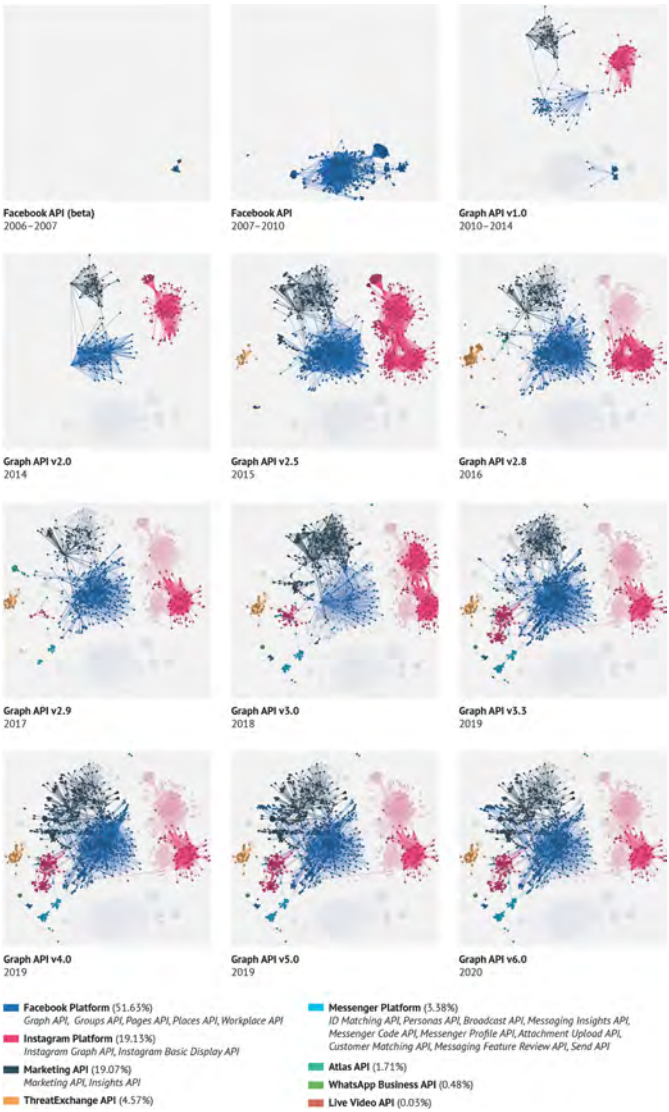


Figure 3: Small multiple visualizations of the evolution of Facebook's API ecosystem between 2006 and 2020 (selection). Visualization by Van der Vlist et al. using the software Gephi and its ForceAtlas2 algorithm.[32]

32 For the complete visualization of the evolution of the Facebook API ecosystem between 2006 and 2020, see Van der Vlist et al.,"API Governance," 7.

A comparative approach was also used in the research project tracing the evolution of Facebook/Meta's ecosystem of APIs. The network graphs in Figure 3 are based on archived versions of the developer documentation of the platforms' various APIs – the Facebook Graph and Marketing APIs, the Instagram API, the ThreatExchange API, the Messenger Platform API, the Atlas API, the WhatsApp Business API, and the Live Video API. APIs offer so-called endpoints that allow access to platform entities, including users, pages, posts, images, comments, reactions, etc., that have potential relations to each other, e.g., images, posts, or comments created by a user. In the language of the Graph API, Facebook/Meta's oldest API, endpoints are nodes and the relationships between nodes are edges. Therefore, visualizing the APIs as graph networks seemed natural. The network data was extracted from the developer documentation for different versions of the API, which we in part retrieved from the Internet Archive's Wayback Machine and in part collected ourselves from the live web. For each version, the documentation consists of multiple webpages that are interlinked. We came to realize that those pages represent the nodes of APIs and the links represent the edge-relationships between nodes. In total 63,027 documents were retrieved and analyzed for creating the network visualization using Gephi and its ForceAtlas2 layout algorithm (Figure 3). For engaging with the development of these APIs over time, we created multiple versions of the same graph. Each view depicts the API ecosystem at a certain point in time or for a certain version of the core Graph API. The different colors in each graph indicate the different APIs Facebook/Meta maintains for its various services and offerings. In a post-processing step, we added a semi-transparent network graph of all nodes and edges known up to that point and arranged the resulting graphs in a small multiple visualization. It was only through this comparative juxtaposition of different temporal views of the same network that it became legible to us, in that we could see the expansion and diversification as well as integration of the platform's API ecosystem.

Scalable Readings

Another important technique in deploying visualization as a research method is the engagement with the objects of study at different scales. The notion of scalable reading was proposed by digital humanities scholars to overcome the apparent antagonism between *close* and *distant* reading: "With Google Earth you can zoom in and out of things and discover that different properties of phenomena

are revealed by looking at them from different distances."[33] Niels-Oliver Walkowski and Johannes Pause adapted the idea of scalability to the digital study of films, using the term "scalable viewing."[34] In our case studies of film circulation on the one hand and platform infrastructures on the other, it also proved pivotal to consider our research objects at different levels or scales. Yet merely zooming into network graphs does not provide additional insights, as is the case with Google Earth. For looking at networks at different scales, it is usually necessary to filter them according to different measures or properties.

Figure 4 shows a filtered view of the relations between the most interconnected film festivals. The connections of several thousand festivals shown in Figures 1 and 2 presented us with a problem of opaqueness, where hardly anything is visible in the visualization made up of several thousand nodes and ties overwriting each other.[35] We therefore opted to zoom in and reduce the nodes and ties to be able to discern specific characteristics. For this we looked at the top two percent of festivals based on the eigenvector centrality measure, i.e., the festivals that are highly connected to festivals that are also highly connected – what might be called "popular" festivals, which are connected to other "popular" festivals.

Due to the reduced number of nodes, we were able to add further contextual information into the visualization and analysis, such as information about the categorization of the festivals based on the accreditation by the International Federation of Film Producers Associations (FIAPF). The FIAPF accreditation gives an indication of the historical ranking and importance of a festival within the sector. A FIAPF-accredited "competitive film festival" (red) corresponds to the term "A-list" festival, which in industry jargon means that it is one of the most important festivals for premieres and film business.

The core of the resulting visualization in Figure 4 shows the festivals that one would expect to see, e.g., A-list festivals such as Cannes, Toronto, Berlin, Venice, and Karlovy Vary. But within the top two percent we also find festivals that, based on industry logic, are not known to be big global players with a big market. For example, we find Filmfest Hamburg and Filmfest München, circled in red in Figure 4, which are only the second or third biggest international film festivals in

33 Martin Mueller, "Scalable Reading," *Scalable Reading* (blog), 2020, accessed October 14, 2023, https://sites.northwestern.edu/scalablereading/2020/04/26/scalable-reading/.

34 Johannes Pause and Niels-Oliver Walkowski, "Scalable Viewing," *Open Media Studies Blog*, 2019, accessed October 14, 2023, https://zfmedienwissenschaft.de/online/scalable-viewing.

35 This is easily illustrated by opening a PDF version of a Gephi network visualization of this network, where it takes several seconds to open the final image and one can observe how each connection line is drawn over the other. See Skadi Loist and Evgenia Samoilova, "Visualization of Festival Run Sequence," *figshare*, 2022, https://doi.org/10.6084/m9.figshare.24588954.v1.

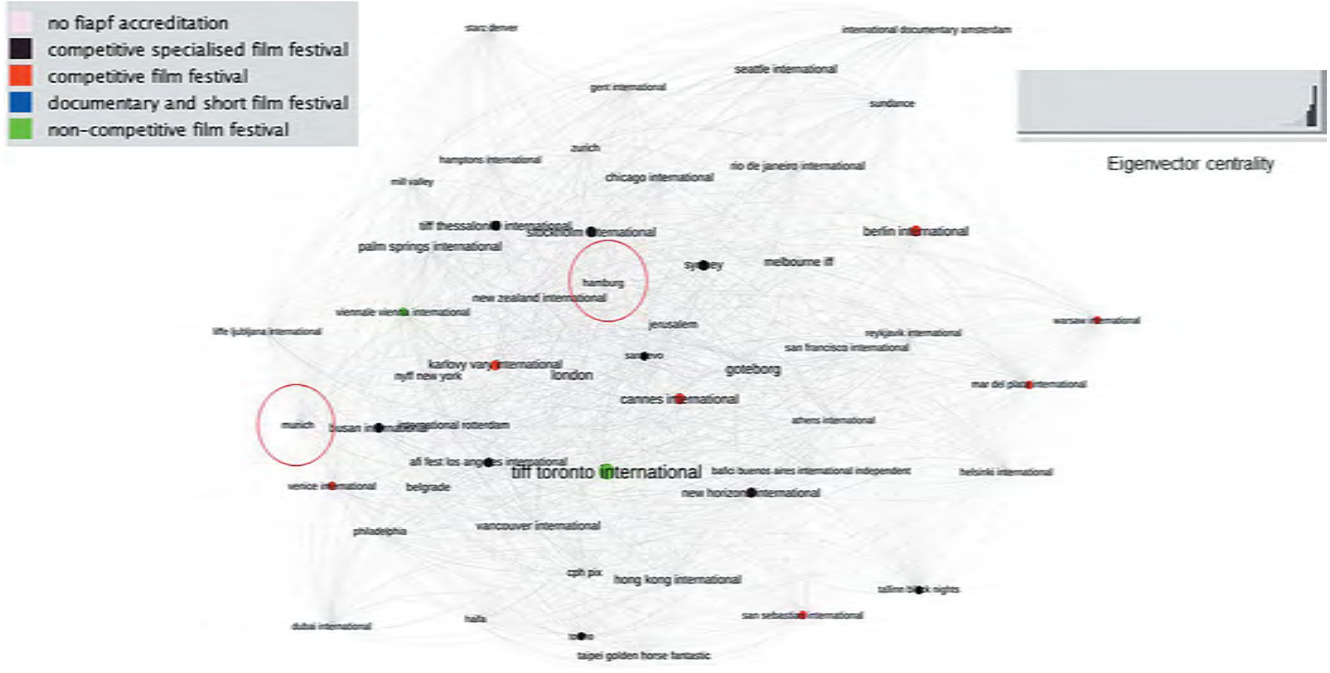


Figure 4: Highly interconnected core: top two percent festivals. Visualization of a directed network of the top 52 festivals (1.64% of the festivals) with the highest eigenvector centrality (52 festivals and 1,326 connections; 0.72% of all the connections), where color of the nodes (festivals) is coded by festival categorization based on FIAPF accreditation. Visualization by Zhenya Samoilova using the software Gephi and its ForceAtlas2 algorithm, https://doi.org/10.6084/m9.figshare.25585653.v1.

Germany. They have mainly regional or national impact and are not commonly thought of as having international impact on the circuit like A-list events. However, this visualization shows their significance for the exchange of films within the network, with Hamburg being the more centrally positioned and thus more connected of the two.

In the Facebook platform study, scalability was achieved by amending the ecosystem view provided by the network graph visualization with other types of visualizations that focus on different aspects of the same data, or draw on additional developer resources as research data. An example of this is the visualization of how single platform entities evolve over time (Figure 5). At the center of Facebook is the user entity that consists of descriptive attributes (so-called fields) and potential relationships (so-called edges), which typically represent user activities on the platform. To analyze how users evolved over time, we extracted the fields and edges of the user object from the documentation for all available versions and generated a flow diagram using the RankFlow visualization tool.[36] Because RankFlow diagrams were originally conceived to visualize changes in the order of search results over time, we had to adapt the visualizations generated by the tool to our purposes and research interests.

In our case, the order in which fields appeared in the developer documentation was irrelevant. What interested us was when fields were added and when fields were removed from the user. We therefore manually changed the heatmap colors of the RankFlow visualization to colors indicating the introduction (green), continuation (grey), and removal (red) of fields (Figure 5). The appropriation of RankFlow allowed us to make the evolution of the Facebook user as a datafied entity visible, so that we could analyze how public controversies affected the platform and its governance. In the process of reworking the RankFlow graph with our data, we ended up creating a similar yet distinct form of graphical expression that does not show the flow of search ranks but instead demonstrates the evolution of categorical variables over time.[37]

36 Bernhard Rieder, "RankFlow," 2016, accessed April 2, 2024, https://labs.polsys.net/tools/rankflow/; Bernhard Rieder, Ariadna Matamoros-Fernández, and Òscar Coromina, "From Ranking Algorithms to 'Ranking Cultures': Investigating the Modulation of Visibility in YouTube Search Results," *Convergence: The International Journal of Research into New Media Technologies* 24, no. 1 (2018): 50–68, https://doi.org/10.1177/1354856517736982.

37 Category Flow visualizations can also be created using the Python library PyCatFlow; see Marcus Burkhardt and Herbert Natta, "PyCatFlow: A Python Package for Visualizing Categorical Data Over Time," *Zenodo*, September 27, 2021, https://doi.org/10.5281/ZENODO.5531785; Marcus Burkhardt, "PyCatFlow: Visualizing Categorical Data Over Time," *Medium* (blog), October 6, 2021, accessed October 10, 2023, https://medium.com/@bumatic/pycatflow-visualizing-categorical-data-over-time-b344102bcce2.

Figure 5: Evolution of Facebook user properties 2006–2010, adapted RankFlow visualization to indicate the introduction (green), continuation (grey), and removal (red) of descriptive fields. Visualization by Van der Vlist et al.[38]

Conclusion

As we discussed at the beginning of the chapter, visualizations have become increasingly important in digital cultures and can help researchers deal with vast amounts of data. As several scholars from art history and philosophy and digital humanities have discussed long before us, images and visualizations are neither self-evident nor self-explanatory. Rather than dwelling on the shortcomings of vi-

38 For a complete view of the evolution of the Facebook user between 2006 and 2020 see Van der Vlist et al., "API Governance," 12.

sual epistemology provided by current visualization platforms and applications, we advocate engaging reflexively with the possibilities they afford for doing research in new ways. Here we take a cue from Venturini et al. to employ forms such as Visual Network Analysis for an exploratory research approach and iterative analysis.[39]

Through the lens of two distinct case studies and practical examples – the comparative analysis of film festivals and the evolution of Facebook/Meta's API ecosystem – we have shown how visualization can serve in practice as a method in digital media studies and as a genuine form of research. Our chapter emphasizes the importance of conceptualization, iterative analysis, appropriation, and comparative approaches. We explored different modes in the use of visualization for each case study in order to be able to account for the complexities of each research area, using comparative as well as scalable perspectives. In addressing different ways to deal with temporality and spatiality in our data and research questions, we highlighted the practices of visualization, i.e., of finding forms of graphical expression and analyzing them as an open-ended process. On occasion this meant going back to the datasets and adding further information and calculations. At other times, it meant adapting the tool itself rather than the data fed into it. Visualization is not an easy or direct undertaking; it is a mode of inquiry full of detours, which help to reframe the data and questions along the way.

References

Bauer, Matthias, and Christoph Ernst. *Diagrammatik: Einführung in ein Kultur- und medienwissenschaftliches Forschungsfeld*. Bielefeld: Transcript, 2010.

Burkhardt, Marcus. "PyCatFlow: Visualizing Categorical Data Over Time." *Medium* (blog), October 6, 2021. Accessed October 10, 2023. https://medium.com/@bumatic/pycatflow-visualizing-categorical-data-over-time-b344102bcce2.

Burkhardt, Marcus, and Herbert Natta. "PyCatFlow: A Python Package for Visualizing Categorical Data Over Time." *Zenodo*, September 27, 2021. https://doi.org/10.5281/zenodo.5531785.

De Valck, Marijke. *Film Festivals: From European Geopolitics to Global Cinephilia*. Amsterdam: Amsterdam University Press, 2007.

Drucker, Johanna. *Visualization and Interpretation: Humanistic Approaches to Display*. Cambridge, MA: MIT Press, 2020.

Drucker, Johanna. *Graphesis: Visual Forms of Knowledge Production*. Cambridge, MA: Harvard University Press, 2014.

39 Venturini, Jacomy, and Jensen, "What Do We See."

Elsaesser, Thomas. “Film Festival Networks: The New Topographies of Cinema in Europe.” In *European Cinema: Face to Face with Hollywood*, 82–107. Amsterdam: Amsterdam University Press, 2005.

Fry, Ben. “Computational Information Design.” PhD diss., Massachusetts Institute of Technology, 2004.

Gombrich, Ernst H. *Art and Illusion: A Study in the Psychology of Pictorial Representation*. London: Phaidon, 1984.

Gombrich, Ernst H. *The Image and the Eye: Further Studies in the Psychology of Pictorial Representation*. Ithaca, NY: Cornell University Press, 1982.

Loist, Skadi. “Studying Film Circulation: Moving Film Festival Research to an Evidence-Based, Global Perspective.” In *Shaping Film Festivals in a Changing World: Practice and Methods*, edited by Dorota Ostrowska and Tamara Falicov, 211–227. Amsterdam: Amsterdam University Press, 2025.

Loist, Skadi. “Zirkulation im Netzwerk: Eine Betrachtung zur Zirkulationskraft von Filmfestivals.” *Zeitschrift für Medienwissenschaft* 12, no. 2 (2020): 55–63. http://dx.doi.org/10.25969/mediarep/14833.

Loist, Skadi, and Evgenia Samoilova. “Visualization of Festival Run Sequence.” *figshare*, 2022. https://doi.org/10.6084/m9.figshare.24588954.v1.

Loist, Skadi, Evgenia Samoilova, Martha Emilie Ehrich, and Katharina Burgdorf. “New Approaches to Studying Film Circulation: Network Analysis in the Festival Ecosystem.” Paper presented at Festivals et dynamiques cinématographiques trans(nationales): formes de production, de circulation et de représentation, Toulouse, March 28–31, 2022. *Zenodo*. https://doi.org/10.5281/zenodo.10004780.

Loist, Skadi, and Zhenya Samoilova. “How to Capture the Festival Network: Reflections on the Film Circulation Dataset.” *NECSUS: European Journal of Media Studies* 12, no. 1 (2023): 363–390. http://dx.doi.org/10.25969/mediarep/19615.

Loist, Skadi, and Zhenya Samoilova. “Digitale Methoden und Open Media Studies: Zur Erforschung von Filmfestivalruns.” *Open Media Studies Blog*, May 8, 2019. https://doi.org/10.5281/zenodo.7934483.

Mersch, Dieter. “Gaze and Withdrawal: On the ‘Logic’ of Iconic Structures.” In *Technology and Desire: The Transgressive Art of Moving Images*, edited by Rania Gaafar and Martin Schulz, 341–355. Bristol: Intellect, 2014.

Mueller, Martin. “Scalable Reading.” *Scalable Reading* (blog), 2020. Accessed October 14, 2023. https://sites.northwestern.edu/scalablereading/2020/04/26/scalable-reading/.

Ostasieninstitut. “Ein Bild sagt mehr als tausend Worte. 一画胜千言。.” Ostasieninstitut. Accessed October 14, 2023. https://ostasieninstitut.com/bibliothek/sprichwoerter-ostasiens/ein-bild-sagt-mehr-als-tausend-worte-一画胜千言。-yi-hua-sheng-qian-yan/.

Pause, Johannes, and Niels-Oliver Walkowski. “Scalable Viewing.” *Open Media Studies Blog*, 2019. Accessed October 14, 2023. https://zfmedienwissenschaft.de/online/scalable-viewing.

Rieder, Bernhard, Ariadna Matamoros-Fernández, and Òscar Coromina. “From Ranking Algorithms to ‘Ranking Cultures’: Investigating the Modulation of Visibility in YouTube Search Results.” *Convergence: The International Journal of Research into New Media Technologies* 24, no. 1 (2018): 50–68. https://doi.org/10.1177/1354856517736982.

Rieder, Bernhard. “RankFlow,” 2016. Accessed April 2, 2024. https://labs.polsys.net/tools/rankflow/.

Stjernfelt, Frederik. *Diagrammatology: An Investigation on the Borderlines of Phenomenology, Ontology, and Semiotics*. Dordrecht: Springer, 2007.

Tukey, John W. *Exploratory Data Analysis*. Reading: Addison-Wesley, 1977.

Van der Vlist, Fernando N., Anne Helmond, Marcus Burkhardt, and Tatjana Seitz. "API Governance: The Case of Facebook's Evolution." *Social Media + Society* 8, no. 2 (2022). https://doi.org/10.1177/20563051221086228.

Venturini, Tommaso, Mathieu Jacomy, and Pablo Jensen. "What Do We See When We Look at Networks: Visual Network Analysis, Relational Ambiguity, and Force-Directed Layouts." *Big Data & Society* 8, no. 1 (2021). https://doi.org/10.1177/20539517211018488.

Verhoeven, Deb. "Visualising Data in Digital Cinema Studies: More Than Just Going Through the Motions?" *Alphaville: Journal of Film and Screen Media* 11 (2016): 92–104. https://doi.org/10.33178/alpha.11.06.

Verhoeven, Deb, Paul S. Moore, Amanda Coles, Bronwyn Coate, Vejune Zemaityte, Katarzyna Musial, Elizabeth Prommer, Michelle Mantsio, Sarah Taylor, Ben Eltham, Skadi Loist, and Alwyn Davidson. "Disciplinary Itineraries and Digital Methods: Examining the Kinomatics Collaboration Networks." *NECSUS: European Journal of Media Studies* 9, no. 2 (2020): 273–298. http://dx.doi.org/10.25969/mediarep/15320.

Julia Noordegraaf

A Scalable Perspective on Historical Cinema Cultures: Studying Movie Going in Amsterdam (1952–1972) with Digital Data and Tools

Introduction

The rapid digitization of archival holdings over the past decades has made a wealth of historical sources available to scholars in the field of media studies. At the same time, a broad variety of tools have become available as open or licensed software that can be employed to search, enrich, visualize, and analyze these sources. Newly emerging research infrastructures, such as the Media Ecology Project, Lantern, Distant Viewing Lab, or CLARIAH Media Suite, are aiming to provide sustainable access to such data and tools for the purpose of research and teaching.[1] And, in tandem, new methodologies are being developed to make sense of these historical data at various levels of analysis, including the suite of tools for cross-disciplinary research in the Media Ecology Project;[2] the use of mapping tools for the analysis of film distribution and consumption;[3] the Visual Annotation Tool to support the qualitative analysis of film color;[4] the "cinemetrics" approach to analyze shot duration as a marker of film style;[5] and the "distant viewing" framework for the automated, quantitative analysis of the content and style of films and television series.[6]

1 See https://mediaecology.dartmouth.edu/wp/; https://lantern.mediahist.org/; https://www.distantviewing.org/; https://mediasuite.clariah.nl/, accessed November 12, 2023.

2 Mark Williams and John Bell, "The Media Ecology Project: Collaborative DH Synergies to Produce New Research in Visual Culture History," *Digital Humanities Quarterly* 15, no. 1 (2021), accessed November 12, 2023, https://www.digitalhumanities.org/dhq/vol/15/1/000524/000524.html.

3 Laura Horak, "Using Digital Maps to Investigate Cinema History," in *The Arclight Guidebook to Media History and the Digital Humanities*, ed. Charles R. Acland and Eric Hoyt (Falmer: REFRAME Books, 2016), 65–102.

4 Gaudenz Halter et al., "VIAN: A Visual Annotation Tool for Film Analysis," *Computer Graphics Forum* 38, no. 3 (2019): 119–129, https://doi.org/10.1111/cgf.13676.

5 Yuri Tsivian, "Cinemetrics: Part of the Humanities' Cyberinfrastructure," in *Digital Tools in Media Studies: Analysis and Review, An Overview*, ed. Michael Ross, Manfred Grauer, and Bernd Freisleben (Bielefeld: Transcript, 2009), 93–100.

6 Taylor Arnold and Lauren Tilton, "Distant Viewing: Analyzing Large Visual Corpora," *Digital Scholarship in the Humanities* 34, issue supplement 1 (2019): i3–i16, https://doi.org/10.1093/llc/

https://doi.org/10.1515/9783111082486-016

While generating exciting new opportunities for researching the history of film, radio, television, and online video and audio at an unprecedented scale and level of complexity, such digital media historical research also presents new challenges. In particular, the process of turning sources into data and processing them with various types of software raises questions about how to account for such transformations in the analysis of findings. As such, digital research demands a new kind of literacy from media scholars, who have to evaluate what historical "truths" emerge, and to what extent these are shaped by digital transformation and processing. In this chapter, I investigate the use of digital data and tools in film historical research, evaluating the opportunities and challenges. In particular, I investigate the epistemological and methodological implications of data-driven media historiography by asking what knowledge it brings and how scholars may negotiate the methodological challenges. I apply a "scalable research framework," which outlines how digital data and tools can be integrated in a research workflow that alternates between the macro level of identifying patterns in large datasets, across space and through time, and the micro level of one particular movie.

In order to explore the opportunities and challenges of digital film historiography in research practice, my contribution focuses on a central case study: the programming of Dutch fiction films in Amsterdam cinemas from 1952 until 1972.[7] The choice of period was partly motivated by the available data, which cover the period 1948–1995 (as an addition to the already available programming data for Dutch cinemas until 1947 included in Cinema Context[8]) and for which we produced clean programming data for the sample years 1952, 1962, and 1972 (in parallel with the sample years for which data are available on the programming of cinemas in Belgium[9]). Historically, this period includes the peak of cinema-going in the postwar period and its gradual decline from the 1960s. The year 1972 marks the end of the model of the single screen cinema, soon followed by multiplex cinemas that entailed new dynamics of programming and movie-going. Whereas the first decades of Dutch film and cinema history have been extensively studied,[10]

fqz013; Taylor Arnold and Lauren Tilton, *Distant Viewing: Computational Exploration of Digital Images* (Cambridge, MA: MIT Press, 2023).

7 This case study is part of the research for my current book project on digital methods for media historiography.

8 See https://cinemacontext.nl/, accessed November 12, 2023.

9 Vincent Ducatteeuw et al., "Critical Reflections on Cinema Belgica: The Database for New Cinema History in Belgium," *Journal of Open Humanities Data* 9, no. 1 (2023), https://doi.org/10.5334/johd.91.

10 Karel Dibbets and Frank van der Maden, *Geschiedenis van de Nederlandse Film en Bioscoop tot 1940* (Houten: Wereldvenster, 1986); Clara Pafort-Overduin, "Hollandse films met een Hol-

the period after WWII has been less well researched, particularly regarding the distribution and exhibition of Dutch films. The case study I am discussing therefore focused on the programming of Dutch fiction films in Amsterdam cinemas in the first decades after WWII in the context of the overall programming, on which data had become available due to the automatic extraction of programming data from the film listings in digitized Dutch newspapers.[11] The analysis focused on discovering geographical, thematic, and longitudinal patterns in the programming of the films in Amsterdam cinemas to investigate the relative share and topics of "national" content presented to local audiences. Following the trajectory of extracting film listings from digitized newspapers via the visualization and analysis of this data with Python notebooks and Gephi network analysis software, up until the interpretation and contextualization of the results, my contribution in the present chapter provides a methodological reflection on the practice of doing digital film historiography and outlines a proposed framework for "scalable film historical research."

Over the past decades, film scholars have critically examined the concept of national cinema.[12] As Marie Cronqvist and Christoph Hilgert have pointed out, media histories are not confined to national borders or media specificity; they plead for a more integrative approach to "entangled media histories."[13] In line with this, my case study is considered in the broader socio-cultural context of movie going, extending the analysis beyond the medium specific into the broader historical context of exhibition and reception. As the following analysis will demonstrate, understanding the programming of Dutch fiction films in Amsterdam in-

lands hart. Nationale identiteit en de Jordaanfilms 1934–1936" (PhD diss., Utrecht University, 2012), http://dspace.library.uu.nl/handle/1874/256372; Judith Thissen, "Understanding Dutch Film Culture: A Comparative Approach," *Alphaville: Journal of Film and Screen Media* 6 (2013): 1–14, accessed February 24, 2023, https://www.alphavillejournal.com/Issue6/HTML/ArticleThissen.html.

11 This was done at the University of Amsterdam's CREATE research program and lab in the context of the DIGIFIL project, which received support from the CLARIAH CORE project. See https://www.clariah.nl/nl/projecten/digifil-digital-film-listings, accessed November 12, 2023.

12 Tim Bergfelder, "National, Transnational or Supranational Cinema? Rethinking European Film Studies," *Media, Culture & Society* 27, no. 3 (2005): 315–331, https://doi.org/10.1177/0163443705051746; Thomas Elsaesser, "ImpersoNations: National Cinema, Historical Imaginaries and New Cinema Europe," *Mise Au Point. Cahiers de l'association Française Des Enseignants et Chercheurs En Cinéma et Audiovisuel* 5 (2013), https://doi.org/10.4000/map.1480; Andrew Higson, "The Concept of National Cinema," *Screen* 30, no. 4 (1989): 36–47, https://doi.org/10.1093/screen/30.4.36; Jerry White, "National Belonging: Renewing the Concept of National Cinema for a Global Culture," *New Review of Film and Television Studies* 2, no. 2 (2004): 211–232, https://doi.org/10.1080/1740030042000276653.

13 Marie Cronqvist and Christoph Hilgert, "Entangled Media Histories," *Media History* 23, no. 1 (2017): 130–141, https://doi.org/10.1080/13688804.2016.1270745.

vites a comparison with films from other production countries, which ensures a more transnational perspective on this local cinema culture. Finally, as the analysis will show, a data-driven approach to such a case raises questions about the definition of national cinema and thus ultimately leads to a deconstruction and rethinking of the whole idea of national film. As such, in addition to exploring the use of digital methods for film historiography, this chapter aims to evaluate how a data-driven research approach may shed new light on the intricacies of researching historical transnational media cultures.

Conceptual and Methodological Frameworks for Digital (Media) Historical Research

The focus on the ways in which digital technologies impact historical research is especially relevant for the field of media studies. As Jentery Sayers argues, it is important to acknowledge that all technologies have values embedded in them that influence how we interpret their output: "practitioners should be cognizant of not only the values and histories embedded in technologies, but also how those values and histories shape interpretation."[14] Technologies "are intricately interlaced with labor and knowledge production in and beyond the academy,"[15] and it is, in fact, hard to separate our analysis from them, as "we are entangled with the media we produce and research."[16] For example, Adrian Mackenzie in his ethnographic study of "machine learners" has shown how specific subject positions are generated in the design and operation of machine learning technologies.[17] Jasmijn Van Gorp et al. also demonstrate how the development of a digital media research infrastructure is the result of a co-development approach between scholars' needs, technological affordances, and development skills, which invites reflection on the mediating role of digital technology.[18] In fact, as Sayers indicates, "media

14 Jentery Sayers, "Introduction: Studying Media Through New Media," in *Routledge Companion to Media Studies and Digital Humanities*, ed. Jentery Sayers (London: Routledge, 2018), 1.

15 Sayers, "Introduction: Studying Media Through New Media," 2.

16 Sayers, "Introduction: Studying Media Through New Media," 2.

17 Adrian Mackenzie, *Machine Learners: Archaeology of a Data Practice* (Cambridge, MA: MIT Press, 2017).

18 Jasmijn Van Gorp, Liliana Melgar Estrada, and Julia Noordegraaf, "Involving Users in Infrastructure Development: Methodological Reflections from the Research Pilot Projects Using the CLARIAH Media Suite," *TMG Journal for Media History* 24, nos. 1–2 (2021): 1–10, https://doi.org/10.18146/tmg.809.

studies and digital humanities work through new media as means and modes of inquiry."[19]

Based on my case study, the present chapter focuses on the practice of doing film historical research with digital data and tools in order to understand the various processes of cultural transcoding that are taking place, and that impact what we can know about historical cinema cultures. In this way, it hopes to contribute to the literacy that working with digital data and tools for media historiography requires. Over recent years, scholars have started developing conceptual frameworks for grasping the impact of digital technology on the access to and use of data in historical research. After outlining two such frameworks, I will in the subsequent section apply them to the case study to demonstrate how they can be used in the context of a digital cinema research project.

In their paper "Data scopes for digital historical research," Rik Hoekstra and Marijn Koolen present the concept of the "data scope" to reflect on and describe the various steps that researchers take in a data-driven research workflow, and the impact these steps have on research findings and their interpretation.[20] As they indicate, when using digital collections, historians interact with their data in an iterative fashion, enriching and enlarging them over the course of the research process. Each step of selection, enrichment, and classification involves choices based on the exploration and interpretation of the data.[21] Data interaction should be seen as an integral part of doing digital research and the creation of data scopes is a crucial part of the digital source criticism that is required to make all forms of digital historiography transparent and accountable.[22]

A data scope "represents the process through which different views on research data are created that are relevant to a specific research question."[23] Data scoping as a research method requires that the researcher answers a set of questions. These include the question of how the dataset itself, as well as the subsets obtained through queries, relates to the original source data; the question of how the transformations observed can be taken into account when analyzing findings in the light of both the research question and the source material; and the question of how to describe the data scope, so that others can retrace the steps taken by the

19 Sayers, "Introduction: Studying Media Through New Media," 2.

20 Rik Hoekstra and Marijn Koolen, "Data Scopes for Digital History Research," *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 52, no. 2 (2019): 79–94, https://doi.org/10.1080/01615440.2018.1484676.

21 Hoekstra and Koolen, "Data Scopes for Digital History Research," 79.

22 Hoekstra and Koolen, "Data Scopes for Digital History Research," 92–93.

23 Hoekstra and Koolen, "Data Scopes for Digital History Research," 80.

researcher.[24] In collecting the answers to these questions, the data scope becomes the place that links "different resources that were compiled with different purposes" and describes how these data have been extended and contextualized.[25]

The second framework for analyzing the impact of digitization on media historical research data was developed in a collaboration among interdisciplinary researchers at the University of Wisconsin-Madison and Concordia University. There, Eric Hoyt, Charles Acland, and their collaborators have built a tool (Arclight) and a critical framework for what they call "Scaled Entity Search" (or SES): the search for and analysis of specific entities in the contents of the Media History Digital Library.[26] They used the knowledge obtained in building and using the Arclight tool to develop a framework that alerts us to elements in a digital workflow that influence the knowledge we can obtain from the data on the phenomenon studied.[27] SES is thus both a method for performing searches and a critical framework for interpreting their results.[28] It is this framework that I will focus on here, as it provides a useful conceptual tool for assessing the impact of digitization on the research process and the interpretation of the findings. The SES framework was developed specifically for search, but its components are also relevant for critically reflecting on the impact of working with film programming data extracted from a digitized newspaper corpus.

The framework focuses on three key elements: the entities (people, objects, places, events), the corpus (the collection of sources from which the data derive), and the digital (the technological processing of the sources and data). For each, it presents a set of questions that alert users to the impact that choices made for each element have on their findings.[29] For the *entities*, the SES framework asks users to reflect on how the list of entities was developed and formed: which terms have been selected, based on which sources, and which ones have been omitted? For the *corpus*, researchers should reflect on the size and scope, the context of its creation, what it covers, and – importantly – what aspects of the topic it does *not* reflect. Concerning the *digital*, the framework invites researchers to question the origin and quality of the metadata used and the impact of the soft-

24 Hoekstra and Koolen, "Data Scopes for Digital History Research," 92.

25 Hoekstra and Koolen, "Data Scopes for Digital History Research," 81.

26 See https://projectarclight.org/; https://mediahistoryproject.org/, accessed August 25, 2023.

27 Eric Hoyt et al., "Scaled Entity Search: A Method for Media Historiography and Response to Critiques of Big Humanities Data Research," in *2014 IEEE International Conference on Big Data (Big Data)* (Washington, DC: IEEE, 2014), 51–59, https://doi.org/10.1109/BigData.2014.7004453.

28 Eric Hoyt et al., "Searching, Mining and Interpreting Media History's Big Data," in *Routledge Companion to Media Studies and Digital Humanities*, ed. Jentery Sayers (London: Routledge, 2018), 414.

29 Hoyt et al., "Searching, Mining and Interpreting Media History's Big Data," 417.

ware and settings employed to transform analogue sources into digital ones (e.g., how the quality of the transcription with Object Character Recognition software (OCR) impacts the search results). Finally, the SES framework asks how these three elements relate to one another and how in combination they impact the search results.

Corpus and Methods

How were the conceptual frameworks of data scopes and Scaled Entity Search operationalized for the research I conducted on the programming of Dutch films in local cinemas? In my case study, the entities are the film screenings in Amsterdam cinemas as advertised in dedicated listings in national and local newspapers. This choice was based on a focus on the film screening event, following Karel Dibbets, who argues that the screening (When) is what brings together all the elements of a local film culture: Who (people, companies), What (films), and Where (cinemas).[30] From the perspective of data scopes, this focus on the film screening event was the basic unit of the model. The sources for these film listings were national newspapers, which, during the research period, contained advertisements for film screenings in dedicated sections in local editions for Amsterdam, Rotterdam, The Hague, and Utrecht. These newspapers had been collected by the National Library of the Netherlands, which provided detailed information on the selection and completeness of the newspaper collection and on the selection for and processing of the corpus for digitization, including the quality of the OCR.[31] For our three sample years, the corpus was complete in its coverage.

The nature of the listings, which are highly structured, and the relatively good quality of the OCR allowed for automatic extraction of the data, although a fair amount of semi-manual cleaning was required to obtain a usable dataset.[32] The screenings as advertised in the local and national newspapers are generally a good proxy for what films historical audiences actually saw in Amsterdam in

30 Karel Dibbets, "Cinema Context and the Genes of Film History," *New Review of Film and Television Studies* 8, no. 3 (2010): 331–342, https://doi.org/10.1080/17400309.2010.499784.

31 See https://www.kb.nl/onderzoeken-vinden/bijzondere-collecties/kranten; https://www.metamorfoze.nl/boeken-kranten-en-tijdschriften/digitalisering-kranten; https://www.kb.nl/en/research-find/datasets/delpher-newspapers; https://www.delpher.nl/over-delpher/wat-zit-er-in-delpher/wat-zit-er-in-delpher/kranten#cc362, accessed August 20, 2023.

32 Ivan Kisjes et al., "DIGIFIL Final Report" (Amsterdam: University of Amsterdam/CREATE, 2020), https://doi.org/10.21942/UVA.12651683.V1.

those three sample years.[33] Together, the process of extraction and cleaning generated a comprehensive dataset documenting the phenomenon. At the same time, it should be acknowledged that the coverage of the cleaned dataset is limited to the three sample years, meaning that screenings of the same film in the periods in between may have been missed.[34] It was also restricted to screenings in the commercial circuit of cinema theatres that operated under the umbrella of the National Cinema Union (NBB). The alternative circuit of screenings that took place in the context of ideologically driven associations (e.g., Catholic, Protestant, and Communist) were not included and thus not visible in this corpus.[35] Thus the modeling choice to focus on the film screening event as the moment when audience members encounter a film with a specific local profile did not fully account for the impact of the organization of the exhibition sector.

In line with the existing approaches to digital source criticism outlined above, I will in the following analysis elaborate on the process of selecting, collecting, cleaning, linking, visualizing, analyzing, and interpreting data on the exhibition of Dutch films in the context of the overall cinema culture in Amsterdam in the period 1952–1972.

Dataset

The starting point for the extraction of the data on film screenings in Amsterdam for the three sample years was the Cinema Context database: a collection of structured data on films, cinemas, people, and companies, which allows one to reconstruct the "DNA" of Dutch film culture.[36] Upon its launch in 2006, Cinema Context contained screening data on films screened in cinemas in major Dutch cities up until 1948. In order to expand this data to the post-WWII period, the research team in the DIGIFIL project (a programmer, four senior researchers (including

33 Except in cases where titles have been wrongly advertised, as in the case of a film starring actress Lizabeth Scott that had been advertised as screened in Hollandia with only her name as the title in the *Algemeen Handelsblad* of August 1, 1952, https://resolver.kb.nl/resolve?urn=KBNRC01:000088512:mpeg21:a0035, accessed November 8, 2023.

34 How we compensated for this in the analysis is outlined in the Dataset section below.

35 See Thunnis van Oort, "Industrial Organization of Film Exhibitors in the Low Countries: Comparing the Netherlands and Belgium, 1945–1960," *Historical Journal of Film, Radio and Television* 37, no. 3 (2017): 475–498, https://doi.org/10.1080/01439685.2016.1157294.

36 Dibbets, "Cinema Context and the Genes of Film History," 336; Thunnis van Oort and Julia Noordegraaf, "The Cinema Context Database on Film Exhibition and Distribution in the Netherlands: A Critical Guide: Arts and Media," *Research Data Journal for the Humanities and Social Sciences* 5, no. 2 (2020): 91–108, https://doi.org/10.1163/24523666-00502008.

myself), and two research assistants), which was conducted between 2018 and 2020, experimented with the automatic extraction of film screening data from the film listings in the digitized newspapers in the Delpher database at the KB National Library of the Netherlands.[37]

We started by sourcing the major national newspapers for the period 1948 (the first year for which we lacked comprehensive screening data in Cinema Context) until 1995 (the last year for which newspapers were available in Delpher). This resulted in a collection of around 20,650,000 articles from *De Tijd*, *De Telegraaf*, *De Volkskrant*, *De Waarheid*, *Algemeen Handelsblad*, *Trouw*, *Het Parool*, and *NRC*. In order to be able to automatically extract the film screenings from the OCR-ed newspaper advertisements, we had to take three subsequent steps, for which a processing pipeline was created: (1) Locating the listings in newspapers: classifying the articles in the corpus as a film listing or other; (2) Parsing the listings: identifying the entities in them (film titles, names of cinemas, screening times, etc.); (3) Linking the film titles: linking extracted titles with entries in external databases such as the Internet Movie Database for identification.[38]

The result was a set of about 200 lists of cinema programs for the four largest cities of the Netherlands (Amsterdam, Rotterdam, The Hague, and Utrecht) for every year from 1948 until 1995. The evaluation of the results of the automatically extracted data against a "golden standard," a manually created dataset of Rotterdam film screenings 1951–1953, showed that the data was far from clean; the identification of film titles was especially problematic. In order to be able to use them for the analysis of the distribution of Dutch films, seen against the total number of films screened, therefore, the data had to be cleaned. For this, we devised a semi-manual process, combining Python scripts for matching extracted titles against reference databases with manual checking of the scans of the newspapers on the Delpher portal or other online sources (e.g., Wikipedia) for the film titles that were unidentified or for which the match seemed problematic.

The amount of work involved in semi-manually cleaning such a big dataset required a selection in the scope of the corpus. We decided to focus on the programming of the cinemas in Amsterdam, the capital city, for which we had programming data up to 1948 but on which academic research for the post-WWII era was still lacking. The three sample years – 1952, 1962, 1972 – were chosen to facilitate comparative research with the Cinema Belgica database, which contains similar

37 Kisjes et al., "DIGIFIL Final Report."

38 All the steps are fully described in Kisjes et al., "DIGIFIL Final Report," which also includes links to the scripts stored in https://gitlab.com/uvacreate/digifil, accessed September 22, 2023.

data on film screenings in Flanders and Brussels for the same three years.[39] In order not to miss films that premiered in the first week of January, we added the newspapers of the last week of the year prior to the sample years. The cleaning operation yielded a dataset with 8,608 lines of film screenings for the cinemas active in Amsterdam in the three sample years.[40] The resulting dataset thus severely restricted the analysis to commercially operating cinemas in the capital city and to three sample years, which may not be representative for the period nor for the country as a whole. So, in the end, this data allowed for a snapshot of a specific, local urban cinema culture.

Methods

In order to analyze the data for trends at the level of the city and the cinemas, a "ranking and counting"[41] method was used, in combination with data visualization for qualitative analysis.[42] For both we ran Python scripts on the data in a Jupyter notebook environment on the Google Colab service.[43] The workflow entailed a close dialogue between a programmer (Ivan Kisjes) and researcher (me) on the questions for which the scripts retrieved relevant subsets of the data and generated visualizations. As my own literacy is limited to a basic level of reading Python code, such a dialogue with the programmer was required to understand how the code retrieves and visualizes which subset of the data was relevant for a specific research question. The Jupyter notebook provided an excellent support for such a dialogue, as it contains both the code and a space to explain in normal language

39 See https://www.cinemabelgica.be/, accessed on September 22, 2023. See Ducatteeuw et al., "Critical Reflections on Cinema Belgica"; Julia Noordegraaf et al., "Discovering Cinema Typologies in Urban Cinema Cultures: Comparing Programming Strategies in Antwerp and Amsterdam, 1952–1972," in *The Palgrave Handbook of Comparative New Cinema Histories*, ed. Daniela Treveri Gennari, Lies Van de Vijver, and Pierluigi Ercole (Cham: Palgrave Macmillan, 2023), 239–262, https://doi.org/10.1007/978-3-031-38789-0_12.

40 The dataset and scripts are published in an open repository: https://gitlab.com/uvacreate/cinema-context/scalable-perspective-on-historical-film-cultures.

41 Dibbets, "Cinema Context and the Genes of Film History"; Julia Noordegraaf, Kathleen Lotze, and Jaap Boter, "Writing Cinema Histories with Digital Databases: The Case of Cinema Context," *TMG Journal for Media History* 21, no. 2 (2018): 106–126, https://doi.org/10.18146/2213-7653.2018.369.

42 Ben Fry, *Visualizing Data: Exploring and Explaining Data with the Processing Environment* (Sebastopol, CA: O'Reilly, 2007).

43 The notebook file is available at https://gitlab.com/uvacreate/cinema-context/scalable-perspective-on-historical-film-cultures.

which function it performs.[44] As such, it also responds to the growing demand for open scholarship, allowing others to inspect and replicate the analyses.[45]

For the meso-level analysis of the content of the films from the different countries of production, we conducted a qualitative Visual Network Analysis[46] of co-occurring plot keywords using Gephi software.[47] This again was a collaboration, with the programmer (Kisjes) proposing the idea and generating the visualizations, and the researcher (me) interpreting the results via qualitative content analysis. The micro-level analysis of the content of the films shown was conducted by myself, using qualitative interpretation of contextual information on the films and their content.

The "Glocality" of Amsterdam Cinema Culture in Numbers

This section analyses the screenings of Dutch films seen against the total of films from other parts of the world screened in Amsterdam in 1952 (35 cinemas), 1962 (38 cinemas), and 1972 (32 cinemas). The dataset contains 8,608 screenings in total, of which 230 are screenings of unidentified film titles. The latter generally relates to films that were screened as part of dedicated afternoon programming for children (e.g., "Woody Woodpecker cartoon festival" at Cineac Reguliersbreestraat in 1972) and night screening programs (e.g., "Highlights from the erotic film festival" at The Movies in 1972), which were left out of the analysis. In total, 2,991 unique known film titles were shown in the three sample years, of which 53 were Dutch films, including five co-productions (with Germany [3], Belgium [1], Brazil [1]). The 53 Dutch films were screened 285 times (amounting to 1.7% of the total

44 Mari Wigham, Liliana Melgar, and Roeland Ordelman, "Jupyter Notebooks for Generous Archive Interfaces," in *2018 IEEE International Conference on Big Data (Big Data)* (Seattle, WA: IEEE, 2018), 2766–2774, https://doi.org/10.1109/BigData.2018.8622203.

45 Bernadette M. Randles et al., "Using the Jupyter Notebook as a Tool for Open Science: An Empirical Study," in *2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL)* (Toronto, ON: IEEE, 2017) 1–2, https://doi.org/10.1109/JCDL.2017.7991618.

46 Mathias Decuypere, "Visual Network Analysis: A Qualitative Method for Researching Sociomaterial Practice," *Qualitative Methods* 20, no. 1 (2020): 73–90, https://doi.org/10.1177/1468794118816613.

47 Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy, "Gephi: An Open Source Software for Exploring and Manipulating Networks," in *Proceedings of the Third International Conference on Weblogs and Social Media* 3, no. 1 (2009): 361–362, accessed February 24, 2023, http://aaai.org/ocs/index.php/ICWSM/09/paper/view/154.

number of unique films and 3.4% of the total number of 8,378 screenings, so the few Dutch films were screened relatively often).

Overall, in those three sample years, Amsterdam cinemas screened films from 49 countries of production (see Table 1). The distribution of the films from those countries is very skewed: films produced in the United States had by far biggest share of the market (1,412 titles, 47.2% of all unique film titles screened in Amsterdam in the sample years), followed by films produced in the United Kingdom (12%), France (11.8%), Italy (10%), and Germany (6.6%), with Sweden (2.1%) and the Netherlands (1.7%) leading the "long tail" of titles from all other countries, many of which had less than ten titles screened in the sample years, and some figuring in the list only as co-producing country (see Figure 1).

Table 1: Number of unique films shown in Amsterdam cinemas in 1952, 1962, 1972, per country of production (co-productions are counted per equal share in the production); omitting films from countries with less than ten screenings in total. The totals per country do not always add up as some films have been screened in multiple years.

Country	1952	1962	1972	Total	% of total # unique films
USA	472.0	613.5	424.0	1,411.5	47.2
GBR	54.5	186.0	138.8	359.2	12.0
FRA	45.5	202.7	120.1	353.8	11.8
ITA	32.0	119.8	160.7	300.0	10.0
DEU	35.5	116.2	50.4	197.6	6.6
SWE	7.0	27.0	39.8	62.8	2.1
NLD	9.0	22.0	29.5	50.5	1.7
ESP	1.5	11.5	29.9	40.4	1.4
SUHH	3.0	12.0	16.2	29.2	1.0
JPN	0.0	6.2	22.2	28.5	1.0
AUT	5.0	17.0	1.7	23.7	0.8
DNK	7.0	8.5	7.0	21.0	0.7
MEX	2.0	7.5	5.5	14.5	0.5
BEL	1.0	3.5	4.8	9.3	0.3
YUG	0.0	1.7	7.5	9.1	0.3
CHE	2.0	2.7	1.5	6.2	0.2
CAN	1.0	0.0	5.5	6.5	0.2
HUN	1.0	1.0	4.5	5.5	0.2
GRC	0.0	3.8	4.5	7.3	0.2
POL	0.0	2.0	4.0	6.0	0.2
GHN	0.0	0.0	5.0	5.0	0.2
PHL	1.0	2.5	1.5	4.0	0.1
ARG	1.0	2.1	1.1	4.2	0.1
HKG	1.0	2.0	1.0	4.0	0.1
ROM	0.0	1.0	1.0	2.0	0.1
DZA	0.0	0.0	1.8	1.8	0.1

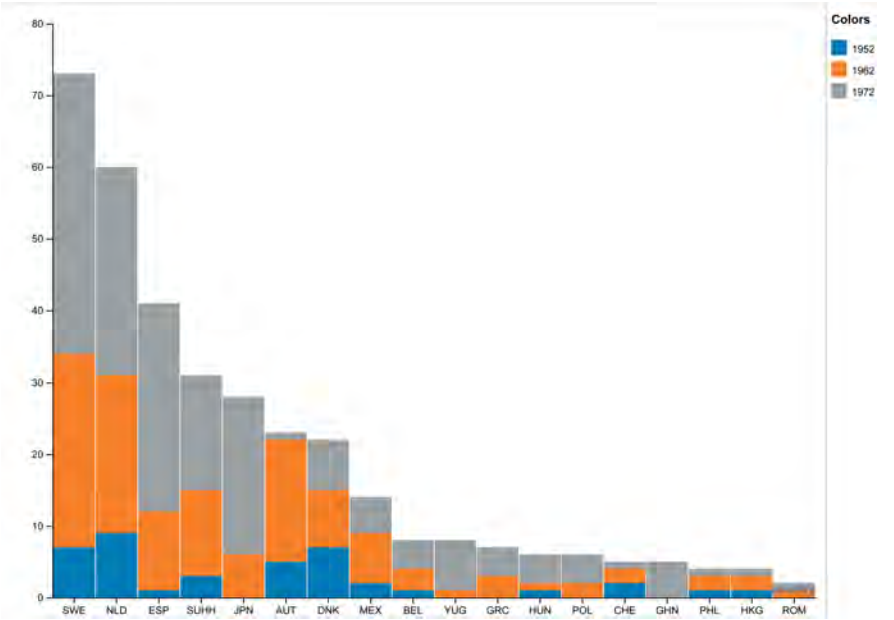


Figure 1: The "long tail" of the number of unique films per country of production (co-productions are split equally over the countries) screened in Amsterdam cinemas in 1952 (blue), 1962 (orange), and 1972 (grey), excluding the USA, GBR, FRA, ITA, DEU and countries with less than 10 screenings.

In addition to the number of unique titles for each country of production, their market share is defined by the number of screenings (Table 2). For example, there are only two Romanian films in our dataset (PUSTIUL [Elisabeta Bostan, 1962] and SERATA [Malvina Ursianu, 1972]) but these have in total been screened 10 times, giving Romania a larger market share than Canada (in total 8 films of which 3 are co-productions, so weighted as 6.5 unique titles, shown only 9 times in total).

Looking at the developments over time in the screenings, we do not observe spectacular changes in the market share of the different countries. However, we can clearly identify a decline of the share of productions from the United States, from 63.1% in 1952 to 45.8% in 1962 and 41.4% in 1972, with a clear growth in the share of films from Italy (from 5.1% in 1952 and 9.8% in 1962 to 11.9% in 1972) and a more modest growth for films from Sweden (from 1.7% in 1952 and 1.9% in 1962 to 3% in 1972) and the Netherlands (from 0.9% in 1952 and 2.9% in 1962 to 5.5% in 1972). If the top five production countries (United States, Great Britain, France, Italy, and Germany) are left out, we see little variation except for the screenings of Dutch films, which steadily increase from 9 films shown in 1952 to 22 in 1962

Table 2: Number of screenings in Amsterdam cinemas in 1952, 1962, 1972, per country of production (co-productions are counted per equal share in the production) and their share (%) of the total amount of screenings.

Country	1952	1962	1972	Total	1952 (%)	1962 (%)	1972 (%)
USA	1,132	1,648	1,237	4,017	63.1	45.8	41.4
FRA	147.5	508	281	937	8.2	14.1	9.4
GBR	168.5	419	434	1021	9.4	11.7	14.5
SWE	31	68	90	189	1.7	1.9	3.0
DEU	129.5	269	120	518	7.2	7.5	4.0
ITA	91	352	357	800	5.1	9.8	11.9
ESP	6.5	27	54	88	0.4	0.8	1.8
DNK	16	28	20	64	0.9	0.8	0.7
NLD	17	104	164	285	0.9	2.9	5.5
MEX	6	11	8	25	0.3	0.3	0.3
CHE	4	8	4	17	0.2	0.2	0.1
PHL	4	6	3	13	0.2	0.2	0.1
CAN	1	0	8	9	0.1	0.0	0.3
SUHH	16	24	39	79	0.9	0.7	1.3
AUT	15	48	4	67	0.8	1.3	0.1
BEL	1	15	17	33	0.1	0.4	0.6
ARG	1	3	3	6	0.1	0.1	0.1
HKG	2	5	3	10	0.1	0.1	0.1
HUN	5	1	19	25	0.3	0.0	0.6
YUG	0	8	16	24	0.0	0.2	0.5
BRA	0	1	2	3	0.0	0.0	0.1
IND	0	3	2	5	0.0	0.1	0.1
GRC	0	9	6	15	0.0	0.2	0.2
JPN	0	10	43	53	0.0	0.3	1.4
IRN	0	1	3	4	0.0	0.0	0.1
CZE	0	2	6	8	0.0	0.1	0.2
EGY	0	4	2	6	0.0	0.1	0.1
POL	0	4	6	10	0.0	0.1	0.2
ZAF	0	1	0	1	0.0	0.0	0.0
HRV	0	2	0	2	0.0	0.1	0.0
ROM	0	2	8	10	0.0	0.1	0.3
AUS	0	1	1	1	0.0	0.0	0.0
NOR	0	3	0	3	0.0	0.1	0.0
ISR	0	1	0	1	0.0	0.0	0.0
DZA	0	0	7	7	0.0	0.0	0.2
LUX	0	0	1	1	0.0	0.0	0.0
PRI	0	0	1	1	0.0	0.0	0.0
TUN	0	0	3	3	0.0	0.0	0.1
TUR	0	0	1	1	0.0	0.0	0.0
HTI	0	0	1	1	0.0	0.0	0.0
BOL	0	0	1	1	0.0	0.0	0.0
PRT	0	0	1	1	0.0	0.0	0.0

Table 2 (continued)

Country	1952	1962	1972	Total	1952 (%)	1962 (%)	1972 (%)
GHA	0	0	2	2	0.0	0.0	0.1
KOR	0	0	1	1	0.0	0.0	0.0
GHN	0	0	15	15	0.0	0.0	0.5
SYR	0	0	1	1	0.0	0.0	0.0
AND	0	0	0	0	0.0	0.0	0.0
IRL	0	0	1	1	0.0	0.0	0.0
Total	1,794	3,594	2,990	8,378	100	100	100

and 32 in 1972 (Table 1). Some films were shown in two of the three sample years: in particular older films from the 1930s, which were shown in both 1952 and 1962: De Jantjes (Jaap Speyer, 1934), Bleeke Bet (Richard Oswald, 1934), and Merijntje Gijzen's Jeugd (Kurt Gerron, 1936). This may be due to the lack of film production in the 1940s and the slow take-up of the industry after the war.[48]

The increase in screenings of Dutch films is also visible in the full programming data for the period 1948–1994 for Amsterdam, The Hague, Utrecht, and Rotterdam (see Figure 2). We can identify the upward trend, with some shared peaks in each city for specific years, which may indicate the release of specific titles. The increase is strongest in the period after 1970, which saw a wave of productions by Dutch filmmakers that emerged from the newly established film school.[49] We also can observe some minor variations between the cities, with Rotterdam having slightly fewer screenings of Dutch films in the period 1980–1994 than the other three cities, suggesting that Rotterdam audiences may have appreciated these national productions less. As such, the visualization is heuristically interesting and allows one to identify points for further analysis. However, the differences are very small and, as the data from this period have not been manually checked, the dataset may contain misidentified titles and thus not be fully representative. Hence, the graph below should be interpreted with care and checked with more in-depth, qualitative analysis for specific sample years.

48 In the period 1940–1959, 21 fiction films were produced in the Netherlands (an average of 1 per year), against 36 in the period 1934–1939 (an average of six per year). Source: Henk van Gelder, *Hollands Hollywood: Alle Nederlandse speelfilms van de afgelopen zestig jaar* (Amsterdam: Luitingh Sijthoff, 1995).

49 Hans Schoots, *Van Fanfare Tot Spetters (1956–1980)* (Amsterdam: Bas Lubberhuizen, 2004).

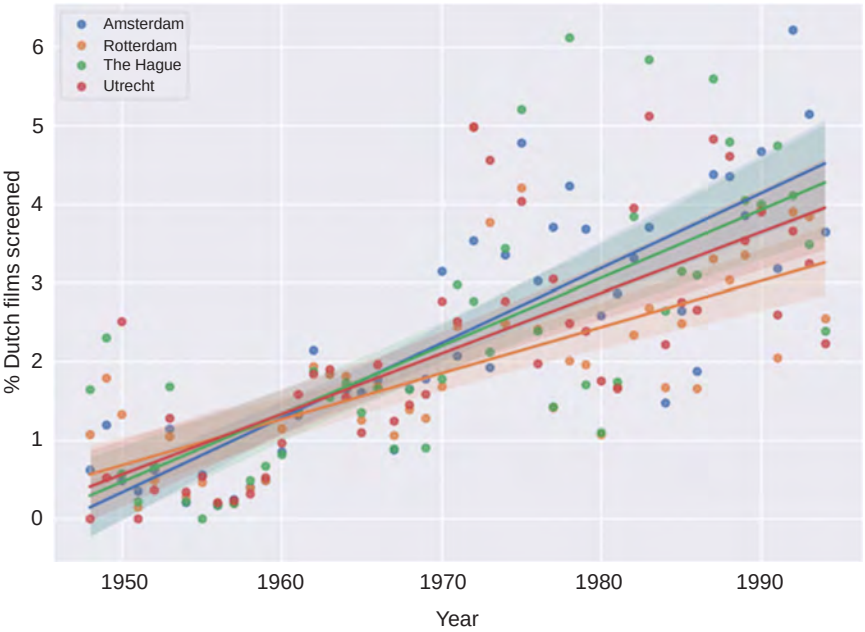


Figure 2: Percentage of screenings of Dutch films between 1948–1994 in Amsterdam (blue), The Hague (green), Utrecht (red), and Rotterdam (orange). As these percentages are very low, differences between the cities are small and the trend lines should be interpreted with care.

Zooming In: Venues

From the macro-level of general statistics on the "glocality" of the Amsterdam film screenings we can zoom in to the level of individual cinemas. The share of Dutch film screenings in the programs of the Amsterdam cinemas ranges from 0% (e.g., Nöggerath in 1952) to 25% (Rialto in 1972). Pie chart visualizations of the screening data provide a sense of the share of the countries of production per theatre and allow for a quick overview of the place of Dutch films in the overall internationality of each one's profile.

In 1952, the Capitol cinema at Rozengracht in the middle of the "popular" Jordaan neighborhood, in addition to a majority of Hollywood films, had a relatively large percentage of Dutch film screenings: 11.1% (see Figure 3). This so-called "neighborhood" cinema, a theatre that catered for local audiences, screened relatively older films, including the already mentioned "Jordaan" films from the 1930s. As most of these films are set in the neighborhood, it is perhaps not surprising that

they had a continued appeal to local audiences. But another neighborhood cinema, the Odeon in Zeeburg in Amsterdam East, also showed older titles such as De Jantjes, Malle Gevallen (Jaap Speyer, 1934) and Pygmalion (Ludwig Berger, 1937) in 1952 and 1962 (although there it amounts to only 3.8% of the screenings in 1952, compared to 11.1% at Capitol).

The pie chart visualizations of arthouse cinemas such as De Uitkijk and Kriterion show that these theatres are clearly compensating for the dominance of Hollywood films by showing films from European countries, in particular Germany, Great Britain, and France (see Figure 4). The arthouse theatres do not program many Dutch films, however. For example, Kriterion did not program any Dutch film in 1952 or 1962, and only two in 1972 (Woensdag [Bas van der Lecq, 1972] and Wat zien ik!? [Paul Verhoeven, 1971]), in which year De Uitkijk only showed one Dutch-Belgian co-production (Mira [Fons Rademakers, 1972]). Perhaps the Dutch films produced in the period did not meet the ambitions of these arthouse theatres to show only "quality films."[50]

The Dutch fiction film production is known to include a large share of children's films, which reached the cinemas once traveling cinema entrepreneur Henk van der Linden started to make children's films with his production company Rex films. Although often critiqued for their lack of quality and one-sided focus on adventure and mischief, his films were very popular among local audiences. For example, De nieuwe avonturen van Dik Trom [The new adventures of Dik Trom] (Henk van der Linden, 1958) was shown in at least one cinema in the Netherlands every week during a period of 28 years, reaching 1,263,250 viewers in the Netherlands and over two million viewers when including the screenings in Surinam and the Dutch Caribbean and 16mm screenings at schools and other venues.[51] In Amsterdam, children's films were shown mostly in neighborhood cinemas, often in afternoon programs alongside evening programming of regular feature films. An example is the Ambassade cinema, a small neighborhood cinema (360 seats) that programed mainstream feature films in the evening (e.g., The Seven Year Itch [Billy Wilder, 1955] with Marilyn Monroe in the week of August 16, 1962), accompanied by Dutch children's films at 2.00 pm in the afternoon (in the week

50 Richard van Bueren, *Saturday Night at the Movies. Het grote Amsterdamse bioscopenboek. Deel 2+3 E–Z* (Amsterdam: Lecuona, 1998), 176, 360.

51 Eye Filmmuseum, "Kleine kijkers, groot publiek: de kinderfilms van Rex Film," Eye Filmmuseum, accessed September 24, 2023, https://www.eyefilm.nl/nl/collectie/collecties/film/dossiers/kleine-kijkers-groot-publiek-de-kinderfilms-van-rex-film. In Amsterdam, the film was shown at Hallen and Rialto in 1962 and again at Rialto and at Victoria in 1972.

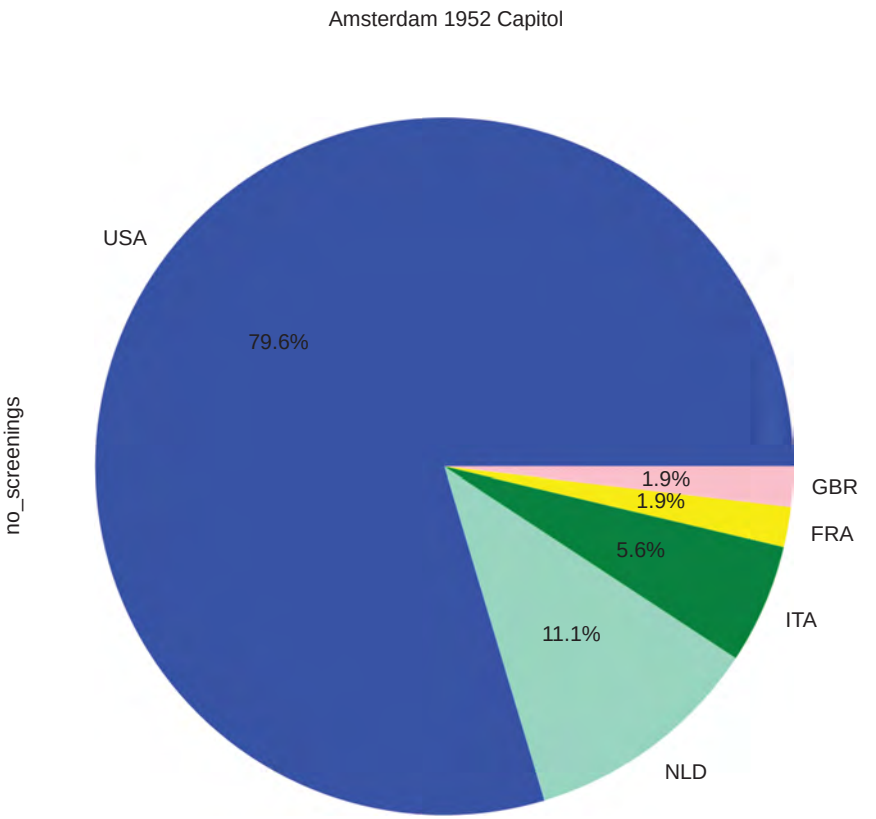


Figure 3: Pie chart visualizations showing the percentage of film screenings per country of production in Capitol and Odeon cinemas in 1952.

of August 16, 1962, Sjors van de Rebellenclub [Henk van der Linden, 1955]).[52] Such films circulated around the neighborhood cinemas owned by the Van Royen family.[53] Some theatres deliberately made "youth cinema" part of their profiles, such as the Rialto cinema at Ceintuurbaan in Amsterdam South and the Bio neighborhood cinema in Amsterdam East, which around the mid-1950s received a "guaran-

52 *De Telegraaf*, August 15, 1962, 8; retrieved from Delpher.nl, https://resolver.kb.nl/resolve?urn=ddd:011204397:mpeg21:p008, accessed September 24, 2023.

53 Noordegraaf et al., "Discovering Cinema Typologies in Urban Cinema Cultures: Comparing Programming Strategies in Antwerp and Amsterdam, 1952–1972."

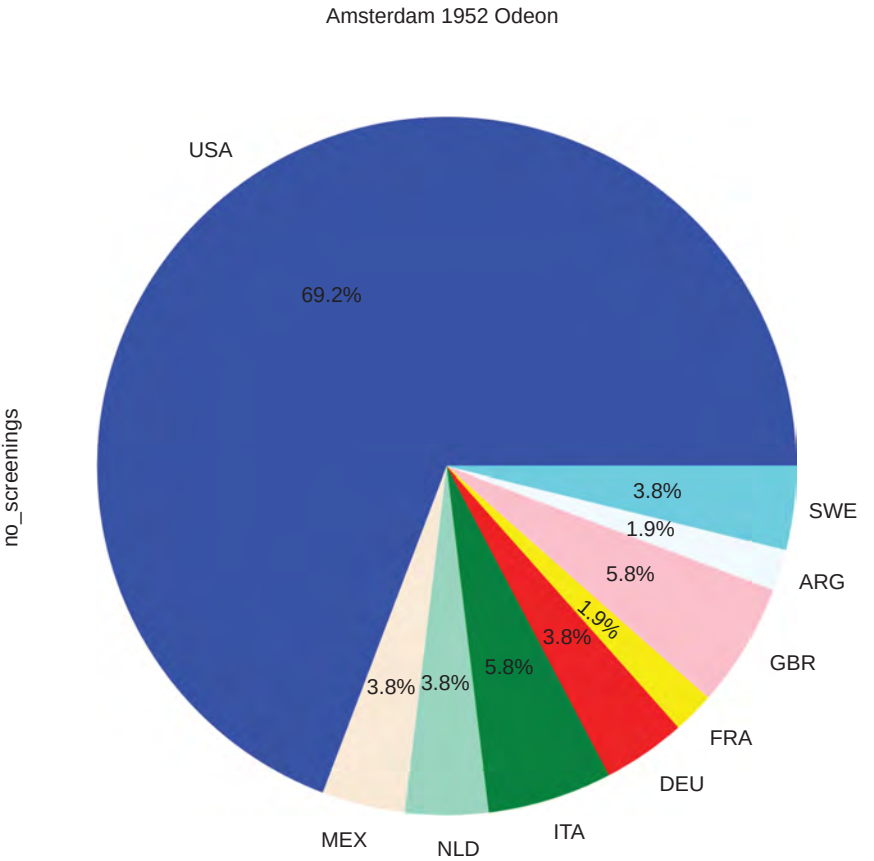


Figure 3 (continued)

tee seal" from the Amsterdam Youth Council.[54] By 1972, the share of Dutch films at Rialto had grown to 25%; these were exclusively children's films (see Figure 5).

Zooming In: Shared Themes

As becomes clear from the above discussion of the (intern)national profile of the Amsterdam cinemas, the visualization of the programming data invites zooming in

54 Richard van Bueren, *Saturday Night at the Movies. Het grote Amsterdamse bioscopenboek. Eerste Deel A–D* (Oss: NCAD uitgeverij, 1996), 46.

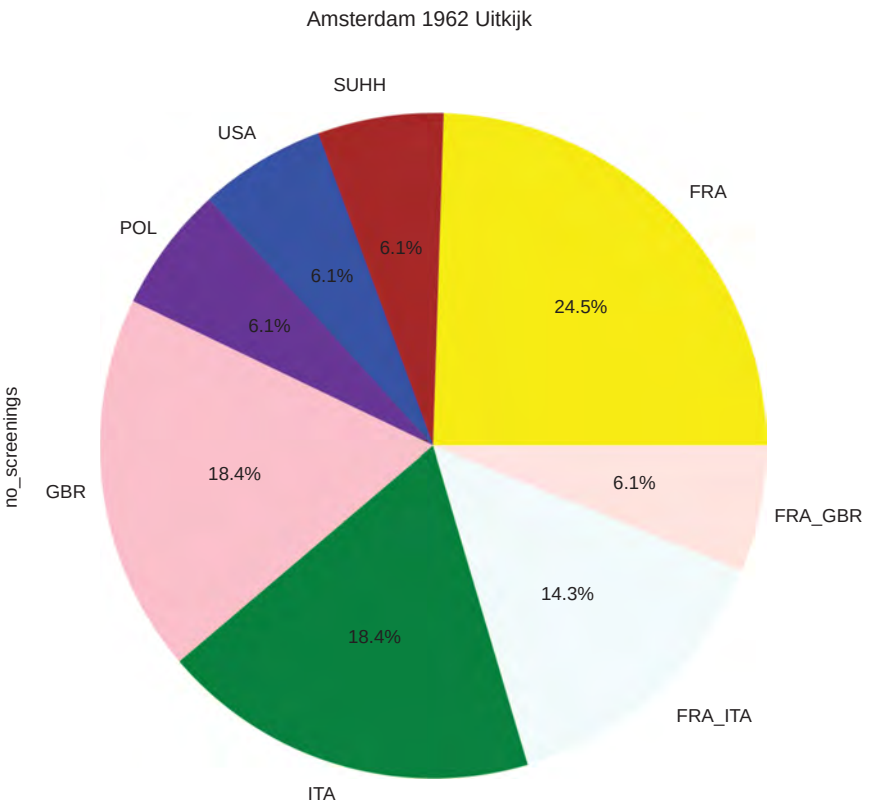


Figure 4: Pie chart visualizations showing the percentage of film screenings per country of production in De Uitkijk cinema in 1962 and 1972.

further to get an impression of the kind of films that make up this profile. Before analyzing the individual film titles in the dataset, we looked at the plot keywords assigned to the films in our dataset in the Internet Movie Database.[55] For this analysis we grouped all films from the same production country that were shown in Amsterdam in any of the sample years (co-productions occur once for each country of production). With the help of the network visualization tool Gephi,[56] we visualized these keywords in a network graph per country of production, where the plot keywords are the nodes, which are linked to each other via keywords shared between

55 This analysis was conducted by Ivan Kisjes at UvA-CREATE. We scraped the plot keywords from the IMDb site and deleted plot keywords that appear in less than three of the films in our dataset.
56 Bastian, Heymann, and Jacomy, "Gephi: An Open Source Software for Exploring and Manipulating Networks."

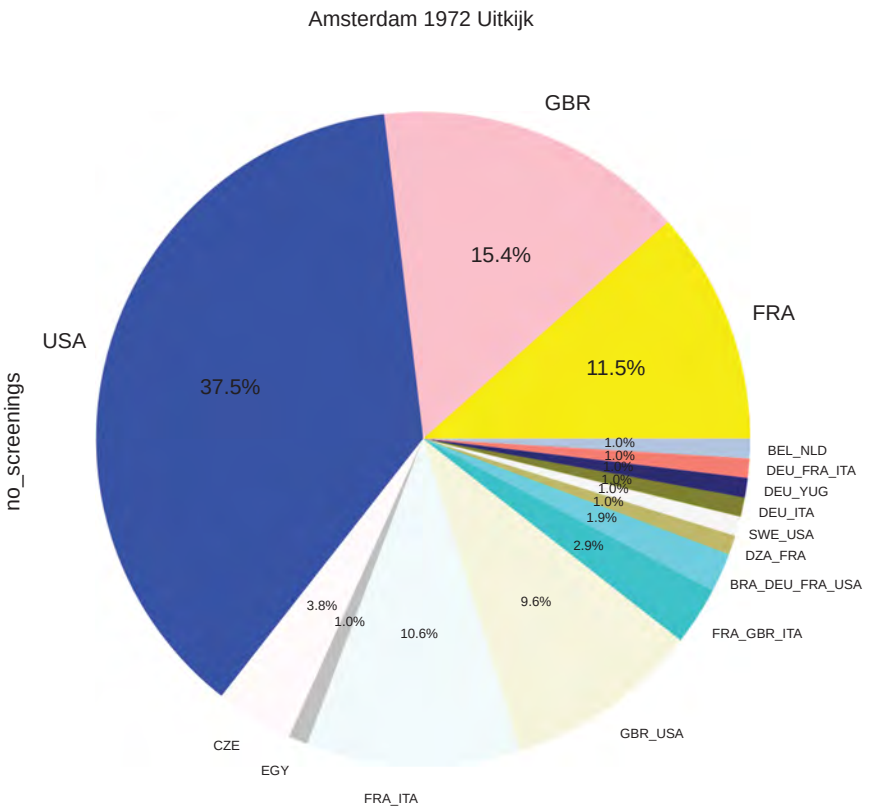


Figure 4 (continued)

the films (edges). The resulting visualization provides a hint of the themes or genres that the films from a specific country of production share, whereby similar keywords shared by various films form clusters with some density, distinguished from each other by different colors, and a larger font indicates keywords shared by more films. The further apart the clusters are positioned, the less overlap there is in the keywords that describe the films in the clusters.

For example, the visualization of the 29 Japanese films of the in total 32 shown in Amsterdam in the sample years for which plot keywords are available on IMDb, shows six large and three smaller separate clusters (Figure 6). The keywords in each cluster give an indication of the thematic of the films they describe. The orange cluster contains "epic" films involving "horseback riding," "campfires," and "forbidden love." It is closely related to the large pink cluster, which at the center appears to reference Samurai martial arts films, with keywords such as "samurai," "battle," "sword," "stabbed with a spear," and "death." On the right-hand side, the pink clus-

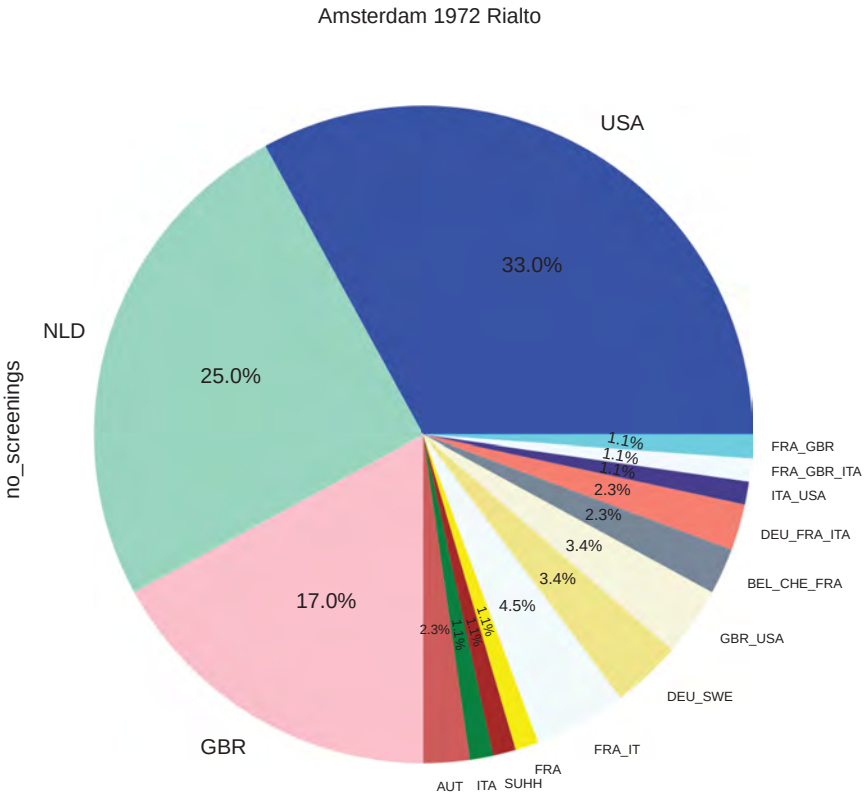


Figure 5: Pie chart visualization showing the percentage of film screenings per country of production at Rialto in 1972.

ter literally references the films from the "Japanese new wave," including "surrealism," "slow motion scene," and naked or undressing characters. The turquois cluster at the top links to the Japanese new wave films but specifically focuses on "neo-noir" films, with "envy," "attempted suicide," and "long takes." The green cluster signals exploitation films including "giant monsters" in the genre of the "psychotronic film"; the fact that it appears separate from the other clusters signals that it concerns a genre of films to which quite distinct keywords apply. The grey cluster contains keywords related to WWII ("u boat," "us soldier"); the blue cluster to "psychological dramas" and "psychological thrillers" (including many drugs-related keywords and style elements such as "long take"); and the light blue cluster to the right Westerns ("saloon," "sheriff's office," "shootout"). The yellow cluster to the right of the green cluster appears to reference dramas set at the beach with the occurrence of broken hearts, bathing suits and a ukulele.

Figure 6: Network visualization of co-occurring plot keywords for 29 (of the in total 32) films (co-) produced in Japan that were screened in Amsterdam cinemas in 1952, 1962, and 1972. Each cluster unites films that are described in similar keywords. Visualization made in Gephi and available at https://gitlab.com/uvacreate/cinema-context/scalable-perspective-on-historical-film-cultures.

The network visualization of the plot keywords for the films produced in the Netherlands shows less well-defined clusters, indicating that the keywords used to describe them are relatively close in meaning (Figure 7). The one outlier is the orange cluster at the bottom that appears to reference war movies and thrillers ("murder," "courtroom," "world war two," "resistance"). Of the 53 Dutch films in our dataset, only 36 have plot keywords on IMDb. Of those keywords, the majority reference nudity and sex ("male nudity," "female nudity," "bondage," "female/male rear nudity," etc.). When examining the dataset, it becomes clear that this signals a clear bias in the IMDb plot keywords: close reading of the actual list of titles reveals that the majority of films screened in Amsterdam theatres in the three sample years were pre-WWII classics or children's films; the plot keywords in the visualization are attributed to the six explicitly sexual films in the corpus. Clearly, the films with sexual content attract most attention by the users of IMDb but that does not provide a reliable picture of what Amsterdam audiences saw of their national film production. This demonstrates the need to conduct proper source and data criticism and to always combine close and distant methods of analysis.

Conclusion and Discussion

What new knowledge may be obtained on the distribution of Dutch films, seen in the context of the overall programming? The scalable analysis of the films screened in Amsterdam presented above shows that Dutch films, although small in number compared to the "big" film countries, were consistently shown to local audiences. Of course, with under 2% of unique titles and just over 3% of the screenings, their market share is comparatively very small, but Dutch films were on the program in every sample year and their absolute number steadily increases over time (the numbers triple between 1952 and 1972). It is striking that the first successful sound films from the 1930s were shown in 1952 and still shown in 1962, indicating that they remained popular for a long time.[57] As discussed, this may partly be explained by the fact that during and after WWII, film production in the Netherlands came almost to a standstill; of the Dutch films screened in 1952 and 1962, only one was made after the war: Een Koninkrijk voor een huis (A Kingdom for a House) (Jaap Speyer, 1949). This had changed by 1972 as a result of the new wave of film makers that had graduated from the newly established film academy.

57 Clara Pafort-Overduin, "Distribution and Exhibition in The Netherlands, 1934–1936," in *Explorations in New Cinema History*, ed. Richard Maltby, Daniel Biltereyst, and Philippe Meers (Hoboken, NJ: Wiley-Blackwell, 2011), 125–139.

Figure 7: Network visualization of co-occurring plot keywords for 36 of the in total 53 films (co-)produced in the Netherlands that were screened in Amsterdam cinemas in 1952, 1962, and 1972. Each cluster unites films that are described in similar keywords. Visualization made in Gephi and available at https://gitlab.com/uvacreate/cinema-context/scalable-perspective-on-historical-film-cultures.

It is also clear that the largest share was made up by films focused on children and families; overall 29 (55%) of the 53 Dutch films screened in the sample years were children's films (in 1972 this applied to 18 out of the 32 films), primarily shown at neighborhood cinemas and self-declared "youth cinemas." This confirms the conventional wisdom that the Dutch are good at making documentaries and children's films and less so at making fiction.[58] The attempt to analyze the content of the films screened via the IMDb plot keyword co-analysis demonstrates the need to carefully assess the source and scope of the data; in the case of the Dutch films, the keywords were heavily skewed to the six sexually explicit films (only 11% of the total). This demonstrates the nature of IMDb as a user-generated database, which does not fully adopt the standards of completeness and transparency of institutional repositories such as the National Library of the Netherlands, which houses the Dutch historical newspaper collection.

When zooming out, the analysis shows that there are differences between cinemas in their profiles; in particular, arthouse cinemas countered the U.S. dominance in the other cinemas.[59] Over time, however, the data for all the cinemas show a decline of the share of Hollywood films, with more room for films produced in European countries (in particular from the United Kingdom, Germany, France, and Italy). Overall, films produced in other parts of the world had a very small market share; Amsterdam audiences were not widely exposed to films produced outside the United States and Europe.

At the same time, the use of country of production as an indicator of the "locality" of the films is quite limited. For example, a film such as WATERLOO (Sergey Bondarchuk, 1970) is listed as an Italian-Russian co-production (involving Dino de Laurentiis and Mosfilm), but, while having a Russian director, a Russian co-producer, and Ukraine as a shooting location, the story is set in present-day Belgium and is in English. New approaches, based on the Linked Open Data approach, provide promising avenues to approach the locality of films in a more nuanced way: connecting the screening data to Wikidata allows for the inclusion of narrative location (the country or region in which the film is set), the filming location, the country of birth and nationality of all known cast members, and the language(s) spoken in the film.[60]

58 Peter Verstraten, *Humour and Irony in Dutch Post-War Fiction Film* (Amsterdam: Amsterdam University Press, 2016), 13–14.

59 The same applies to the sex cinemas Parisien and Centraal, see Noordegraaf et al., "Discovering Cinema Typologies in Urban Cinema Cultures: Comparing Programming Strategies in Antwerp and Amsterdam, 1952–1972."

60 For a first attempt, see Julia Noordegraaf et al., "Cinema Context HoMER 2023," University of Amsterdam Library Linked Open Data, July 2023, accessed November 12, 2023, https://lod.uba.uva.nl/Cinema-Context/-/stories/cinema-context-homer-2023.

To what extent then do digital data and "scalable" research methods contribute to media historiography? On the one hand, a data-driven approach entails a renewed focus on national film culture. This is so, first, because easily extractable data are mostly available for large cities in countries with a high level of digitization, such as the Netherlands. In this sense, a data-driven approach reinstates the focus on the center at the expense of the periphery (the [capital] city versus the province; a Western European country versus the rest of the world). Secondly, the available datasets for film history also reduce the complexity of film as an artistic medium to the level of clearly demarcated data points; e.g., using the label of country of production as a marker of locality of the films. In this sense, the data used in this analysis, such as the country of production or plot keywords in IMDb, are clearly "captured" in the sense that Johanna Drucker describes,[61] and should be approached with care, as the bias in the plot keywords for the Dutch films demonstrates.

My analysis shows how difficult it is to define and capture a national cinema culture with digital data and tools. The quantitative approach, here primarily a matter of ranking and counting structured data on the films screened in Amsterdam cinemas in the three sample years, provides a macro-level image of how Dutch national film production reached local audiences in comparison to the films from other countries of production and how this trended over time. The plot keywords analysis generates a very rough indication of national film production which, as the Dutch case shows, is heavily biased towards specific genres, and as such can only be used heuristically, as an invitation to zoom in to the level of the actual films shown, which can then be analyzed qualitatively. The Linked Data analysis showed avenues for further research that complicate the idea of national cinema by taking into account more fine-grained aspects of locality, such as the nationality of cast members, the language spoken, and the narrative and filming locations. And even then the need for proper source criticism in the traditional historiographical sense remains a requirement, as user-generated content (IMDb, Wikidata) is never complete nor flawless.

A scalable research framework should allow researchers to navigate between different levels of analysis, from the detailed level of the scanned source to the visualization of trends in the data in graphs and charts. At the same time, it should be transparent in providing information on the origin and processing of the data used for such visualizations. It should function as a heuristic tool, identifying areas to explore via in-depth, qualitative research. The Arclight tool developed by Hoyt et al. is based on such an approach. Arclight presents graphs that show how entities

61 Johanna Drucker, "Humanities Approaches to Graphical Display," *Digital Humanities Quarterly* 5, no. 1 (2011), http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html.

trend in the overall Media History Digital Library corpus, over time and in combination with other entities, while its connection with the search environment Lantern allows users to see results lists with snippets of the texts and the option to view the scan of the original source page. As such, Arclight facilitates a research process that "combines abstraction and granularity; users can read fine details while situating them within the larger corpus and in relation to other entities."[62] In this sense, Arclight has incorporated the need of most humanities scholars to be able to do some form of reading of the data, "zooming in and out of details [. . .] alternating between distant reading and close reading."[63] As I have shown in this chapter, such a scalable research process is essential for understanding local historical cinema cultures in a digital workflow.

References

Arnold, Taylor, and Lauren Tilton. "Distant Viewing: Analyzing Large Visual Corpora." *Digital Scholarship in the Humanities* 34, issue supplement 1 (2019): i3–i16. https://doi.org/10.1093/llc/fqz013.

Arnold, Taylor, and Lauren Tilton. *Distant Viewing: Computational Exploration of Digital Images*. Cambridge, MA: MIT Press, 2023.

Bastian, Mathieu, Sebastien Heymann, and Mathieu Jacomy. "Gephi: An Open Source Software for Exploring and Manipulating Networks." In *Proceedings of the Third International Conference on Weblogs and Social Media* 3, no. 1 (2009): 361–362. Accessed February 24, 2023, http://aaai.org/ocs/index.php/ICWSM/09/paper/view/154.

Bergfelder, Tim. "National, Transnational or Supranational Cinema? Rethinking European Film Studies." *Media, Culture & Society* 27, no. 3 (2005): 315–331. https://doi.org/10.1177/0163443705051746.

Bueren, Richard van. *Saturday Night at the Movies. Het grote Amsterdamse bioscopenboek. Deel 2+3 E–Z*. Amsterdam: Lecuona, 1998.

Bueren, Richard van. *Saturday Night at the Movies. Het grote Amsterdamse bioscopenboek. Eerste Deel A–D*. Oss: NCAD uitgeverij, 1996.

Cronqvist, Marie, and Christoph Hilgert. "Entangled Media Histories." *Media History* 23, no. 1 (2017): 130–141. https://doi.org/10.1080/13688804.2016.1270745.

Decuypere, Mathias. "Visual Network Analysis: A Qualitative Method for Researching Sociomaterial Practice." *Qualitative Methods* 20, no. 1 (2020): 73–90. https://doi.org/10.1177/1468794118816613.

Dibbets, Karel. "Cinema Context and the Genes of Film History." *New Review of Film and Television Studies* 8, no. 3 (2010): 331–342. https://doi.org/10.1080/17400309.2010.499784.

Dibbets, Karel, and Frank van der Maden. *Geschiedenis van de Nederlandse Film en Bioscoop tot 1940*. Houten: Wereldvenster, 1986.

62 Hoyt et al., "Searching, Mining and Interpreting Media History's Big Data," 417.

63 Hoekstra and Koolen, "Data Scopes for Digital History Research," 82.

Drucker, Johanna. “Humanities Approaches to Graphical Display.” *Digital Humanities Quarterly* 5, no. 1 (2011). Accessed February 24, 2023. http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html.

Ducatteeuw, Vincent, Daniel Biltereyst, Philippe Meers, Christophe Verbruggen, Dries Moreels, Julia Noordegraaf, Sally Chambers, Pieterjan De Potter, Tamar Cachet, Nicolas Franck, and Florian Deroo. “Critical Reflections on Cinema Belgica: The Database for New Cinema History in Belgium.” *Journal of Open Humanities Data* 9, no. 1 (2023). https://doi.org/10.5334/johd.91.

Elsaesser, Thomas. “ImpersoNations: National Cinema, Historical Imaginaries and New Cinema Europe.” *Mise Au Point. Cahiers de l’association Française Des Enseignants et Chercheurs En Cinéma et Audiovisuel* 5 (2013). https://doi.org/10.4000/map.1480.

Eye Filmmuseum. “Kleine kijkers, groot publiek: de kinderfilms van Rex Film.” Eye Filmmuseum. Accessed September 24, 2023. https://www.eyefilm.nl/nl/collectie/collecties/film/dossiers/kleine-kijkers-groot-publiek-de-kinderfilms-van-rex-film.

Fry, Ben. *Visualizing Data: Exploring and Explaining Data with the Processing Environment*. Sebastopol, CA: O’Reilly, 2007.

Gelder, Henk van. *Hollands Hollywood: Alle Nederlandse speelfilms van de afgelopen zestig jaar*. Amsterdam: Luitingh Sijthoff, 1995.

Halter, Gaudenz, Rafael Ballester-Ripoll, Barbara Flueckiger, and Renato Pajarola. “VIAN: A Visual Annotation Tool for Film Analysis.” *Computer Graphics Forum* 38, no. 3 (2019): 119–129. https://doi.org/10.1111/cgf.13676.

Higson, Andrew. “The Concept of National Cinema.” *Screen* 30, no. 4 (1989): 36–47. https://doi.org/10.1093/screen/30.4.36.

Hoekstra, Rik, and Marijn Koolen. “Data Scopes for Digital History Research.” *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 52, no. 2 (2019): 79–94. https://doi.org/10.1080/01615440.2018.1484676.

Horak, Laura. “Using Digital Maps to Investigate Cinema History.” In *The Arclight Guidebook to Media History and the Digital Humanities*, edited by Charles R. Acland and Eric Hoyt, 65–102. Falmer: REFRAME Books, 2016. Accessed February 24, 2023. https://projectarclight.org/book/.

Hoyt, Eric, Kit Hughes, Derek Long, Tony Tran, and Kevin Ponto. “Scaled Entity Search: A Method for Media Historiography and Response to Critiques of Big Humanities Data Research.” In *2014 IEEE International Conference on Big Data (Big Data)*, 51–59. Washington, DC: 2014. https://doi.org/10.1109/BigData.2014.7004453.

Hoyt, Eric, Tony Tran, Derek Long, Kit Hughes, and Kevin Ponto. “Searching, Mining and Interpreting Media History’s Big Data.” In *Routledge Companion to Media Studies and Digital Humanities*, edited by Jentery Sayers, 413–422. London: Routledge, 2018.

Kisjes, Ivan, Thunnis van Oort, Kathleen Lotze, Ieva Staliunaite, Joris Verbeek, Kaspar Beelen, and Julia Noordegraaf. “DIGIFIL Final Report.” Amsterdam: University of Amsterdam/CREATE, 2020. https://doi.org/10.21942/UVA.12651683.V1.

Mackenzie, Adrian. *Machine Learners: Archaeology of a Data Practice*. Cambridge, MA: MIT Press, 2017.

Noordegraaf, Julia, Kathleen Lotze, and Jaap Boter. “Writing Cinema Histories with Digital Databases: The Case of Cinema Context.” *TMG Journal for Media History* 21, no. 2 (2018): 106–126. https://doi.org/10.18146/2213-7653.2018.369.

Noordegraaf, Julia, Thunnis van Oort, Kathleen Lotze, Daniel Biltereyst, Philippe Meers, and Ivan Kisjes. “Discovering Cinema Typologies in Urban Cinema Cultures: Comparing Programming Strategies in Antwerp and Amsterdam, 1952–1972.” In *The Palgrave Handbook of Comparative New Cinema Histories*, edited by Daniela Treveri Gennari, Lies Van de Vijver, and Pierluigi Ercole, 239–262. Cham: Palgrave Macmillan, 2023. https://doi.org/10.1007/978-3-031-38789-0_12.

Noordegraaf, Julia, Leon Wissen, Ivan Kisjes, and Thunnis van Oort. “Cinema Context HoMER 2023.” University of Amsterdam Library Linked Open Data, July 2023. Accessed November 12, 2023. https://lod.uba.uva.nl/Cinema-Context/-/stories/cinema-context-homer-2023.

Oort, Thunnis van. “Industrial Organization of Film Exhibitors in the Low Countries: Comparing the Netherlands and Belgium, 1945–1960.” *Historical Journal of Film, Radio and Television* 37, no. 3 (2017): 475–498. https://doi.org/10.1080/01439685.2016.1157294.

Oort, Thunnis van, and Julia Noordegraaf. “The Cinema Context Database on Film Exhibition and Distribution in the Netherlands: A Critical Guide: Arts and Media.” *Research Data Journal for the Humanities and Social Sciences* 5, no. 2 (2020): 91–108. https://doi.org/10.1163/24523666-00502008.

Pafort-Overduin, Clara. “Distribution and Exhibition in The Netherlands, 1934–1936.” In *Explorations in New Cinema History*, edited by Richard Maltby, Daniel Biltereyst, and Philippe Meers, 125–139. Hoboken, NJ: Wiley-Blackwell, 2011.

Pafort-Overduin, Clara. “Hollandse films met een Hollands hart. Nationale identiteit en de Jordaanfilms 1934–1936.” PhD diss., Utrecht University, 2012. Accessed February 24, 2023. http://dspace.library.uu.nl/handle/1874/256372.

Randles, Bernadette M., Irene V. Pasquetto, Milena S. Golshan, and Christine L. Borgman. “Using the Jupyter Notebook as a Tool for Open Science: An Empirical Study.” In *2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, 1–2. Toronto, ON: IEEE, 2017. https://doi.org/10.1109/JCDL.2017.7991618.

Sayers, Jentery. “Introduction: Studying Media Through New Media.” In *Routledge Companion to Media Studies and Digital Humanities*, edited by Jentery Sayers, 1–6. London: Routledge, 2018.

Schoots, Hans. *Van Fanfare Tot Spetters (1956–1980)*. Amsterdam: Bas Lubberhuizen, 2004.

Thissen, Judith. “Understanding Dutch Film Culture: A Comparative Approach.” *Alphaville: Journal of Film and Screen Media* 6 (2013): 1–14. Accessed February 24, 2023. https://www.alphavillejournal.com/Issue6/HTML/ArticleThissen.html.

Tsivian, Yuri. “Cinemetrics: Part of the Humanities’ Cyberinfrastructure.” In *Digital Tools in Media Studies: Analysis and Review, An Overview*, edited by Michael Ross, Manfred Grauer, and Bernd Freisleben, 93–100. Bielefeld: Transcript, 2009.

Van Gorp, Jasmijn, Liliana Melgar Estrada, and Julia Noordegraaf. “Involving Users in Infrastructure Development: Methodological Reflections From the Research Pilot Projects Using the CLARIAH Media Suite.” *TMG Journal for Media History* 24, nos. 1–2 (2021): 1–10. https://doi.org/10.18146/tmg.809.

Verstraten, Peter. *Humour and Irony in Dutch Post-War Fiction Film*. Amsterdam: Amsterdam University Press, 2016.

White, Jerry. “National Belonging: Renewing the Concept of National Cinema for a Global Culture.” *New Review of Film and Television Studies* 2, no. 2 (2004): 211–232. https://doi.org/10.1080/1740030042000276653.

Wigham, Mari, Liliana Melgar, and Roeland Ordelman. “Jupyter Notebooks for Generous Archive Interfaces.” In *2018 IEEE International Conference on Big Data (Big Data)*, 2766–2774. Seattle, WA: IEEE, 2018. https://doi.org/10.1109/BigData.2018.8622203.

Williams, Mark, and John Bell. “The Media Ecology Project: Collaborative DH Synergies to Produce New Research in Visual Culture History.” *Digital Humanities Quarterly* 15, no. 1 (2021). Accessed February 24, 2023. https://www.digitalhumanities.org/dhq/vol/15/1/000524/000524.html.

Filmography

Bleeke Bet, dir. Richard Oswald. Cinetone Studios, Duivendrecht, the Netherlands, 1934.
De Jantjes, dir. Jaap Speyer. Cinetone Studios, Duivendrecht, the Netherlands, 1934.
De nieuwe avonturen van Dik Trom, dir. Henk van der Linden. Rex Film, Schinnen, the Netherlands, 1958.
Een Koninkrijk voor een huis, dir. Jaap Speyer. Monopole Film, Rotterdam/Amsterdam, the Netherlands, 1949.
Malle Gevallen, dir. Jaap Speyer. Cinetone Studios, Duivendrecht, the Netherlands, 1934.
Merijntje Gijzen's Jeugd, dir. Kurt Gerron. Filmstad, Wassenaar, the Netherlands, 1936.
Mira, dir. Fons Rademakers. Fons Rademakers Productie, Amsterdam, the Netherlands, 1972.
Pustiul, dir. Elisabeta Bostan. Studioul Cinematografic Bucuresti, Bucharest, Romania, 1962.
Pygmalion, dir. Ludwig Berger. Cinetone Studios, Duivendrecht, the Netherlands, 1937.
Serata, dir. Malvina Urşianu. Studioul Cinematografic Bucuresti, Bucharest, Romania, 1971.
Sjors van de Rebellenclub, dir. Henk van der Linden. Rex Film, Schinnen, the Netherlands, 1955.
The Seven Year Itch, dir. Billy Wilder. 20th Century Fox, Los Angeles, CA, United States, 1955.
Wat zien ik!?, dir. Paul Verhoeven. Rob Houwer Film Holland, The Hague/Amsterdam, the Netherlands, 1971.
Waterloo, dir. Sergey Bondarchuk. Dino de Laurentiis Cinematografica/Mosfilm, Rome/Moscow, Italy/Soviet Union, 1970.
Woensdag, dir. Bas van der Lecq. Hilvaria Film, Amsterdam, the Netherlands, 1972.

Matthias Grotkopp

Catastrophe or Pointillism of Disaster? Annotating and Visualizing Patterns of Ecological Imagination

Introduction

Considering the importance of audiovisual images in the realm of our contemporary digital public spheres, the relationship between the digital humanities and media studies in general, and film studies in particular, is still surprisingly unconsolidated and mobile. One of the main challenges, as compared to the available standards and commonly shared practices dealing with text, is to create access to the perceptual properties of the temporal unfolding of audiovisual images in a way that makes sense as a digital data structure. There are many projects and initiatives dealing with metadata about audiovisual objects, about persons, places and times, screenings, and instances[1] – even though here too the different archival and research institutions are still looking for common standards. There are also huge advances in computer vision-gathering data about objects and persons depicted in the image, detecting words written on the screen or on buildings and other objects, or recognizing concepts and persons mentioned in spoken dialogue. But the moving images and sounds as spatiotemporal scenarios of perception are still obstinate objects when it comes to formatting them as datasets.

The answers to this obstinacy are diverse and come from very different directions, so it certainly is necessary to multiply the tools and methods for a multitude of possible case studies. Such studies may lie in focusing on those aspects of audiovisual composition that offer themselves for quantification like shot length,[2] or for algorithmic processing like colorimetry,[3] or the use of face recognition to determine field sizes and screen times.[4] They may lie in algorithms that create

1 The other contributions in this book are ample proof of the diversity and creativity in this field.

2 Yuri Tsivian, "Cinemetrics, Part of the Humanities' Cyberinfrastructure," in *Digital Tools in Media Studies: Analysis and Research*, ed. Michael Ross, Manfred Grauer, and Bernd Freisleben (Bielefeld: Transcript, 2009), 93–100.

3 Barbara Flückiger, "A Digital Humanities Approach to Film Colors," *The Moving Image: The Journal of the Association of Moving Image Archivists* 17, no. 2 (2017): 71–94.

4 Taylor Arnold and Lauren Tilton, "Distant Viewing: Analyzing Large Visual Corpora," *Digital Scholarship in the Humanities* 34, issue supplement 1 (2019): i3–i16, https://doi.org/10.1093/llc/fqz013.

https://doi.org/10.1515/9783111082486-017

new – mostly visual but also sonic – objects out of the moving images whose contemplation or comparison can lead to analytic findings.[5] Yet there is also this specific part of our *déformation professionelle* that media scholars are quite often much more inclined to question and critique the epistemologies and blind spots of these answers than to deploy them prolifically.

In this chapter I want to reflect on a method for annotating audiovisual images that tries to retain the full spectrum of parameters present in film analysis, and tries to retain their synchronous and diachronous relationships, while still dissecting the moving images and sounds into discrete entities. One core facet of this method is the re-formatting of annotation data as human readable visualizations and, in what follows, I want to focus on three concrete use scenarios of finding and comparing patterns of audiovisual composition via patterns in visualized annotations.

Visualizing Audiovisual Structures

What I want to present in this chapter is a case study from a project that looks at fiction films, documentaries and activist films, nature series, and videos on social media platforms that deal with the climate crisis.[6] What we want to do is to describe the concrete experiential patterns used to convey both the science and the moral imperatives of the climate crisis, and how these patterns, metaphors, and affects intervene in audiovisual discourse and thus in our common everyday perception of the world and our entanglements with technologies, materialities, and other existences. One of the problems for thought and everyday perception in this context is that scientific knowledge and ordinary experience, the planetary and the local, the political now and geological deep time, are not easily mapped onto each other. How do we make sense of the ecological multi-crisis as both a

5 Kevin L. Ferguson, "Volumetric Cinema," *[in]transition: Journal of Videographic Film and Moving Image Studies* 2, no. 1 (2015); Jason Mittell, "Videographic Criticism as a Digital Humanities Method," in *Debates in the Digital Humanities 2019*, ed. Matthew K. Gold and Lauren F. Klein (Minneapolis: University of Minnesota Press, 2019), 224–242; Michael J. Kramer, "What Does a Photograph Sound Like? Digital Image Sonification as Synesthetic Audiovisual Digital Humanities," *Digital Humanities Quarterly* 15, no. 1 (2021), accessed June 30, 2023, http://digitalhumanities.org/dhq/vol/15/1/000508/000508.html.

6 "Intervening World Projections: Audiovisuality of Climate Change" project within the CRC 1512 "Intervening Arts" (Freie Universität Berlin, 2022), project members: Matthias Grotkopp, Yvonne Pfeilschifter, and Leona Schleicher, https://www.sfb-intervenierende-kuenste.de/en/teilprojekte/C/C05/index.html, accessed March 15, 2024.

dispersed, diffuse, and confusing number of local and individual events, injustices, and pollutions here and there and, at the same time, a global emergency that is more than just the sum of these events?[7] Why are we constantly oscillating between the image of one massive global catastrophe and the pointillism of many, unevenly distributed disasters that together result in the multi-crisis? The task is to comprehend this difficult nature of form, scale, and spatiotemporal pattern when imagining climate change and other ecological damage caused by the assemblage of humans, fossil fuels, technology, western capitalism, and extractivism.[8]

Visualizations come into play at this point, namely, where we try to find the gestalt of these processes in audiovisual patterns as they ground the metaphors, affects, and schemata of our understanding and feeling. What is the melting ice sheet? Not as a piece of information, an image, or a point in the terabytes of climate data, but as a force, a movement, a gesture that is realized in our embodied experience and that can be felt thousands of kilometers away from its actual occurrence.

When thinking about visualizing cinematic patterns there are of course clear historical precursors. In his essay on "Vertical Montage,"[9] Sergei Eisenstein uses a graph that shows the common gestural qualities of the visual and the acoustic, dissecting a scene from his film ALEXANDER NEVSKY (SU 1938) and making visible the interplay between the movement dynamics of Prokofiev's musical score and the shapes and movements of the image compositions. Dziga Vertov developed several methods of score-keeping in order to structure the editing process accord-

7 When I am not referring to a concrete project team, I use the pronoun "we" in order to refer to people from western industrialized societies and the interweaving of an everyday use of technologies and infrastructure based on fossil energy and a certain heterogenous but still largely coherent cultural and epistemological framework that goes along with it. This "we" is connected to, but distinguishable from a "we" comprising human and non-human, contemporary and future beings that are all impacted by the climate crisis and biodiversity crisis, yet in different degrees and temporalities.

8 The amount of literature on the topic is now so vast that a comprehensive overview is impossible; to mention just a few publications: Tom Cohen, ed., *Telemorphosis: Theory in the Era of Climate Change* (Ann Arbor: Open Humanities Press, 2012); Donna J. Haraway, *Staying with the Trouble: Making Kin in the Chthulucene* (Durham, NC; London: Duke University Press, 2016); Anna Tsing et al., eds., *Arts of Living on a Damaged Planet* (Minneapolis: University of Minnesota Press, 2017); Kathryn Yussof, *A Billion Black Anthropocenes or None* (Minneapolis: University of Minnesota Press, 2018); Eva Horn and Hannes Bergthaller, *The Anthropocene. Key Issues for the Humanities* (New York: Routledge, 2020); Dipesh Chakrabarty, *The Climate of History in a Planetary Age* (Chicago, London: University of Chicago Press, 2021).

9 Sergei Eisenstein, "Vertical Montage" [1940], in *Selected Works, Vol. 2: Towards a Theory of Montage 1937–1940*, ed. Michael Glenny and Richard Taylor (London: I. B. Tauris, 2010), 337.

ing to recurring motifs and framings.[10] And there are of course the Cinemetrics-tables of shot length and their statistical evaluation.[11]

What we wanted to do, however, had the challenge to be more inclusive regarding the parameters that figure in the visualization and, at the same time, to be much more generic in its values, not already deduced by a specific object and therefore open to all kinds of audiovisual data and all kinds of film analytical research questions. So far, the setup has been applied with satisfactory results to fiction films, documentary films and series, and TV news, as well as social media videos from YouTube and TikTok. Most of the individual and collaborative projects or publications further appropriated and developed the setup for approaches to affect and rhetoric and to modes of persuasion and dissemination of knowledge, but it can also be used for subjects like representation of race, class, and gender,[12] or areas like stylometry and genre studies. As a predominantly formal aesthetic method, however, it always has to be supplemented with other forms of enquiry if questions of power, normativity, positionality, and specificity are to be taken into account within the field of cultural studies.[13]

The method and the Ada Filmontology were initially developed[14] in a project on the financial crisis, a collaboration between film scholars and information scientists in Berlin and Potsdam.[15] They build on a previous methodological frame-

10 Cf. Adelheid Heftberger, *Digital Humanities and Film Studies: Visualizing Dziga Vertov's Work* (Cham: Springer Nature, 2018).

11 Yuri Tsivian et al., https://cinemetrics.uchicago.edu, accessed March 15, 2024. For an exemplary case study of statistical analysis, see Mike Baxter, Daria Khitrova, and Yuri Tsivian, "Exploring Cutting structure in Film, with Applications to the Films of D. W. Griffith, Mack Sennett, and Charlie Chaplin," *Digital Scholarship in the Humanities* 32, no. 1 (2017), 1–16, https://doi.org/10.1093/llc/fqv035.

12 Tamara Drummond and Janina Wildfeuer, "The Multimodal Annotation of Gender Differences in Contemporary TV Series: Combining Qualitative Questions and Quantitative Results," in *Annotations in Scholarly Editions and Research: Functions, Differentiation, Systematization*, ed. Julia Nantke and Frederik Schlupkothen (Berlin, Boston: De Gruyter, 2020), 35–58.

13 Cf. Stuart Hall, "Cultural Studies and Its Theoretical Legacies," in *Cultural Studies*, ed. Larry Grossberg, Cary Nelson, and Paula Treichler (London: Routledge 1992), 277–294.

14 An extended introduction can be found in Jan-Hendrik Bakels et al., "Matching Computational Analysis and Human Experience: Performative Arts and the Digital Humanities," *Digital Humanities Quarterly* 14, no. 4 (2020), accessed June 30, 2023, http://www.digitalhumanities.org/dhq/vol/14/4/000496/000496.html.

15 Junior research group "Audio-Visual Rhetorics of Affect" (Freie Universität Berlin, Hasso-Plattner-Institut, funded by the German Ministry of Education and Research, 2015–2021. Project Leader: Jan-Hendrik Bakels; project members: Henning Agt-Rickauer, Christian Hentschel, Thomas Scherer, Jasper Stratil; student assistants: Anton Buzal, Yvonne Pfeilschifter, João Pedro Prado, Rebecca Zorko; mentors: Harald Sack, Hermann Kappelhoff; associated members: Mat-

work for the analysis of cinematic affect and cinematic metaphor, which consists of a standardized sequence of segmentation, free text annotation, descriptive qualification, and synthesis.[16] This framework in turn stands in relation to long established uses of analogue and digital annotation in film studies' introductory courses as well as in elaborated case studies like Raymond Bellour's analysis of THE BIRDS (Alfred Hitchcock, USA 1963), in which he demonstrates the system of describing shots according to chosen parameters and then determining patterns and segmentations, from whose structural relationships (and the gaps within) meaning derives.[17]

One strategic decision that we made from the outset was that we did not want to restrict the parameters of annotation to what was considered automatable at the start of the project development or to a specific aesthetic category. Compared to the high functionality and versatile spatial display modes of colorimetric analyses and figure-ground extraction in the VIAN system,[18] our approach did not aim at a direct visualization of singular features but only their interplay via their formatting as semantic data. The other automatic forms of processing audiovisual images that were most common at the start of the project were speech recognition and diverse variations of object recognition, which can be applied for domain specific queries but do not offer direct access to questions of audiovisual composition. However, in relation to the existing taxonomy described below, future consideration could be given to adapting automatic feature recognition for film aesthetic parameters with recourse to creative proxy reasoning, using face recognition in relation to the frame as an inference for field size, for instance,[19] or using keypoint detection in order to grasp body posture and move-

thias Grotkopp, Olivier Aubert), accessed June 30, 2023, https://projectada.github.io, https://www.ada.cinepoetics.fu-berlin.de/en/ada-toolkit/index.html.

16 Thomas Scherer, Sarah Greifenstein, and Hermann Kappelhoff, "Expressive Movements in Audiovisual Media: Modulating Affective Experience," in *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction*, ed. Cornelia Müller et al., vol. 2 (Berlin/New York: De Gruyter Mouton, 2014), 2081–2092; Cornelia Müller and Hermann Kappelhoff, *Cinematic Metaphor. Experience – Affectivity – Temporality* (Berlin, Boston: De Gruyter, 2018).

17 Raymond Bellour, *The Analysis of Film* [1973] (Bloomington: Indiana University Press, 2000), 28–67.

18 Barbara Flückiger and Gaudenz Halter, "Methods and Advanced Tools for the Analysis of Film Colors in Digital Humanities," *Digital Humanities Quarterly* 14, no. 4 (2020), accessed June 30, 2023, http://www.digitalhumanities.org/dhq/vol/14/4/000500/000500.html.

19 Arnold and Tilton, "Distant Viewing." With all algorithmic tools it is of course necessary to be critical of possible biases and constrictions.

ment.[20] Another principle that we chose as starting point was to privilege a time-based and tiered approach, as opposed to other ways of creating or coding metadata.[21] Observations of duration, rhythm, and multimodal co-occurrence were to be foregrounded at every step of creating or reviewing annotations.

Based on the elementary course of procedure of segmentation, annotation, and qualification, our project has focused on analyzing audiovisual composition as multimodal, temporal patterns. In order to facilitate its future evaluation according to rule-based reasoning (and with the possibility of using interfaces for automatic and semi-automatic annotations in mind), our mode of annotation had to be considered not only from a film studies perspective but from the point of view of data sciences as well. This also meant that, even though we always left room for unstructured, free-speech annotation, the backbone of annotations had to come from a shared formal taxonomy. Therefore, we devised the following setup:

First, we developed a structured film analytical vocabulary, a glossary defining the basic terms in English and in German. This may sound trivial, but the analysis of film as the composition of multimodal and temporally unfolding perceptual scenarios is still far from even remotely reaching the levels of standardization and interoperability that is already taken for granted in the study of textual data through TEI XML and other mark-up languages and that goes beyond unstructured collections of tags,[22] which are, however, important steps towards such a shared data model.

20 Nanne van Noord et al., "Automatic Annotations and Enrichments for Audiovisual Archives," in *ICAART 2021: Proceedings of the 13th International Conference on Agents and Artificial Intelligence: February 4–6, 2021 – Volume 1: ARTIDIGH 2021*, ed. A. P. Rocha et al. (Vienna: SciTePress, 2021), 638, https://doi.org/10.5220/0010387706330640.

21 For a consideration of the advantages and disadvantages of time-based tools and QDAs that enable the integration of multiple resources, see Liliana Melgar Estrada et al., "Film Analysis as Annotation: Exploring Current Tools," *The Moving Image* 17, no. 2 (2017): 40–70, https://doi.org/10.5749/movingimage.17.2.0040; Liliana Melgar Estrada and Marijn Koolen, "Audiovisual Media Annotation Using Qualitative Data Analysis Software: A Comparative Analysis," *The Qualitative Report* 23, no. 13 (2018): 40–60, https://doi.org/10.46743/2160-3715/2018.3035.

22 Cf. Susan Aasman et al., "Tales of a Tool Encounter: Exploring Video Annotation for Doing Media History," *VIEW Journal of European Television History and Culture* 7, no. 14 (2018): 73–87, https://doi.org/10.18146/2213-0969.2018.jethc154; Allison Cooper, Fernando Nascimento, and David Francis, "Exploring Film Language with a Digital Analysis Tool: the Case of Kinolab," *Digital Humanities Quarterly* 15, no. 1 (2021), accessed June 30, 2023, http://digitalhumanities.org/dhq/vol/15/1/000515/000515.html.

Second, this glossary is formatted as a machine-readable semantic ontology,[23] so that every definition and every annotated value is represented as a triple, which can be published as linked open data. This means that every annotation is formatted as "*this* temporal segment of this film has *this* value within *this* type." What in simple free speech would be described as a "slow tracking shot to the left" will be annotated on several parameters: the value "slow" within the type CameraMovementSpeed, the value "tracking shot" within the type CameraMovementType, and the value "left" within the type CameraMovementDirection.[24] Most other applications of annotation tools that we have encountered so far either use taxonomies only on the level of "types," where the content of the single annotation is then only human readable, or they use a one-dimensional list of tags, which flattens observations on different levels of complexity onto the same plane and does not make annotations evaluable as relations.

Third, we implemented this ontology in the desktop annotation tool Advene, which offers a tiered, time-based view, as well as several other interfaces for entering or reading annotations.[25] In our current workflow, most of the types are still based on labor intensive manual annotations, with the exception of (manually corrected) shot detection and soundwave algorithm. But the semantic structure is built in such a way that the inclusion of interfaces for further automatic or semi-automatic annotations is possible and desired, and the first experiences with optical flow analysis and automatic segmentation of the audio track into music, voice, sounds, and silence are proving very promising. The fact that the large part of annotations is created manually has meant that, for collaborative work, intercoder reliability cannot be regarded as a given but must be brought

23 Henning Agt-Rickauer, Christian Hentschel, and Harald Sack, "Semantic Annotation and Automated Extraction of Audio-Visual Staging Patterns in Large-Scale Empirical Film Studies," *Proceedings of the Posters and Demos Track of the 14th International Conference on Semantic Systems (SEMANTiCS 2018)*, Vienna, Austria, September 10–13, 2018, accessed June 30, 2023, https://projectada.github.io/publications/semantics2018.pdf.

24 One might add that this translation of the perceptual qualities of audiovisual movement images into "digital objects" is less a question of human semantics or a logic of syntax but rather an organization of relations. Cf. Yuk Hui, *On the Existence of Digital Objects* (Minneapolis: University of Minnesota Press, 2016).

25 See https://advene.org, accessed March 15, 2024, originally designed by Olivier Aubert, Yannick Prié, and Pierre-Antoine Champin beginning in 2002. Based on a collaboration with Olivier Aubert, Advene has been further adjusted and extended to meet the specific requirements, not only regarding the (manual) annotation process but also interfaces for the film analytical ontology, video retrieval, and the support of RDF. These interfaces as well as in-application generation of visualizations and multimedia publications were the main factors for choosing Advene. The Ada-film ontology could in principle be implemented in other annotation tools like VIAN or ELAN.

about in an initial project phase. Some conclusions and recommendations[26] have proven central. The information that a given shot is either a "medium long shot" or a "long shot" does not follow a digital either/or discrimination and so it can very well be that different annotators assign different values. But, given the main interest in temporal and multimodal patterns, what is important are the dynamics of coming closer or moving away that follow a similar curve, even though single annotations might deviate. This also means that the parameters that record speeds of movement or intensities of displayed emotion in distinct levels have to be adjusted according to the thresholds of an individual film or scene. What might technically be a similar speed of camera movement could be annotated as the highest level in a slow cinema drama, but only as a slower movement if occurring in the hyperactive action sequence of a super-hero movie. What is important also to keep in mind is that even though many annotations take the shot as the basic temporal segment, this is not necessarily the default mode for all. Some segments are located at smaller units of impulses, accents, and gestures, while others can encompass larger elements of musical pieces, brightness, or contrast of the image over extended sequences. In the end, we must not forget that the graphical user interface of the annotation tool and the affordances of slowing down and repeating shots and frames, as well as the mental framing device of the ontology itself, shape and inform the perceptual processes that in turn form the basis of the annotation processes. This means that a certain decidophobia at the detail level must be countered by tracing the annotations back to an intuitive, embodied, and temporally unfolding viewing process.

And finally, because the data is machine readable, it was possible to implement a function in the application, where the annotated data is displayed in a manner resembling Eisenstein's: each modality or parameter is shown as a horizontal line, and the interplay between different levels of cinematic staging can be read vertically like the interplay between instruments and different voices on a music score sheet. Furthermore, the displayed parameters can be freely selected and ordered, and one can choose between a certain array of display types and color schemes in order to foreground the specific dynamics of the annotated segment.

Finding the right method of display for the single annotation values was not a trivial task, because we wanted a generic base but also wanted to retain room for customization depending on different analytical hypotheses. If one takes Jac-

26 One of the results of the Junior research group "Audio-Visual Rhetorics of Affect" is an extended manual for the annotation (see n. 15), which also includes – so far only in German – a short guideline for collaborative annotation: https://www.ada.cinepoetics.fu-berlin.de/media/ada-toolkit/Annotationshilfe_23_07_2021.pdf, accessed March 15, 2024.

ques Bertin's basic parameters[27] as a guideline (the different combinations of size, value, texture, color, orientation, and form), our display uses position, value, and size as the main variables, with colors used to distinguish data values within parameters, without them having distinct significance in themselves. In our experimental phase, we came to the conclusion that forms and shapes did not work as well and that, even for data that refer to directions like camera movements and movements in the image, the use of directions as variables was not productive either. This is possibly because these may be useful in horizontal displays of a single parameter but as soon as one is looking at both vertical interplays and horizontal developments – and therefore peripheral vision of patterns is needed – position, value, and size seem to be the most suitable.

Before I go on to explain how we use these visualized annotations, I want to anticipate or preempt two objections. First, this way of spatializing the temporality of cinematic images is of course subject to linear clock time. It does not attempt to grasp felt time, experienced time as visualized data. This aspect has to be inserted into the interpretation of the diagram. And, second, there is still the problem of retaining the work of shaping the data in the final result, as Johanna Drucker has pointed out in several publications – here amongst others:

> The standard approach to information visualization is to generate a graphic from live or static data. [. . .] The interpretative work of shaping the data disappears from view in the final result. The image displayed on screen, in print, or through other output devices appears as a statement of fact. The interpretative dimensions of the activity that shaped the data are rendered invisible, not so much concealed as simply missing from view, absent without a trace.[28]

This work of visibility is, on the one hand, the task of the textual framing of the visualizations but, on the other, is also a recurring tendency in case studies that have been working with this setup in so far as the visualizations are not a simple output left on its own but are themselves objects of inscriptions and interactions (as can be seen in Figure 3).

There are many different use scenarios for these visualizations, three of which I want to present here.[29] The first is to use them as evidence, as supporting material for an observation, an interpretation that one wishes to present to an

27 Jacques Bertin, *Semiology of Graphics: Diagrams, Networks, Maps* (Madison, WI: University of Wisconsin Press, 1983).

28 Johanna Drucker, *Visualization and Interpretation: Humanistic Approaches to Display* (Cambridge, MA; London: MIT Press, 2020), 1.

29 Many thanks to Yvonne Pfeilschifter, Pablo Tobaria, and Rebecca Zorco who prepared the basic annotation packages for my analyses in this chapter.

audience, for example, in transdisciplinary exchange where qualitative description often lacks the same impact or *evidentia* as visual accessories. The second scenario is the use of visualized annotations to support comparative arguments, demonstrating that scenes that one hypothesizes to have similar intended assertions and effects actually follow similar audiovisual dynamics, perhaps on different levels of staging, but with a common spatiotemporal gestalt. And the third scenario is the explorative use within a hermeneutic method, where annotations may help to foreground compositional patterns that can then be used to delve deeper into the structure of a film.

Fabricating Evidence

In order to show how Ada-visualizations can be used as supporting material to make evident an analytical observation, I am re-visiting an analysis of the film THE AGE OF STUPID (Franny Armstrong, UK 2009). It is a film that I had worked on in my dissertation,[30] and I knew more or less what the structure of the selected scene was. After a few shots showing news footage of a flood, incrementing year dates are shown growing in size on a black screen. The single ciphers are filled with diverse images of disaster and conflict in quick succession. A collage of short snippets of fictional news announcements is accompanied by a musical ostinato that slowly intensifies. Finally, the music achieves a crescendo, the ciphers jump towards us and we see an animated, expressionist view of panicking apes in a burning jungle. Thus the scene builds a linear progression on several modal layers of sound and image, an increase of tension and contrasts, that reaches a point of culmination and – in the sense of Eisenstein's ideas of pathos and the organic composition[31] – a leap from this heightened quantity into a different quality.

Looking at the pattern of the visualized annotation, these audiovisual dynamics are clear in the different annotated parameters displayed on the score sheet

30 Matthias Grotkopp, *Filmische Poetiken der Schuld. Die audiovisuelle Anklage der Sinne als Modalität des Gemeinschaftsempfindens* (Berlin, Boston: De Gruyter, 2017). This case study is missing in the abridged English translation: *Cinematic Poetics of Guilt, Audiovisual Accusation as a Mode of Commonality* (Berlin, Boston: De Gruyter, 2021).

31 Sergei Eisenstein, "The Structure of the Film" [1939], in *Film Form. Essays in Film Theory*, ed. Jay Leyda (New York, London: Harcourt, 1977), 150–178.

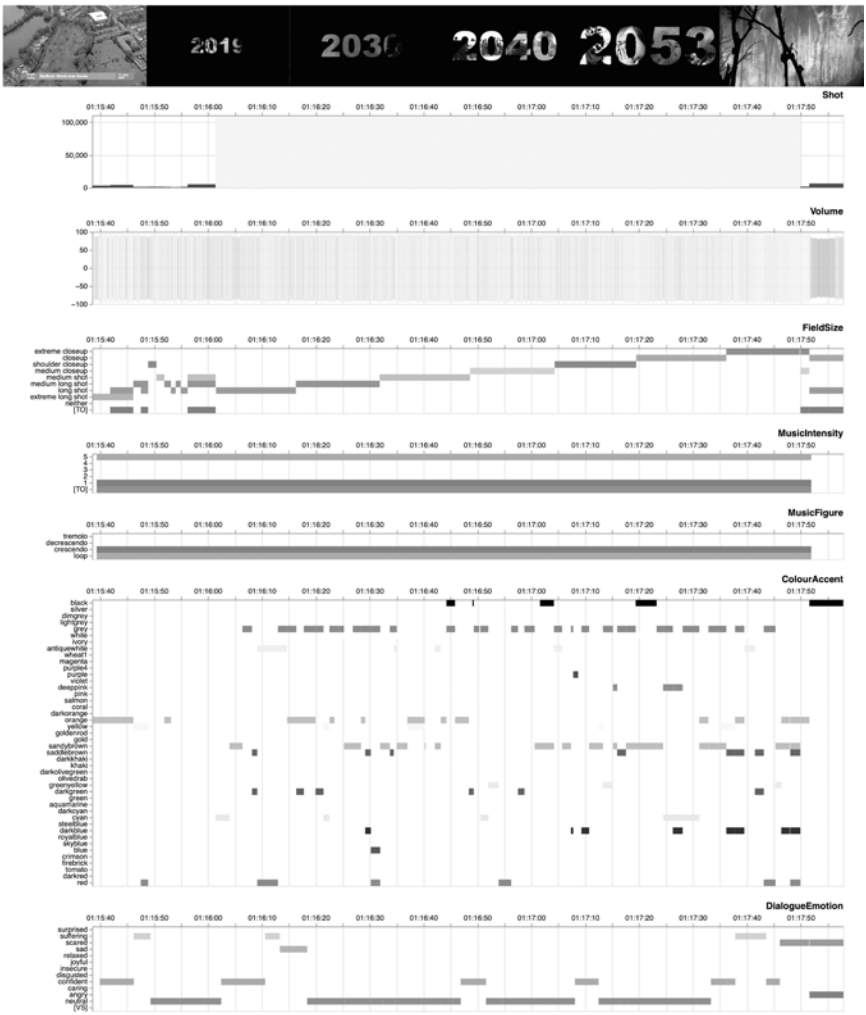


Figure 1: Ada Timeline-view of a scene from THE AGE OF STUPID (Franny Armstrong, UK 2009), generated in Advene.

(Figure 1).[32] Reading it from the top to the bottom, we see the sequence of year dates as a long, single shot, taking the faciality of the numbers as the dominant

32 The visualization printed in this volume shows a selection of seven layers of cinematic staging out of the almost 80 types predefined in the Ada Filmontology and out of the 25 that have been automatically and manually annotated for the preparation of this publication.

factor (instead of annotating the fast changes within the ciphers, acting as frames for a split screen montage of diverse footage from environmental catastrophes and social breakdown). The linear rise in the Volume is readable in the size and the growing brightness of the soundwave visualization and the FieldSize shows the linear approaching of the year dates. The musical intensity has been annotated as rising from very low to very high volume (1 [TO] 5) and the MusicFigure, the dynamic pattern of the music, is characterized as a loop that forms a crescendo, a repeated phrase in a continuous increase. This linear development in field size and musical intensity is accompanied by a chaotic, unpredictable change of colors within the ciphers. As the quantitatively dominant color in this shot would be the black background, the best access to this kind of movement is the annotation of ColorAccent that shows how the values for this parameter are shifting all the time. And finally, the DialogueEmotion, that is the voice-over from diverse found-footage sources, shifts its tone from for the most part "neutral" and "confident," "sad" and "suffering" to "scared" and "angry" as the sequence finally switches into a different quality, a different kind of image: the animated view of a burning forest.

Taken together, this methodology allows us to demonstrate how in this scene a specific dimension of feeling and thinking about the tipping points of the earth systems is staged as an audiovisual pattern – a relentless escalation of audiovisual intensity that reaches a point of impossible further progression and then switches completely.

This way of employing the method can be used in many circumstances, especially in transdisciplinary dialogue when one wants to show that the subjective experience of audiovisual images can be attributed to concrete patterns that can be made objectifiable in visualizations. In this sense, the overall methodological claim of this setup is to use digital humanities in order to bring together a formalist, empirical description of audiovisual images with a phenomenological description of subjective experience, in the vein of Vivian Sobchack, Jennifer Barker, and others. Visualizations are a tool that helps to "describe and explicate the general or possible structures and meanings that inform the experience and make it potentially resonant and inhabitable for others."[33] This does not mean, however, that the annotated data and the patterns identified in them make a naïve claim to objectivity or universality – the annotations are already interpretations of the images as expressive perception. But our hope is that the combination of visualization and

33 Vivian Sobchack, *Carnal Thoughts: Embodiment and Moving Image Culture* (Berkeley, Los Angeles: University of California Press, 2004), 5.

description "is resonant and the experience's structure sufficiently comprehensible"[34] to others.

It's a Match!

The main argument in our project, however, is not just that specific scenes generate specific ways of meaning-making and affective experience but that, throughout the circulation of films and videos on the climate crisis, biodiversity loss, and pollution, a genuine audiovisual discourse is formed where not only iconographic images (like polar bears on small ice floats) and keywords are re-appearing but also patterns of cinematic composition. The second use scenario for our visualized annotations therefore consists in finding out whether patterns like the one I have condensed from THE AGE OF STUPID (Figure 2, top) can be regarded as recurring types and what can be learned from comparing specimens. What I did next was to heuristically collect a few scenes – from an extended corpus of films dealing with the climate crisis in the last 20 years or so – that have as a theme the question of escalation, of reaching audiovisual tipping points as equivalents to the tipping points of earth's ecosystems, even though – or exactly because – they work with very different kinds of iconographies and different modes of rhetoric.

As a first comparison (Figure 2, middle), I have chosen a scene from ANTHROPOCENE – THE HUMAN EPOCH (Jennifer Baichwal et al., CAN 2018) which has a completely different approach on the level of its depicted contents: not the incrementing view on global, dispersed disasters of all kinds but literally zooming in on the demise of coral reefs and one individual coral in particular. But still the experience of aggravated demise towards a tipping point follows a similar movement pattern beginning at about 1:04:45 – the increasing Volume, the linear closing in of FieldSizes, which leads to an almost suffocating feeling of constriction and an increase of ImageIntrinsicMovement.

And then there is a sudden break, a shift into a different quality at 1:05:48, indicated by the switch of RecordingPlaybackSpeed from timelapse imagery back to normal playback speed, underlined by shots where the ImageIntrinsicMovement is characterized by a contrast between motion and stasis, which means that the intensity of movement in front of the camera is annotated as "1" or "2" (on a scale of 0 to 3) and as "0" at the same time, with the syntax element "[VS]" indicating a contrast. It is a different scale of place and time but as a cinematic pattern it is strikingly similar: escalation, tipping point, and flatlining.

34 Sobchack, *Carnal Thoughts*, 5.

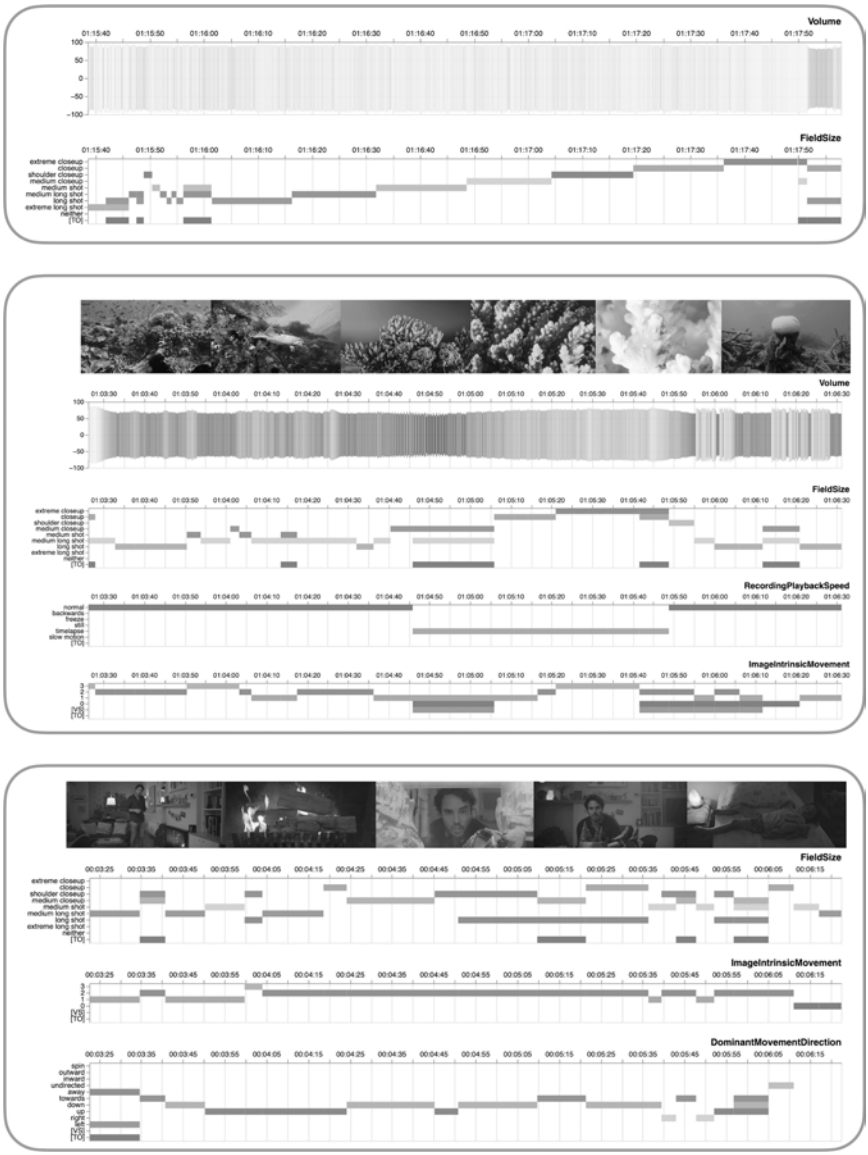


Figure 2: Ada Timeline-view of scenes (top to bottom) from The Age of Stupid (Franny Armstrong, UK 2009), Anthropocene – The Human Epoch (Jennifer Baichwal et al., CAN 2018), and 2040 (Damon Gameau, AUS 2019), generated in Advene.

The second exemplar of this kind of scene shows a very distinct re-working of this pattern (Figure 2, bottom). The explanatory exposition of 2040 (Damon Gameau, AUS 2019) is not working via montage or zooming-in but through a composite of mise-en-scène and computer-generated imagery. Its editing pattern is quite regular, with longer shots of the filmmaker and presenter talking into the camera alternating with shorter inserts. Here again the global problems are closing in on us in the type FieldSize. The linearity of this movement is agitated by a second voice of irregularity, like the color accents in THE AGE OF STUPID, but here it is the DominantMovementDirection in front of the camera that alternates between upwards and downwards movements until, at the end, different movements coalesce into one shot before the scene turns towards static shots.

The specific appropriation of this pattern in 2040 can be linked to the counterbalancing of the linearity of the FieldSizes getting closer by the simultaneous presence of long shots within these close-ups and medium close-ups. This is because the film here shows composite shots that insert miniature versions of large-scale landscapes and processes – an extraction site, calving glaciers, a forest – into the interior of a family home – an open-hearth fireplace, a freezer compartment, a coffee table. The stabilizing anchor of the presentation, who calmly tells us that the problems are solvable while the events around him are escalating, therefore has its equivalent on the visual level, when the escalating processes become manageable and thus the tipping points become turning points. The scene ends with two shots without movement of the camera or movement in the image, showing the filmmaker's sleeping daughter and the calm exterior of the house. Here too the quantitative aspect of rising intensity shifts into a different quality. But the calming moment at the end of the scene, which is the equivalent to the catastrophic explosion in THE AGE OF STUPID or the flatlining of extinction in ANTHROPOCENE – THE HUMAN EPOCH, is reinterpreted as a return to peaceful equilibrium. So the inherent narrative or history of "escalation" as a mode of intervention seen through these three films is its re-working of alarm and alarmism into either loss and mourning or hope and solution-focused optimism.

With the help of the visualizations, it is possible to show precisely how these scenes, even if they are very different on the level of depicted content or spoken or written information, are part of a shared kind of topic model on the level of audiovisual dynamics. One hypothesis that is beyond the scope of this short chapter is that these kinds of audiovisual topoi travel beyond the limits of the format of documentary film-making, and are also shared and appropriated by fictional films and social media videos.

What Am I Looking At?

Comparing scenes that were already heuristically grouped together by informed intuition is one thing. But our idea was also to use the visualized annotations in an exploratory manner. This next section tries to re-construct the process of interpreting an annotated scene without a preformed idea of its pattern or a certain typology as a scaffold.

In this case it is the visualization of a scene from another documentary film, ThuleTuvalu (Matthias van Gunten, CH 2014). The film as a whole rather straightforwardly crosscuts between its two main settings: Greenland, where the ice disappears, and Tuvalu, where the rising sea levels make the islands disappear. The specific scene was chosen because there is an interesting (and for this film non-recurring) match–cut of camera movement between the two places that makes them fuse in a different quality, and I was curious to see whether the visualized annotation could show that there was something more happening.

What one can see by looking at the types CameraMovementDirection, CameraAngle, and DominantMovementDirection in the image is that the scene is not split in two: the Arctic here and the South Pacific there. Rather it has four parts. Following the gathering of momentum through diverse movements as a dynamic of departure after the Narwhal hunt in the previous scenes, a continuous movement from left to right is established as the sledge dogs cross a crack in the thinning ice (1). This crossing is interrupted; a movement to the left and backwards shows a dog unable to make it to the other side in a shot that is at 58 seconds by far the longest in this sequence (2). Then a new vector comes into play, a travelling forward of sleighs and camera (3), supported by an intensifying slow, somber electronic music. This increasing intensity transports the split in the ice towards the islands on the other side of the world as the camera tilts down while moving backwards from an extreme low angle view into the palm trees and the sky to a high angle view of people on the back of a pickup truck (*). And there, in the South Pacific island, there are no more distinct movement vectors to the left or right, forward or backward, but only “undirected” movements (4), the camera gently wavering, people moving in all directions. The water is coming from everywhere.

What happened in my descriptions is that of course I could not help making the instant feedback loop into the hermeneutic understanding of the scene and the film, making sense of the annotated data in terms of experience and meaning-making. Coming back to Johanna Drucker’s admonition quoted earlier, it is important that the effect of making the features of audiovisual composition discrete and quantifiable, in the process of generating annotations and visualizations, is held in check by making them neither the end product nor by claiming

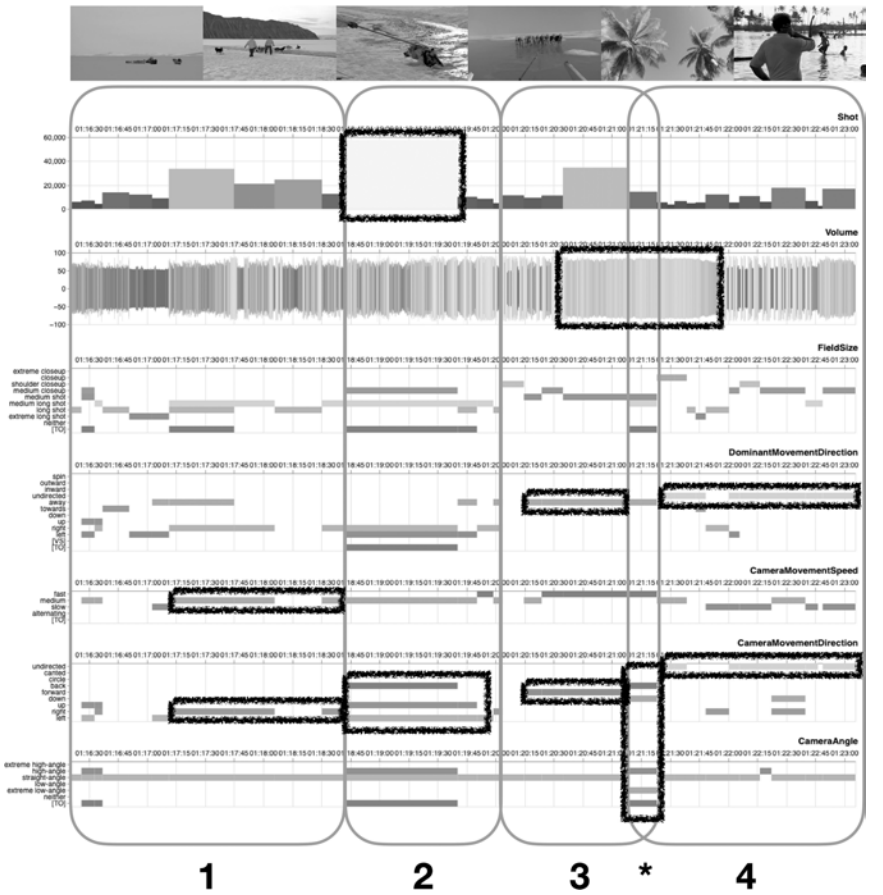


Figure 3: Ada Timeline-view of a scene from THULETUVALU (Matthias van Gunten, CH 2014), generated in Advene.

that they are given data. The process of making the annotations and the creative, enunciative act of inscribing into the visualizations – segmenting, highlighting, foregrounding – are just as important, and all of them are bracketed by the goal of reconstructing the actual primary data: the embodied experience of the audio-visual images from a concrete research perspective, which tries to make itself objective without a claim to universality.

Perspectives

My conclusion to this short outline of our method and the sketches of case studies is three-fold. The first aspect comes from the perspective of this concrete project, looking at the interventionist potential of cinematic world projections in fictional films, documentaries, and activist films on climate change. What does this intervention look and feel like? Comparing the visualizations of annotation data from the films, series, and videos helps us recognize and demonstrate the recurring patterns and their variations, like the dynamics of "escalation towards tipping points." Other candidates for this kind of pattern are scenes that show the growing instability of environments or scenes that make different spatial, temporal, and agential scales clash with each other. We are optimistic that it will be possible to show how these patterns move between fiction and non-fiction, and between the dissemination of scientific knowledge and activism, and that there are perhaps shifts and differences in the development of the genres of climate fiction and climate documentary, following different priorities from communicating scientific facts in earlier films to emphasizing modes of action and activism at a later stage.

The second conclusion regards the potential of further development of the tools and principles. Starting with the generation of data, a push toward more automated feature detection would have the benefit of easing the workload of individual case studies, and it would open the method towards bigger corpora and quantitative evaluation. While insisting on a qualitative empirical approach, however, it will remain necessary to implement an accessible mode of manual correction for automatically generated data, because the error rate of well-performing algorithms may be negligible and often equalizing for large corpora but is not so for individual close readings.

Even more interesting and computationally challenging is the evaluation of the annotated data as pattern. So far, the vertical and horizontal connections between the annotated values are literally "drawn" by us as scholars. Obviously, the amount of available data is still far beneath the threshold of what is needed for machine learning, but we are still looking at the possibility of querying a collection of annotations not just for single annotated values but for diachronous or synchronous occurrences of specific values and, most importantly, for combinations of both, like the scenes from the second case study that were all characterized by a linear approaching of field sizes. How far can these kinds of rule-based reasonings be taken? And what use scenarios can be imagined for a general database of film annotations across corpora defined by topics, formats, genres, or periods once shared data models have been achieved, which makes exchange between projects and instances possible? And, in this respect, we think of the Ada Filmontology as an offer-

ing for others to improve and broaden. Whereas archival data and film historical data about distribution and exhibition[35] are increasingly transformed into Open Data and Linked Open Data,[36] which enable the comparison and exchange of information about entities, persons, works, variations, and events, a level of interchange between the results from the Distant Viewing Lab, the VIAN platform, Cinemetrics, the Ada Filmontology, and others would only be a start toward making this a reality for aesthetic parameters. Finally making metadata about audiovisual works connect to metadata about their audiovisuality would be a further step toward combining qualitative and quantitative research on a historical poetics of film, television, and video.

This leads me to the third conclusion, looking at the future of this particular methodological set-up beyond the case studies that have been conducted so far. The financial crisis and the climate crisis have been investigated across different genres and formats but both projects look at the more or less homogenous timeframe of the early twenty-first century with a strong bias towards North America and Europe. A logical next step would be to expand this with both historical and transnational cross-sections, inviting scholars to appropriate the model and contribute their particular perspectives. Another step would be the further integration of the many different projects and communities with similar but not identical ideas on how to use digital methods for film analytical annotation with a focus not too much on single tools and software, which may become obsolete or superseded by new developments,[37] but on standards that can be transferred, passed on, versioned, related. Moving annotation data between one desktop application and another, from Advene to VIAN or ELAN, or to a browser-based tool (ideally one that enables collaborative and simultaneous annotation) and back again, benefiting from their interfaces for different recognizers for movement, sound, color, or language or using different forms of display, can only be done reliably on a larger scale with a machine-readable data framework that can link the different vocabu-

35 See, for example, Vincent Ducatteeuw et al., "Critical Reflections on Cinema Belgica: The Database for New Cinema History in Belgium," *Journal of Open Humanities Data* 9 (2023): 1–16, accessed June 30, 2023, https://doi.10.5334/johd.91.

36 Adelheid Heftberger and Paul Duchesne, "Cataloguing Practices in the Age of Linked Open Data: Wikidata and Wikibase for Film Archives," FIAF – International Federation of Film Archives, June 2020, accessed June 30, 2023, https://www.fiafnet.org/pages/E-Resources/Cataloguing-Practices-Linked-Open-Data.html.

37 Christine Barats, Valérie Schafer, and Andreas Fickers, "Fading Away . . . The Challenge of Sustainability in Digital Studies," *Digital Humanities Quarterly* 14, no. 3 (2020), accessed June 30, 2023, http://www.digitalhumanities.org/dhq/vol/14/3/000484/000484.html.

laries.[38] Looking for standards in this sense does not necessarily mean that one project's ontology fits the needs and requirements of another, but that there is a shared principle that makes (selections of) their findings transferable. In the spirit of the FAIR-principles of Open Science, the annotation data of our projects will be published in a repository so that they can be viewed and reused by others. This kind of approach to basic film analytical data is not yet established for questions of film aesthetics, but their integration into an eco-system of film analytical ontologies, repositories, and databases would make these annotations and visualizations reusable, and open them up to a wide range of scales and questions, from statistical analysis to historical phenomenology, that is, the reconstruction of the experiential aspects of audiovisual poetics in different contexts, modes, and practices.

Postscript

The irony of using resource intensive infrastructure to research the cultural production of the climate crisis does not escape me – even though the laptops, streaming, and server capacities in this project are miniscule in comparison to large scale machine learning setups. It is part of the inescapable contradictions of living and working as an academic in today's western industrial world.[39] This does not release us from the task of not just raising awareness (and alarm) but also identifying and eliminating unnecessary or excessive use of fossil energy and extracted resources in every step of teaching, research, and outreach.

38 One pioneering project in this field is https://onomy.org (accessed June 30, 2023) that was developed by John Bell as a part of Dartmouth's Media Ecology Project. However, one can see that only a few of the published taxonomies on this platform are making use of the possibilities of the RDF framework instead of "only" creating a list of labels, which makes it more cumbersome to map annotations created with different taxonomies onto each other.

39 Cf. Anne Baillot et al., "Digital Humanities and the Climate Crisis: A Manifesto," https://dhc-barnard.github.io/dhclimate/, accessed June 30, 2023; Bethany Nowviskie, "Digital Humanities in the Anthropocene," *Digital Scholarship in the Humanities* 30 (2015), i4–i15, https://doi.org/10.1093/llc/fqv015.

References

Aasman, Susan, Liliana Melgar Estrada, Tom Slootweg, and Rob Wegter. "Tales of a Tool Encounter: Exploring Video Annotation for Doing Media History." *VIEW Journal of European Television History and Culture* 7, no. 14 (2018): 73–87. https://doi.org/10.18146/2213-0969.2018.jethc154.

Agt-Rickauer, Henning, Christian Hentschel, and Harald Sack. "Semantic Annotation and Automated Extraction of Audio-Visual Staging Patterns in Large-Scale Empirical Film Studies." *Proceedings of the Posters and Demos Track of the 14th International Conference on Semantic Systems (SEMANTiCS 2018)*, Vienna, Austria, September 10–13, 2018. Accessed June 30, 2023. https://projectada.github.io/publications/semantics2018.pdf.

Arnold, Taylor, and Lauren Tilton. "Distant Viewing: Analyzing Large Visual Corpora." *Digital Scholarship in the Humanities* 34, issue supplement 1 (2019): i3–i16. https://doi.org/10.1093/llc/fqz013.

Baillot Anne, James Baker, Madiha Zahrah Choksi, Alex Gil, Kaiama L. Glover, Ana Lam, Alicia Peaker, Walter Scholger, Torsten Roeder, and Jo Lindsay Walton. "Digital Humanities and the Climate Crisis: A Manifesto." Accessed June 30, 2023. https://dhc-barnard.github.io/dhclimate/.

Bakels Jan-Hendrik, Matthias Grotkopp, Thomas Scherer, and Jasper Stratil. "Matching Computational Analysis and Human Experience: Performative Arts and the Digital Humanities." *Digital Humanities Quarterly* 14, no. 4 (2020). Accessed June 30, 2023. http://www.digitalhumanities.org/dhq/vol/14/4/000496/000496.html.

Barats, Christine, Valérie Schafer, and Andreas Fickers. "Fading Away . . . The Challenge of Sustainability in Digital Studies." *Digital Humanities Quarterly* 14, no. 3 (2020). Accessed June 30, 2023. http://www.digitalhumanities.org/dhq/vol/14/3/000484/000484.html.

Baxter, Mike, Daria Khitrova, and Yuri Tsivian. "Exploring Cutting Structure in Film, with Applications to the Films of D. W. Griffith, Mack Sennett, and Charlie Chaplin." *Digital Scholarship in the Humanities* 32, no. 1 (2017): 1–16. https://doi.org/10.1093/llc/fqv035.

Bellour, Raymond. *The Analysis of Film* [1973]. Bloomington: Indiana University Press, 2000.

Bertin, Jacques. *Semiology of Graphics: Diagrams, Networks, Maps*. Madison, WI: University of Wisconsin Press, 1983.

Chakrabarty, Dipesh. *The Climate of History in a Planetary Age*. Chicago, London: University of Chicago Press, 2021.

Cohen, Tom, ed. *Telemorphosis: Theory in the Era of Climate Change*. Ann Arbor: Open Humanities Press, 2012.

Cooper, Allison, Fernando Nascimento, and David Francis. "Exploring Film Language with a Digital Analysis Tool: The Case of Kinolab." *Digital Humanities Quarterly* 15, no. 1 (2021). Accessed June 30, 2023. http://digitalhumanities.org/dhq/vol/15/1/000515/000515.html.

Drucker, Johanna. *Visualization and Interpretation: Humanistic Approaches to Display*. Cambridge, MA; London: MIT Press, 2020.

Drummond, Tamara, and Janina Wildfeuer. "The Multimodal Annotation of Gender Differences in Contemporary TV Series. Combining Qualitative Questions and Quantitative Results." In *Annotations in Scholarly Editions and Research: Functions, Differentiation, Systematization*, edited by Julia Nantke and Frederik Schlupkothen, 35–58. Berlin, Boston: De Gruyter, 2020.

Ducatteeuw, Vincent, Daniel Biltereyst, Philippe Meers, Christophe Verbruggen, Dries Moreels, Julia Noordegraaf, and Sally Chambers. "Critical Reflections on Cinema Belgica: The Database for New Cinema History in Belgium." *Journal of Open Humanities Data* 9, no. 1 (2023): 1–16. https://doi.10.5334/johd.91.

Eisenstein, Sergei. “Vertical Montage” [1940]. In *Selected Works, Vol. 2: Towards a Theory of Montage 1937–1940*, edited by Michael Glenny and Richard Taylor, 327–399. London: I. B. Tauris, 2010.

Eisenstein, Sergei. “The Structure of the Film” [1939]. In *Film Form: Essays in Film Theory*, edited by Jay Leyda, 150–178. New York, London: Harcourt, 1977.

Estrada, Liliana Melgar, Eva Hielscher, Marijn Koolen, Christian Gosvig Olesen, Julia Noordegraaf, and Jaap Blom. “Film Analysis as Annotation: Exploring Current Tools.” *The Moving Image* 17, no. 2 (2017): 40–70. https://doi.org/10.5749/movingimage.17.2.0040.

Estrada, Liliana Melgar, and Marjin Koolen. “Audiovisual Media Annotation Using Qualitative Data Analysis Software: A Comparative Analysis.” *The Qualitative Report* 23, no. 13 (2018): 40–60. https://doi.org/10.46743/2160-3715/2018.3035.

Ferguson, Kevin L. “Volumetric Cinema.” *[in]transition: Journal of Videographic Film and Moving Image Studies* 2, no. 1 (2015). Accessed June 30, 2023. http://mediacommons.org/intransition/2015/03/10/volumetric-cinema.

Flückiger, Barbara. “A Digital Humanities Approach to Film Colors.” *The Moving Image: The Journal of the Association of Moving Image Archivists* 17, no. 2 (2017): 71–94.

Flückiger, Barbara, and Gaudenz Halter. “Methods and Advanced Tools for the Analysis of Film Colors in Digital Humanities.” *Digital Humanities Quarterly* 14, no. 4 (2020). Accessed June 30, 2023. http://www.digitalhumanities.org/dhq/vol/14/4/000500/000500.html.

Grotkopp, Matthias. *Filmische Poetiken der Schuld. Die audiovisuelle Anklage der Sinne als Modalität des Gemeinschaftsempfindens*. Berlin, Boston: De Gruyter, 2017. Abridged translation: *Cinematic Poetics of Guilt, Audiovisual Accusation as a Mode of Commonality*. Berlin, Boston: De Gruyter, 2021.

Hall, Stuart. “Cultural Studies and Its Theoretical Legacies.” In *Cultural Studies*, edited by Larry Grossberg, Cary Nelson, and Paula Treichler, 277–294. London: Routledge, 1992.

Haraway, Donna J. *Staying with the Trouble: Making Kin in the Chthulucene*. Durham, NC; London: Duke University Press, 2016.

Heftberger, Adelheid. *Digital Humanities and Film Studies: Visualizing Dziga Vertov’s Work*. Cham: Springer Nature, 2018.

Heftberger, Adelheid, and Paul Duchesne. “Cataloguing Practices in the Age of Linked Open Data: Wikidata and Wikibase for Film Archives.” FIAF – International Federation of Film Archives, June 2020. Accessed June 30, 2023. https://www.fiafnet.org/pages/E-Resources/Cataloguing-Practices-Linked-Open-Data.html.

Horn, Eva, and Hannes Bergthaller. *The Anthropocene: Key Issues for the Humanities*. New York: Routledge, 2020.

Hui, Yuk. *On the Existence of Digital Objects*. Minneapolis: University of Minnesota Press, 2016.

Kramer, Michael J. “What Does a Photograph Sound Like? Digital Image Sonification as Synesthetic Audiovisual Digital Humanities.” *Digital Humanities Quarterly* 15, no. 1 (2021). Accessed June 30, 2023. http://digitalhumanities.org/dhq/vol/15/1/000508/000508.html.

Mittell, Jason. “Videographic Criticism as a Digital Humanities Method.” In *Debates in the Digital Humanities 2019*, edited by Matthew K. Gold and Lauren F. Klein, 224–242. Minneapolis: University of Minnesota Press, 2019.

Müller, Cornelia, and Hermann Kappelhoff. *Cinematic Metaphor: Experience – Affectivity – Temporality*. Berlin, Boston: De Gruyter, 2018.

Noord, Nanne van, Christian Gosvig Olesen, Roeland Ordelman, and Julia Noordegraaf. “Automatic Annotations and Enrichments for Audiovisual Archives.” In *ICAART 2021: Proceedings of the 13th International Conference on Agents and Artificial Intelligence: February 4–6, 2021. Volume 1:*

ARTIDIGH 2021, edited by Ana Paula Rocha, Luc Steels, and Jaap van den Herik, 633–640. Vienna: SciTePress, 2021. https://doi.org/10.5220/0010387706330640.

Nowviskie, Bethany. "Digital Humanities in the Anthropocene." *Digital Scholarship in the Humanities* 30 (2015): i4–i15. https://doi.org/10.1093/llc/fqv015.

Scherer, Thomas, Sarah Greifenstein, and Hermann Kappelhoff. "Expressive Movements in Audiovisual Media: Modulating Affective Experience." In *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction*, edited by Cornelia Müller, Alan Cienki, Ellen Fricke, Silva H. Ladewig, David McNeill, and Jana Bressem, vol. 2, 2081–2092. Berlin, New York: De Gruyter Mouton, 2014.

Sobchack, Vivian. *Carnal Thoughts: Embodiment and Moving Image Culture*. Berkeley, Los Angeles: University of California Press, 2004.

Tsing, Anna, Heather Swanson, Elaine Gan, and Nils Bubandt, eds. *Arts of Living on a Damaged Planet*. Minneapolis: University of Minnesota Press, 2017.

Tsivian, Yuri. "Cinemetrics, Part of the Humanities' Cyberinfrastructure." In *Digital Tools in Media Studies: Analysis and Research*, edited by Michael Ross, Manfred Grauer, and Bernd Freisleben, 93–100. Bielefeld: Transcript, 2009.

Yussof, Kathryn. *A Billion Black Anthropocenes or None*. Minneapolis: University of Minnesota Press, 2018.

Susan Aasman and Tom Slootweg

"Pure Information, Not the Real Thing": Digital Hermeneutics and Nelson Sullivan's Videographic Legacy (1983–1989)

Introduction

This chapter explores the need for a clear digital methodology that can support film and media historical research. We believe that digital hermeneutics can be a key approach for the analysis of large-scale historical audiovisual materials, while simultaneously acknowledging the complexities of this domain of research. For some time now, film and media historiography has been in flux. Not only the digitization of sources, but also the various proposals for methodological innovation have spurred conversation on the status of media historiography in the twenty-first century, as is the case in the present volume. In this chapter, we aim to expand on previous research in this regard with a special focus on digital hermeneutics to reflect on the interpretative challenges when dealing with the use of computational methods and large-scale audiovisual data.[1]

Digital hermeneutics has emerged in the humanities and social sciences as a response to the fact that, when we deal with data and digital tools, we are not only dealing with "texts," but also with the socio-materiality of the tools used.[2] As Silke Schwandt suggests, we need to understand the tools and procedures used by opening up "the 'black box' of our interpretation processes."[3] This observation follows Andreas Fickers, Juliane Tatarinov, and Tim van der Heijden's plea for a "reflexive turn" in developing critical approaches in all stages of (media) historical research, whether it is search, analysis, or presentation of results.[4] In addition,

1 See, for example, Susan Aasman et al., "Tales of a Tool Encounter: Exploring Video Annotation for Doing Media History," *VIEW Journal for European Television History and Culture* 7, no. 14 (2018): 73–87, https://doi.org/10.18146/2213-0969.2018.jethc154.

2 Alberto Romele, Marta Severo, and Paolo Furia, "Digital Hermeneutics: From Interpreting with Machines to Interpretational Machines," *AI & Society* 35 (2020): 73–86, https://doi.org/10.1007/s00146-018-0856-2.

3 Silke Schwandt, "Opening the Black Box of Interpretation: Digital History Practices as Models of Knowledge," *History and Theory* 61, no. 4 (2022): 82, https://doi.org/10.1111/hith.12281.

4 Andreas Fickers, Juliane Tatarinov, and Tim van der Heijden, "Digital History and Hermeneutics. Between Theory and Practice: An Introduction," in *Digital History and Hermeneutics: Between Theory and Practice*, ed. Andreas Fickers and Juliane Tatarinov (Berlin: De Gruyter Oldenbourg, 2022), 9.

https://doi.org/10.1515/9783111082486-018

we take inspiration from critical hermeneutics developed by Johanna Drucker. We do so to emphasize that much of our interpretative work, as readers and/or as media scholars, is informed by and positioned through "screen, device, platform, and projection environments."[5]

As pointed out by several scholars, digital hermeneutics can also be helpful in bridging qualitative and quantitative approaches, and developing more multilayered analyses.[6] In this, we follow Ben Jackson and Tim Hitchcock, who, on the well-known Old Bailey Voices website, argue for:

> a more generic approach to "big data" as a whole – an approach in which large scale patterns can be rapidly explored, and their constituent elements identified; or where single words or phrases (or just datum), can be radically contextualized within the full body of evidence available.[7]

This implies an investigative attitude that engages in various forms of scaled readings, zooming in and out of data, thereby applying forms of distant and close readings, but also, as we will show, modes of hyper reading.

From a digital media historiographical point of view, the possibilities and constraints of computer-assisted analysis have been explored and discussed for quite some years now. Melvin Wevers and Thomas Smits made prominent contributions regarding the automation of visual analysis and the "visual digital turn."[8] They demonstrate how convolutional neural networks enable the analysis of large-scale historical visual media collections to explore content and style. We also extend the insights yielded by Lauren Tilton and Taylor Arnold, who pioneered the analysis of audiovisual sources by means of computer vision algorithms, which is explored further in this volume by Tim van der Heijden, Taylor Arnold, and Lauren Tilton.[9] Lev Manovich's work on cultural analytics is another

5 Johanna Drucker, *Visualization and Interpretation: Humanistic Approaches to Display* (Cambridge, MA: MIT Press, 2020), 7.

6 See, for example, Paolo Gerbaudo, "From Data Analytics to Data Hermeneutics: Online Political Discussions, Digital Methods and the Continuing Relevance of Interpretive Approaches," *Digital Culture and Society* 2 (2016): 95–111, https://doi.org/10.14361/dcs-2016-0207.

7 Ben Jackson and Tim Hitchcock, "Old Bailey Voices," accessed March 1, 2023, https://oldbaileyvoices.org.

8 Melvin Wevers and Thomas Smits, "The Visual Digital Turn: Using Neural Networks to Study Historical Images," *Digital Scholarship in the Humanities* 35, no. 1 (2020): 194–207, https://doi.org/10.1093/llc/fqy085.

9 Taylor Arnold and Lauren Tilton, "Distant Viewing: Analyzing Large Visual Corpora," *Digital Scholarship in the Humanities* 34, issue supplement 1 (2019): i3–i16, https://doi.org/10.1093/llc/fqz013; Tim van der Heijden, Taylor Arnold, and Lauren Tilton, "Distant Viewing the Amateur Film Platform," this volume.

important point of reference.[10] Manovich argues that the explosion of sources generated in our digital culture necessitates a re-evaluation of the basic premise on which we do our research in the humanities, and cultural studies specifically. According to him, "a systematic use of *large-scale computational analysis and interactive visualization of cultural patterns* will become a basic research method."[11] Although Manovich focuses predominantly on the synchronic analysis of digital visual culture, the ongoing efforts to digitize the "back catalog" of our media heritage also raises the question of what basic research methods are required for media historians.

In our contribution, we conduct a reflexive investigation of the consequences of digital hermeneutics for media historical research, with a focus on the challenges that come with media sources that are multi-layered, multimodal, and polyvocal. We explore the ramifications of deploying computer vision algorithms in the historical analysis of digitized audiovisual sources, as part of a larger set of tools and methods. In our chapter, we will clarify these ambitions with reference to audiovisual sources that pertain to historical autobiographical and amateur media production. These sources also represent transformative changes that took place in the ways in which amateur audiovisual media were produced and shared in different contexts.[12] In addition, as many scholars have emphasized, this material is notoriously difficult to reduce to neatly categorized units of analysis because it does not necessarily adhere to formal aesthetic or narrative conventions.[13] Our object of study, furthermore, challenges traditional concepts of archival collections. Either based in personal collections, or floating around online, unmanaged large-scale user-generated collections gain volume and impact, while we still need to work out how to approach them methodologically.[14]

10 Lev Manovich, *Cultural Analytics* (Cambridge, MA: MIT Press, 2020).

11 Manovich, *Cultural Analytics*, 30, italics original.

12 See, for example, Annamaria Motrescu-Mayes and Susan Aasman, *Amateur Media and Participatory Culture: Film, Video and Digital Media* (London: Routledge, 2019).

13 See, for example, Susan Aasman, Tim van der Heijden, and Tom Slootweg, "Amateurism: Exploring Its Meaning in the Age of Film, Video and Digital Media," in *Digital Roots: Historicizing Media and Communication Concepts of the Digital Age*, ed. Gabriele Balbi, Nelson Ribeiro, Valérie Schafer, and Christian Schwarzenegger (Berlin: De Gruyter, 2021), 261; Patricia R. Zimmermann, "Speculations on Home Movies: Thirty Axioms for Navigating Historiography and Psychic Vectors," in *Private Eyes and the Public Gaze: The Manipulation and Valorisation of Amateur Images*, ed. Sonja Kmec and Viviane Thill (Trier: Kliomedia, 2009), 21.

14 See, for example, Susan Aasman, "Unlocking Multiple Histories of Amateur Media: From Micro- to Macro-Histories," *Screen* 61, no. 1 (2020): 157–166, https://doi.org/10.1093/screen/hjaa011.

Sullivan's Queer Fish Eye

We opted for a case study that fits our fascination for historicizing pervasive modes of first-person media production, of which "vlogging" seems to be the most dominant example nowadays. Recently, the work of American videographer Nelson Sullivan (1948–1989) has garnered attention in popular and scholarly discourse in that regard. As a chronicler of 1980s queer life in New York City, Nelson Sullivan captured his everyday life and encounters with prominent members of the bustling club scene, against the backdrop of the unfolding, yet rarely explicitly acknowledged, HIV/AIDS epidemic of the 1980s. The hours of footage, shot by Sullivan on videotape between 1983 and 1989, are currently under the care of NYU's Fales Library.[15] Moreover, his work was made available to the public, in curated form, since 2008 on the YouTube channel 5ninthavenueproject, created by Dick Richards (1946–2018).[16] Matthew Terrell, a writer for *Slate.com*, has characterized Sullivan's legacy as a "rare cultural treasure trove" and "some of the most thorough documents of queer history at a pivotal and challenging time."[17] What stands out in particular is the idiosyncratic aesthetic of his videographical work in "a time long before iPhones," as Sullivan "often turned the camera to himself and, with a fisheye lens, captured a full spectrum of the world around him."

Several poignant aspects thus converge with the videographic legacy of Nelson Sullivan. First, we identify the retrospective status given to Sullivan's work in popular discourse as one of the first "vloggers" in the United States. From a media historical perspective, this observation can come across as an anachronism, certainly when considering the long history of similar or adjacent media practices and aesthetics by Maya Deren, Stan Brakhage, Jonas Mekas, Nam June Paik, and others in New York's avant-garde circles from roughly the 1950s onwards.[18] Moreover, Sullivan's work can also be understood within the specific sociocultural and media-historical context of the 1980s. As media theorist James Moran wrote, the arrival of the video camcorder in the 1980s signified a shift in the characteristics

15 Nelson Sullivan Collection, Fales Library NYU, http://dlib.nyu.edu/findingaids/html/fales/mss_357/, accessed February 12, 2023. The collection comprises various source types: objects, film, photographs, and objects. For this chapter we focus predominantly on his videographic work because of its characterization as early "vlogging."

16 See 5ninthavenueproject, https://www.youtube.com/@5ninthavenueproject, accessed February 12, 2023.

17 Matthew Terrell, "About Last Night: Nelson Sullivan's Camcorder Videos of 1980s Queer New York are a Rare Cultural Treasure," *Slate*, June 19, 2019, accessed April 2, 2023, https://slate.com/human-interest/2019/06/nelson-sullivan-videos-queer-new-york.html.

18 See, for example, Jeffrey K. Ruoff, "Home Movies of the Avant-Garde: Jonas Mekas and the New York Art World," *Cinema Journal* 30, no. 3 (1991): 6–28, https://doi.org/10.2307/1224927.

of amateur media production.[19] According to Moran, the introduction of video created new possibilities for otherwise underrepresented communities, such as those historical social formations that preceded the current LGBTQI+ community in the United States.[20] Sullivan voiced his hopes and ambitions on tape so that his videographical work might be picked up for broadcast by one of the (local) cable networks in and around NYC.

Second, his work, made in the 1980s with a VHS camera and recorder, and later with a Video8 camcorder, has been digitized and uploaded on YouTube by his friend and artistic partner Dick Richards, well after Sullivan's death. The YouTube channel 5ninthavenueproject contains around 700 videos, comprising many hours of footage freely available to viewers. Communication scholar Joseph Deleon has characterized these efforts as part of a "queer archive effect," which he considers "a term that signals both a remediation of past media chronicling queer lives and a queer practice of collecting."[21] Third, the autobiographical dimensions of Sullivan's work are salient. These are described in various ways. Sociologists John Goodwin and Laurie Parsons, for example, regard his videos as a form of "diary-letter hybridity" and foreground the "aesthetics of walking" through public, private, and queer spaces as the most salient features.[22] This appraisal is shared by Ricardo Montez, who furthermore highlights that "Sullivan's life on tape is an index of bodily gesture, a document of technologized embodiment that animates viewers in its performative capacities."[23] Further, Montez regards the act of walking as a "narrative performance."[24]

From a conceptual and analytical point of view, Sullivan's videographical work offers challenges, especially when considering the methodological ambitions of our contribution. Many of the above-mentioned historical characteristics of non-professional video productions are explored more in-depth by Tom Slootweg in 2018 but were also discussed by various scholars over the last four decades.[25]

---

19 James M. Moran, *There's No Place Like Home Video* (Minneapolis: University of Minnesota Press, 2002), 40.

20 Moran, *There's No Place Like Home*, 27.

21 Joseph Deleon, "Nelson Sullivan's Video Memories: YouTube Nostalgia and the Queer Archive Effect," *The Velvet Light Trap* 86 (2020): 17, https://doi.org/10.7560/VLT8603.

22 John Goodwin and Laurie Parsons, "A Life in Motion: Exploring Auto/Biographical Exchanges by 'Walking With' Nelson Sullivan," *Sociological Research Online* 27, no. 3 (2022): 724–744, https://doi.org/10.1177/13607804211025541.

23 Ricardo Montez, "Virtuosic Distortion: Nelson Sullivan's Queer Hand," *ASAP/Journal* 2 (2017): 407, https://doi.org/10.1353/asa.2017.0030.

24 Montez, "Virtuosic Distortion," 411.

25 Tom Slootweg, "Resistance, Disruption, Belonging: Electronic Video in Three Amateur Modes" (PhD diss., University of Groningen, 2018). For a variety of publications concerning the character-

For example, in the early 1990s, media theorist Sean Cubitt broadly characterized video in terms of its "ubiquity and sheer noisiness" and foregrounded its ability to convey "complex modes of storytelling" that challenge "spatial orientation," "time orientation," and conventional notions of "truth and truthfulness" in a media landscape still dominated by institutional and commercial television broadcast.[26] Jon Dovey, almost ten years later, pointed out that camcorder footage:

> has become the privileged form of TV "truth telling," signifying authenticity and an indexical reproduction of the real world; indexical in the sense of presuming a direct and transparent correspondence between what is in front of the camera lens and its taped representation. Secondly, the camcorder text has become the form that most relentlessly insists upon a localised [sic], subjective and embodied account of experience.[27]

This specific entanglement of indexicality and authenticity results in layers of complexity that are furthermore complicated by the autobiographical dimensions in Sullivan's work. To conceptualize this, we propose to also take into consideration the notion of *deixis*. This notion originates from linguistic theory, to theorize how people, time, and space are construed relationally in natural language and face-to-face communication.[28] Media scholar Pepita Hesselberth also adopted this concept to understand the way in which embodiment and spatiotemporal relationships are communicated to audiences in cinemas. Drawing from film scholar Francesco Casetti, she discusses three deictic categories that are predominantly at play: "an 'I' (i.e., the enunciator, or filmmaker), a 'you' (i.e., the addressee, or spectator), and a 'he' (a character, or a 'thing', that is, the film itself)."[29] In addition to Cassetti's three pronoun-based deictic categories, Hesselberth highlights the im-

istics of video, see: Roy Armes, *On Video* (London: Routledge, 1988); Deidre Boyle, "From Portapak to Camcorder: A Brief History of Guerrilla Television," *Journal of Film and Video* 44 (1992): 67–79; John Belton, "Looking Through Video: The Psychology of Video and Film," in *Resolutions: Contemporary Video Practices*, ed. Michael Renov and Erika Suderberg (Minneapolis: University of Minnesota Press, 1996), 61–72; David Buckingham, Rebekah Willet, and Maria Pini, *Home Truths? Video Production and Domestic Life* (Ann Arbor: University of Michigan Press and University of Michigan Library, 2011); Janna Houwen, *Film and Video Intermediality: The Question of Media Specificity in Contemporary Moving Images* (London: Bloomsbury, 2017); Motrescu-Mayes and Aasman, *Amateur Media and Participatory Culture*.

26 Sean Cubitt, *Timeshift: On Video Culture* (London: Routledge, 1991), 4.

27 Jon Dovey, *Freakshow: First Person Media and Factual Television* (London: Pluto Press, 2000), 55.

28 See, for example, Anja Stuckenbrock, *Deixis in der face-to-face-Interaktion* (Berlin: De Gruyter, 2015).

29 Pepita Hesselberth, *Cinematic Chronotopes: Here, Me, Now* (Amsterdam: Amsterdam University Press, 2014), 11. For the classical study by Casetti, see: Francesco Casetti, *Inside the Gaze: The Fiction Film and its Spectator* (Bloomington: Indiana University Press, 1999).

portance of how the camera conveys the "me," and furthermore underlines the importance of two more spatiotemporal categories: "here" and "now."[30]

The deictic complexity sketched above applies in full to Sullivan's autobiographical and embodied narrative strategies. This deictic complexity is further increased by the fact that an understanding of these audiovisual texts is situated in time and space, which makes any hermeneutic interpretation user-dependent, as Johanna Drucker would emphasize: "the dimensions of time and space are influenced by factors of experience, while any understanding of a text is always produced from a position marked by historical, cultural, and individual factors."[31] In the case of the Sullivan collection, it is important to emphasize that these spatiotemporal dimensions matter, particularly so when one remembers how his videographic legacy speaks and spoke to different users at different junctures in time, through various media landscapes and formats over the last 30 years. In other words, how does the historical trajectory of the collection align with what Drucker terms the "discursive heterogeneity" of the cultural record?[32] This question consequently forces us to acknowledge that our interpretative model should allow for a multilayered digital hermeneutical approach.

Moreover, as discussed elsewhere, amateur video's affordances to create long takes and record environmental sound, in combination with a distinctly unpolished aesthetics, resulted in less-controlled, polyvocal, and lengthy recordings.[33] Thus, the inherent multimodality of the material at hand, combined with the lack of strictly delineated formal features and running times, add to the complexity of this type of data when exploring computational methods to conceptualize and map this unique collection. As Lisa Lebow summarized some of the most poignant characteristics of the closely aligned documentary genre of first-person film, they are "foremost about a mode of address: these films 'speak' from the articulated point of view of the filmmaker who readily acknowledges her subjective position."[34]

In what follows, we will propose a way to cope with these layers and their multimodal and deictic complexity and explore what digital hermeneutics could potentially offer to further our understanding of such a unique collection – a collection, moreover, that has existed in various iterations and guises over the last three decades, during various moments of media technological change.

30 Hesselberth, *Cinematic Chronotopes*, 13.

31 Drucker, *Visualization and Interpretation*, 5.

32 Drucker, *Visualization and Interpretation*, 7.

33 Slootweg, "Resistance, Disruption, Belonging," 214.

34 Lisa Lebow, "Introduction," in *The Cinema of Me: The Self and Subjectivity in First-Person Documentary Film*, ed. Lisa Lebow and Michael Renov (New York: Columbia University Press, 2012), 1.

Layering the Cake with Digital Hermeneutics

So, then, how should we assess the Sullivan collection? As indicated earlier, we doubt whether it is fruitful or accurate to identify his work as "vlogging," although "pre-vlogging" may be a preliminary, less anachronistic manner to approach his work. We do however acknowledge, as so many in the comments section of his videos on YouTube seem to suggest, that his work showcases a mode of cultural expression and medial communication that was ahead of its time; yet, at the same time, Sullivan's work, attitude, and aesthetics can also be regarded as a 1980s exponent of earlier and parallel media practices in New York City and beyond.[35] When studying Nelson Sullivan's videographic work, it thus becomes clear that its multifaceted qualities present both opportunities and challenges from a traditional hermeneutic point of view, as well as from the perspective of computational approaches.

Pertaining to the latter, computational approaches become useful whenever the volume of data exceeds a certain number that is impossible for humans to cope with, or when the material used allows for computational analysis due to its digitality or retroactive digitization. It is only then that computational approaches can, for example, produce visualizations, or enable other forms of distant reading that can serve as a first step in dealing with the collection. Moreover, it can help in identifying instances of interest, but should always be seen as "a primary mode of knowledge production, not a secondary expression of preexisting data."[36] The distinction or similarities between data and sources have been discussed before. As Lisa Gitelman and Virgina Jackson pointed out, once original sources are transformed into machine-readable data, a shift occurs from historically and materially contextualized information into "individual, separate and separable" data.[37] For instance, it is important to reiterate that video has visual and auditory dimensions, as is the case with many other audiovisual media. These aspects can be studied separately, but, ultimately, they need to be analyzed in relation to each other.

In addition, we need to not only consider the content or style of the videos, but also the socio-material embeddings of the data, which in our case study will mean that we must understand the various historical contexts in terms of changing distribution and viewing contexts of the footage. In this case, it is pertinent to understand that the original videos became available online almost two decades after they were produced. In addition, it is crucial to know that these videos were

35 See, for example, Ruoff, "Home Movies and the Avant-Garde."

36 Drucker, *Visualization and Interpretation*, 2.

37 Lisa Gitelman and Virgina Jackson, "Introduction," in *"Raw Data" is an Oxymoron*, ed. Lisa Gitelman (Cambridge, MA: MIT Press, 2013), 8.

uploaded on YouTube in slightly revised formats, but with additional titles added by the channel owners.[38]

For all these reasons, we believe that a multi-layered digital hermeneutics is crucial, one that allows a more data-driven exploratory step, but also allows for a contextual reading. N. Katherine Hayles has also underlined the methodological advantage of digital hermeneutics as a form of scaled reading, moving between close, hyper, and machine reading.[39] Inge van der Ven and Tom van Nuenen extended on this by introducing another level, namely "platform hermeneutics," to foreground the role of platforms or media environments in data production, distribution, and reception.[40] If this can indeed be achieved with a full hermeneutic circle, aligning computational methods with more traditional hermeneutics, we believe that a richer and deeper understanding of the data we are exploring can be achieved. In what follows, we will start with platform hermeneutics as a first contextual step to distant reading, followed by hyper reading, and then close reading before we close the hermeneutic cycle.

Platform Hermeneutics

For our case study, we selected six from the collection of videos that were uploaded on YouTube by Dick Richards between 2008 and 2018. We selected these videos based on their titles (see Videography below). As explained elsewhere, we were particularly interested in the "I" (i.e., the enunciator, or filmmaker) and how this is performed through some form of medial self-reflexivity. We therefore selected videos whose titles made some reference to either the media technologies or the "self." As our first step, we engaged in a platform hermeneutics of YouTube and the channel. Van der Ven and van Nuenen explain that "platform hermeneutics" enables a focus on the architecture of the websites studied and which modes of sociality are afforded by those websites.[41] Doing platform hermeneutics, they argue, is a necessary step to ground and contextualize data before

38 To fully grasp the scope and extent of the changes made to Sullivan's original videos would require an in-depth comparative analysis of the content on the channel and the tapes held at the Fales Library. Such an analysis lies beyond the scope of this contribution.

39 N. Katherine Hayles, "How We Read: Close, Hyper, Machine," *ADE Bulletin* 150 (2010): 62–79, https://doi.org/10.1632/ade.150.62.

40 Inge van der Ven and Tom van Nuenen, "Digital Hermeneutics: Scaled Readings of Online Depression Discourses," *Medical Humanities* 48 (2022): 335–346, https://doi.org/10.1136/medhum-2020-012104.

41 Van der Ven and van Nuenen, "Digital Hermeneutics," 339.

their eventual decontextualization in subsequent steps. This fits well with Johanna Drucker's notion of "frame analysis," which implies that, for instance, a selection of uploaded videos requires an act of reading and interpretation that takes into consideration their digital or networked environment.[42] Thus this approach does not only fit well with textual analysis, but also touches on theories of multimodality, where mediated expressions are approached as practices in which meaning making occurs in various semiotic modes that are "socially shaped and evolve to meet the communicative needs of a community."[43]

In our case, what were the communicative needs and semiotic modes of Sullivan's work after taking up the video camera in the early 1980s? And what about his habit of turning his camera 180 degrees to engage in what we would nowadays term a form of "selfie aesthetics"?[44] We are interested in how the specific architecture and affordances of YouTube became instrumental in building a queer archive that also became meaningful to contemporary users. Although his body of work came into being well before the rise of YouTube, it has gained a new relevance to a new generation of users via this platform. To understand this historical trajectory, we need to adopt a media historical approach to chart how this kind of grassroots archivalization occurred across various media platforms and landscapes.

So, despite YouTube's status as the main point of entry for a dedicated community of people who are appreciative of Sullivan's work, it is important to acknowledge that it was certainly not its first distribution or viewing platform. The YouTube collection contains edited versions of Super8 footage (mostly from the 1970s), and videos captured directly on VHS with a separate recorder and camera (first half of the 1980s), which were later replaced by Video8 footage shot on a camcorder (second half of the 1980s). Initially, Sullivan shared his materials with friends who featured in his videos by copying the tapes. These were typical ways of doing things in the alternative media ecosystems of the 1980s and 1990s, when public access to established mass media outlets was still rather limited. Content creators, but also fans, started exploring the new affordances of the video re-

42 Drucker, *Visualization and Interpretation*, 101.

43 Tuomo Hiippala, "Distant Viewing and Multimodality Theory: Prospects and Challenges," *Multimodality & Society* 1, no. 2 (2021): 139, https://doi.org/10.1177/26349795211007094. See also: John A. Bateman, "The Decomposability of Semiotic Modes," in *Multimodal Studies: Multiple Approaches and Domains*, ed. Kay L. O'Halloran and Bradley A. Smith (London: Routledge, 2011), 12–38; Elisabetta Adami and Gunther Kress, "Introduction: Multimodality, Meaning Making, and the Issue of the 'Text'," *Text & Talk* 34, no. 3 (2014): 231–237, https://doi.org/10.1515/text-2014-0007.

44 Nicole Erin Morse, *Selfie Aesthetics: Seeing Trans Feminist Futures through Self-Representational Art* (Durham, NC: Duke University Press, 2022).

corder and began to copy, circulate, and preserve via videotaping, extending on what Lisa Gitelman termed the "cultural logic of the copy."[45]

Because of this logic, we surmise that many active users, intentionally or unintentionally, have built "personal non-canonical archives."[46] Although the outreach of Sullivan's work during his life was limited and stayed within local media outlets and videotape copy culture in local communities in NYC and Atlanta, after his sudden death in 1999, Sullivan's work gained more and more attention. According to Terrell, Sullivan was quite aware of the value of his video archive and optimistic that it would continue to find an audience.[47] Indeed, this is what happened when his friend Dick Richards adopted his videotapes and eventually decided to digitize them.

The YouTube channel initiated a transformational re-appraisal of Sullivan's work. 5ninthavenueproject provided an opportunity to connect with a new generation of enthusiasts as the affordances of YouTube seemed to work rather well as an open video sharing platform that offered easy access for uploaders and viewers alike, and thus went beyond the limitations of the media distribution and viewing of the 1980s and 1990s. What is more, by the time the channel started, YouTube had gained both a notorious yet also optimistic reputation as the place to be for grassroots DIY videos as well as archival initiatives.

In 2008, YouTube had been around for three years and developed from a video repository, which facilitated easy uploading and sharing of moving images, into a platform that many of its users regarded as a culturally significant audiovisual archive. As Jean Burgess and Joshua Green noted, new models of media entrepreneurship became embedded in what they describe as YouTube's grassroots culture.[48] As pointed out by Rick Prelinger, the prompt assessment of this social media platform as a "default archive" has disrupted the exclusivity of scholarly access via institutional archives, and transformed everyday conceptions of archives.[49] Annamaria Motrescu-Mayes and Susan Aasman elaborated on this idea and foregrounded the much more fluid, non-hierarchical, and pluralistic nature of current digital archives.[50]

---

45 Lisa Gitelman, *Paper Knowledge: Toward a Media History of Documents* (Durham, NC: Duke University Press, 2014).

46 Lucas Hilderbrand, *Inherent Vice: Bootleg Histories of Videotape and Copyright* (Durham, NC: Duke University Press, 2009).

47 Terrell, "About Last Night."

48 Jean Burgess and Joshua Green, "The Entrepreneurial Vlogger: Participatory Culture Beyond the Professional-Amateur Divide," in *The YouTube Reader*, ed. Pelle Snickars and Patrick Vonderau (Stockholm: National Library of Sweden, 2009), 90.

49 Rick Prelinger, "The Appearance of Archives," in *The YouTube Reader*, ed. Pelle Snickars and Patrick Vonderau (Stockholm: National Library of Sweden, 2009), 269.

50 Motrescu-Mayes and Aasman, *Amateur Media and Participatory Culture*, 135.

It took Richards around ten years to upload the more than 700 Sullivan videos. The original videotapes, but also some Super8 films from the 1970s, were digitized and re-edited. To each video, a title was added, including the name of the maker. Reactions in the comment section demonstrate the deep-felt appreciation for this type of material, whether it is by people from the queer community and club scene of the 1980s, or by younger generations enthralled by the cheerful exuberance of that period. Although the collection has been a closed corpus since Richards passed away in 2018, the videos remain accessible and popular. In addition, the channel's popularity continues to stimulate users to add material; old friends upload their copies of Sullivan's tapes, or fans re-use the footage in short documentaries. It shows the unique traits of platforms such as YouTube offering a much more dynamic concept of corpus or collections.[51] A recent development is the circulation of shortened clips taken from the YouTube videos on other social media platforms, such as TikTok and Instagram.

The trajectory of this video collection, in terms of changing formats (VHS, DVD, streaming video), platforms (copy culture, public access TV, cable TV, and YouTube), and audiences, illustrates various forms of DIY-curation and archivalization, mainly occurring outside cultural heritage institutions. Abigail De Kosnik conceptualizes this as the formation of "rogue archives," or counter initiatives that represent an entirely new model of online archiving, one that entails collecting data in an idiosyncratic manner.[52] Alternatively, Katie Eichhorn has explored how, traditionally, archives were sites and repositories devoted to starting knowledge production by stimulating new interpretations, especially connected to history. She observes that digital DIY-archivalization "unmoors subjects from the historical moment" and in that process new opportunities emerge for cultural materials that were produced by hitherto marginalized groups.[53] Hence, this can lead to authorizing and legitimizing voices that would otherwise remain unheard: once they entered the digital "archive," they became "social agents outside their designated time."[54]

In summary, the corpus we are looking at – the YouTube collection as our unit of observation – is a mix of various historical media formats: Super8, VHS, and Video8. These productions were originally created between the 1970s and the

51 We should, however, always be aware of YouTube's instability as an archive: the material may be taken down any time, by the original uploader or by YouTube. This underscores the importance of the Fales Library in preserving the original films and videos.

52 Abigail De Kosnik, *Rogue Archives: Digital Cultural Memory and Media Fandom* (Cambridge, MA: MIT Press, 2016), 2.

53 Katie Eichhorn, *The Archival Turn in Feminism: Outrage in Order* (Philadelphia: Temple University Press, 2013), 79.

54 Eichhorn, *The Archival Turn in Feminism*, 80.

late 1980s, but, on YouTube, the various "regularities" in the semiotic modes of these historical media were revitalized and recognized in a new media ecosystem, with a new community of users.[55] To make some headway, we will continue to explore these "regularities" in the semiotic modes through different scales of *reading*, starting with the visual exponent of distant reading: distant *viewing*.

Distant Viewing

The entirety of Sullivan's YouTube collection makes up for a relatively small "big data" set. The channel contains around 700 videos, which contains in total some 2,500 hours of video, which can be split into millions of frames. Apart from this, there is dialogue, sounds, music, and on-screen titles, but also metadata such as URL IDs, video descriptions, keywords, tags, titles, view and like counts, and sometimes hundreds of comments per video. For the sake of this discussion, we focused on the content, and selected videos to explore the opportunities and challenges of distant viewing. Distant viewing is a term that is a variation of the famous concept "distant reading" that was introduced by Franco Moretti.[56] He used the term to go beyond the close reading of only a few canonized literary texts, by using quantitative methods as a strategy to reverse the traditional hierarchy of privileged texts and instead focus on average published works.

Although Moretti's distant reading, as an antidote to close reading, has become both influential and controversial, it did have a significant impact on what Lev Manovich calls the study of cultural data at scale.[57] Manovich introduced cultural analytics to conceptualize the use of computational methods in the analysis of massive cultural datasets and flows. The goal of extracting patterns and trends took hold, and variations in distant reading became an important approach in the digital humanities. When the focus shifted from written text to also include visual material, the term cultural analytics – finding patterns through data mining of large datasets – gained attention. Manovich explicitly aimed to keep visual dimensions in his analytical scope, rather than focusing solely on metadata. His use of visualization techniques, such as generating image grids, offered an alternative to distant "reading" of textual data.

It furthermore opened opportunities for "distant viewing," a term that was expanded upon by Arnold and Tilton to conceptualize and apply computational

55 See, for example, Bateman, "The Decomposability of Semiotic Modes," 23.

56 Franco Moretti, *Distant Reading* (London: Verso, 2013).

57 Manovich, *Cultural Analytics*, 30.

methods to large-scale datasets of images.[58] They understand "viewing" as an interpretative act by either a person or a computer model.[59] Based on training datasets, models are developed that extract features found in images. Arnold and Tilton further developed distant viewing from the context of media history by focusing on audiovisual data, which is even more complex because of the temporal dimension (although neglecting sound as another layer of meaning making). For this case study we used the open-source data mining tool Orange to analyze video images.[60] With this tool, which uses machine learning and data visualization, it is possible to perform data analysis workflows by using widgets that can process, model, and visualize data. This makes it relatively easy to perform complex tasks such as image analysis without requiring programming skills. According to its developers, the tool has a "democratizing potential."[61] To perform image analytics, Orange makes use of pre-trained deep convolutional networks such as Inception v-3, VGG-16, and VGG-19 to profile images with vectors of features. Based on those vector representations, it is possible to cluster images and make classifications. The toolbox is quite helpful for non-experts to experiment with various forms of clustering and do explorative inspection of the data. However, the tool is modest in that its main goal is to assist in the "analysis of smaller image sets where the starting point is image embedding using a pre-trained deep network."[62]

To be able to upload our selected videos in the Orange tool, some preprocessing was needed, transforming the video files into a sequence of images. With the help of a freely available converter tool, the video frames were extracted according to a preselected image–recording ratio. This first step entailed a serious reduction of the dataset, as we now only worked with a short selection of still images, not only leaving out many frames, but also leaving out temporality and sound. The result for one individual video resulted in the clustering of images in Figure 1.

In the next step, we upscaled this process by converging six folders of frames related to the selected videos, by combining them into one image grid. The Orange distant viewing tool regrouped all frames from the six videos according to

58 The term distant viewing was already introduced in 2015 by art historian K. Bender to describe a quantitative approach to art history. See K. Bender, "Distant Viewing in Art History: A Case Study of Artistic Productivity," *International Journal for Digital Art History* 1 (2015): 101–110, https://doi.org/10.11588/dah.2015.1.21639.

59 Arnold and Tilton, "Distant Viewing," i4.

60 Orange Data Mining, https://orangedatamining.com/, accessed February 1, 2023.

61 Primož Godec et al., "Democratized Image Analytics by Visual Programming Through Integration of Deep Models and Small-Scale Machine Learning," *Nature Communications* 10, no. 1 (2019): 2, https://doi.org/10.1038/s41467-019-12397-x.

62 Godec et al., "Democratized Image Analytics," 6. See also Wevers and Smits, "The Visual Digital Turn."

compositional similarities, which offers the researcher a quick overview of multiple images. The neural network was able to make a convincing overview that helped to identify patterns. For instance, it was interesting to note that there is not much variety, but a fairly stable set of similar compositions. Because the model analyzes the uploaded frames as one dataset, it does not discriminate based on other aspects such as narrative, chronology of the videos, etc. This advantage comes with some great challenges as well because the original temporal integrity of the videos is negated. The narrative is excluded, but editing, time duration, and audio are not part of this analysis either.

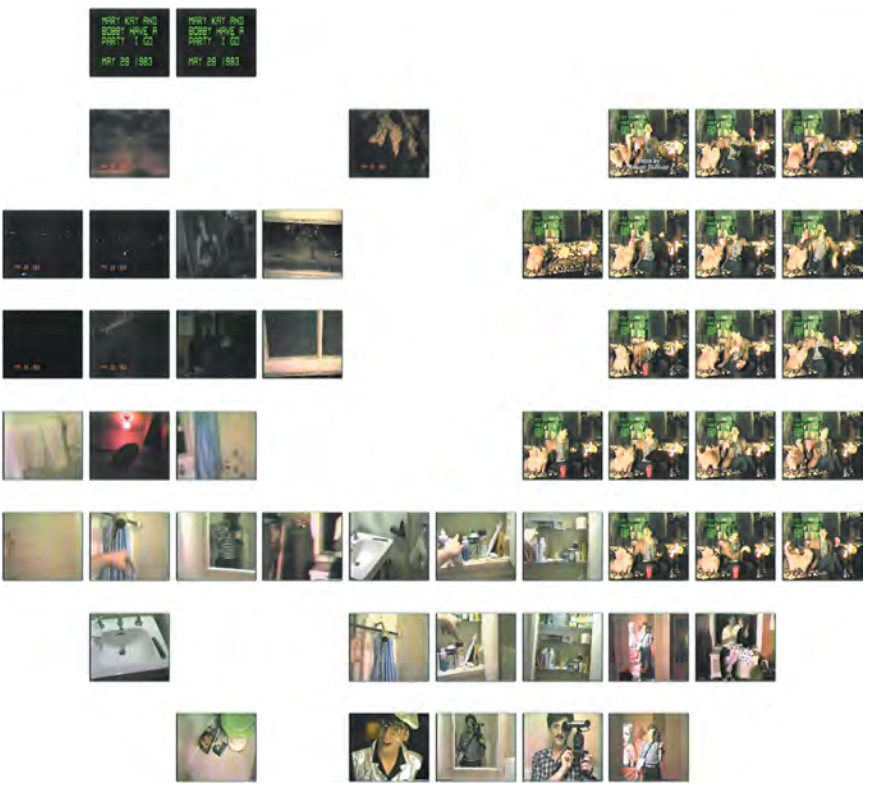


Figure 1: Image grid, based on VGG-19, from the video "Nelson Sullivan's First Video Blog in 1983," https://www.youtube.com/watch?v=8Qc9sISu5yc.

In addition, on the level of the corpus, the chronology is lost as well in this container of images. The image grid classifies based on similarity, and this does not automatically reflect a potential historical development in Sullivan's style. Of

course, this can be remedied by making the procedure into an iterative process; for instance, we can make new selections year by year and, through hierarchical clusters, tease out statistics, which would then show an increasing number of close ups of Nelson Sullivan occupying the screen.

Hyper Reading

Other tools can be helpful as well. For example, the Distant Viewing Toolkit, developed by Taylor Arnold and Lauren Tilton, can identify shot length, and can also identify the same person in a series of videos.[63] Other software can be used for automatic speech recognition and identifying linguistic patterns. Ultimately, the various tools, each with its own strength in producing relational data, can be put into use for a second level of analysis, namely hyper reading. It was Hayles who introduced the term to describe nonlinear, screen-based, and computer-assisted modes of reading, including search queries, skimming, and scanning. Hyper reading allows for a more intuitive and associative relationship with data, which can help to trace specific points of interest. Hayles regards hyper-reading as a way "to identify passages or to hone in on a few texts of interest, whereupon close reading takes over," while also overlapping with machine reading in the quest for identifying patterns.[64] According to Hayles, hyper reading bridges machine and close reading, because it "allows the user both to see large-scale patterns and to zoom in to see a particular cover in detail, thus enabling analyses across multiple scale levels."[65]

So far, as previously explained, the output of visualizations can be reductive, yet also highly productive for hyper reading. Skimming through the image grids reveals some interesting image clustering that evokes rather interesting associations. However, before we discuss some of these associations, we would first like to point out that this research step resembles what Jentery Sayers has termed "a liminal space between standardized and experimental practice, where the consequences remain uncertain."[66] In other words, regardless of the data's status as an

63 Distant Viewing Toolkit, https://github.com/distant-viewing/dvt, accessed February 10, 2023.

64 Hayles, "How We Read," 74.

65 Hayles, "How We Read," 77. The various uses of terms such as machine reading, distant reading, and distant viewing, or, for that matter, hyper and close reading, are still not settled. Distant viewing works for large scale analysis of visual material, but hyper and close analysis also allow other modalities to be taken into account.

66 Jentery Sayers, "Bringing Trouvé to Light: Speculative Computer Vision and Media History," in *Seeing the Past with Computers: Experiments with Augmented Reality and Computer Vision for History*, ed. Kevin Kee and Timothy Compeau (Ann Arbor: University of Michigan Press, 2019), 37.

important historical trace of marginalized communities of the 1980s, we believe that we should nevertheless be thoughtful of how hyper reading, through computer vision, can potentially open "a variety of ways to perceive, engage, and question the stuff of history," but without forgetting that it might also "be simultaneously procedural and subjunctive."[67] This becomes even more relevant when using algorithms that could potentially detect and cluster biometric features, such as faces, gender, ethnicity, and so on. Sayers rightfully warns about potential racial bias, while others have pointed at the broader problems with how algorithms deal with diversity and inclusion.[68]

With this important parenthesis in mind, we return to our material. Our associative skimming of the material has revealed a plethora of clusters that feature close-ups of Sullivan, with some remarkable contextual differences. We see Nelson with a camera, and also without a camera, and we see endless close-ups of him in a car. In one of the first videos, made in 1983, Sullivan explicitly introduced himself onscreen by pointing the camera towards himself (Figure 2, top). To include himself in a long take, he used a mirror to make this happen.[69]

As the history of filmmaking has shown, this aesthetic choice is typical for many amateurs, but also for the earlier discussed documentary filmmakers who engaged in first person and autobiographical filmmaking. The clunky video camera, plus the videotape recorder hanging over the shoulder, has become an iconic image of the early videographer and first-person filmmaker. At the end of the decade, Sullivan is continuously visible in the videos, but this time without any sign of the camera (Figure 2, bottom). This is, as we highlighted before, due to Sullivan's decision to make use of a fish-eye lens on his Video8 camcorder. The newly acquired equipment allowed for different ways of doing things, which immediately affected the communicational aesthetics and clustering of the stills processed in Orange. In the next section, we will take these findings as a next step in the process of closing the circle of our experiment in digital hermeneutics.

67 Sayers, "Bringing Trouvé to Light," 40.

68 See, for example, Eduard Fosch-Villaronga and Adam Poulsen, "Diversity and Inclusion in Artificial Intelligence," in *Law and Artificial Intelligence*, ed. Bart Custers and Eduard Fosch-Villaronga (The Hague: T. M. C. Asser Press, 2022), 109–134; Anna Lauren Hoffmann, "Terms of Inclusion: Data, Discourse, Violence," *New Media & Society* 23 (2021): 3539–3556, https://doi.org/10.1177/1461444820958725.

69 The title of the YouTube video is "Nelson Sullivan's First Video Blog in 1983," but the video itself has the onscreen title: "Mary Kay and Bobby have a Party. I go. May.28 1983," https://www.youtube.com/watch?v=8Qc9sISu5yc, 7:29.

Figure 2: Stills from "Nelson Sullivan's First Video Blog" and "Nelson Sullivan explains his camcorder technique to Lahoma" on the YouTube channel "5ninthavenueproject".

Close Reading

After the process of hyper reading, where we identified several remarkable frames and clusters, close reading facilitates a recontextualization of the data by considering again those aspects that were neglected for the sake of distant read-

ing. By this we mean that: (1) the frames taken from the videos will be reconnected to other modalities such as sound and speech; (2) a recontextualization of the frames within the individual video's narrative but also the chronology of the corpus is required; and (3) a reconsideration of their meaning in relation to its original context of production and reception in the 1980s can take place, while simultaneously taking into account the new modes of distribution and reception on YouTube, almost two decades later.

Close reading can be supported by digital tools, such as video annotation. Within film and media studies, multiple-tier annotation can provide layered descriptions based on variables such as shot-length, speech, or camera movements but can also add descriptions that go beyond neatly defined categories. Tiers afford groupings of tags which can be used for a more complex and interrelated analysis.[70] There are quite a few video annotation tools available for this. In our case, the ELAN tool was used to set-up a multimodal analytical framework containing elements of Point of View (POV), interaction, framing, and sound, adding up to a total of eight tiers (Figure 3).[71] In this stage of our multilayered digital hermeneutics approach, it is especially important to identify the deictic categories "I," "you," and "s/he" in both spoken words and images. Although this approach was initially used within communication studies, here we would like to recommend taking steps to integrate such methodologies into media historical research as well.

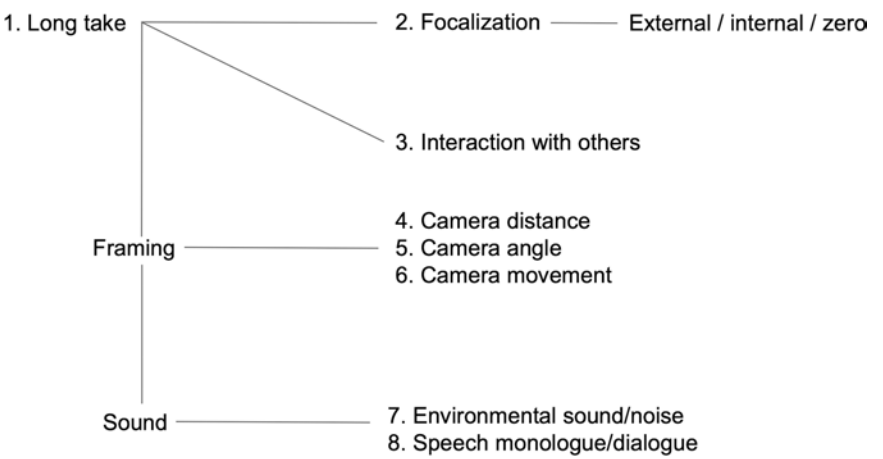


Figure 3: The formalized scheme to structure a close reading.

70 See, for example, Aasman et al., "Tales of a Tool Encounter," 7.

71 ELAN (Version 6.4), Computer software, 2022. Nijmegen: Max Planck Institute for Psycholinguistics. Retrieved from https://archive.mpi.nl/tla/elan.

As described previously, close reading aims to contextualize the individual and aggregated frames by connecting them to the original audio tracks, which contain various dimensions such as speech (dialogue) and noise. This presents a challenge for more quantitative or data-driven qualitative approaches, and thus requires further reflection. As previously observed in our hyper reading, it appears that the selfie style of Nelson Sullivan gradually emerged in his collection. Initially, Sullivan needed a mirror to create a self-portrait, but he was able to acquire new equipment, the camcorder, which was far less bulky and easier to use. In addition, he opted to use a fish-eye lens, which gave a wide-angle scope to record himself and his environment.

Media historical research tends to focus on the effects of these techniques in visual terms, but we argue that speech and (environmental) sound are just as important. With the help of video annotation, the visual and the auditory can come together in a multimodal analysis. For instance, in Sullivan's video "39th Birthday – Waiting for Eric and Liz" (1987?), he recorded himself sitting outside in front of his house, waiting for his friends, with only the camera as his companion. While capturing the street, he suddenly pans and turns the camera 180 degrees. What started out as a shot pointed at the street, ended as a selfie-shot. As he makes the idiosyncratic camera movement, Sullivan narrates:

> I got to learn how to do this. How to point the camera at myself. This is my birthday resolution. Thirty-nine-years-old today, and today I'm going to learn to point the camera at myself. Talk to it . . . Like, I guess, you know, I should talk to it, like there is some information I'm trying to get across to it.

It is in the combination of these two modalities that we can capture important shifts in Sullivan's development as a videographer. His explorations in f/vocalizing his perspective not only occurs through the lens, but works in tandem with the conversation with himself ("I") and the viewer (an implied "you"). The explorative use of the camera thus goes hand in hand with a dialogue with the viewer, in itself a typical feature of first-person cinema since the 1980s and 1990s.

Sullivan kept experimenting. For instance, in 1989 he bought a new camcorder to record his video "Nelson Sullivan and his new Camcorder – Part 1." On a car trip from Atlanta to his hometown, he experimented with using the camera in the car, pointed at himself and the driver (his friend Dick); the wide-angle lens allowed both men to be framed in the shot at the same time. While shooting, Sullivan reflected on the novelty of the process of being in front of the camera. He questioned whether he should talk to the camera as if it were a person, or simply announce his plans, or say nothing at all. These instances of self-reflection about learning the craft of self-portraiture on-camera are historically situated phenomena in both amateur-made video diaries and professional autobiographical documentaries, but

also became pervasive in many vlogs made in the twenty-first century. Choices related to how to address the camera/audience, through sound and image, speak to new generations of video users.

As Michael Renov noted, from the 1970s until the 1990s, "documentary explorations of gay and lesbian identities have exhibited a particular dynamism and vitality."[72] Of particular importance here is the observation that many of these documentaries did not adhere to a specific formal template. This also applies to Nelson Sullivan. Over the years, Sullivan refined his style, focusing on embodied, mobile, and unscripted modes of videography, capturing moments of movement and street life with friends. This offered new ways of mediation that countered prevailing heteronormative representational styles in mainstream media by reconfiguring notions of belonging, home, self, family, and community.[73] These alternative media practices thus ran in tandem with the emerging trend in documentary filmmaking, identified by Renov, to enhance the voice and visibility of marginalized communities. Not only the visual aspects of Sullivan's deictic approach can be situated in the changing dynamics of documentary filmmaking and event videography: the use of sound also signifies new ways of representing the self and others. As Jay Ruby noted, this approach to "subject-generated" content emphasizes yet another aspect of cooperation and collaboration in media that explores new ways of not only speaking about communities, but also speaking for, with, and alongside them.[74]

Through our close reading it became clear that Sullivan introduced some interesting approaches in exposing his life and his thoughts, while also connecting to his community of friends, as well as an assumed audience. He employed experimental tactics for the innovative use of consumer media technologies, which allowed him to mediate himself, his immediate social circles, and the spaces they inhabited in new ways. He went on a trajectory from using mirrors and passing on the camera to combining a fish-eye lens with his camcorder to achieve what we would nowadays call a selfie mode. Through these DIY-practices, Sullivan was consistently seeking to present himself and his companions in his videos. Only later in his videographical work, when he had fully mastered the use of wide-angle shots, was Sullivan able to merge his environment and point of view with that of others. In conclusion, it could thus be argued that in these instances the "I," "you," and "s/he" became highly fluid and dynamic deictic categories.

---

72 Michael Renov, *The Subject of Documentary* (Minneapolis: University of Minnesota Press, 2004), 180.

73 Moran, *There's No Place Like Home Video*, 47.

74 Jay Ruby, "Speaking For, Speaking About, Speaking With, or Speaking Alongside: An Anthropological and Documentary Dilemma," *Visual Anthropology Review* 7 (1991): 42, https://doi.org/10.1525/var.1991.7.2.50.

Closing the Circle with Units of Analysis, Interest, and Observation

In this chapter we investigated the potential of digital hermeneutics in excavating the media historical trajectory of a highly evocative video collection that was originally produced in the 1980s to resurface with force on YouTube in the twenty-first century. As we have shown, the materials produced by Nelson Sullivan showcase a complexity that was astounding for its day and age but has advanced in significance during its digital "afterlife." Not only did we delve deeper into the content and style, resembling many features of what we would nowadays term vlogging, we also highlighted that many of the same aspects can similarly be situated in media practices of the last quarter of the twentieth century, such as event videography, first person film, and autobiographical documentaries.

By adopting digital hermeneutics, with various heuristic levels, as our main approach, we furthermore hope to have contributed to debates on the innovations and skills necessary to adapt to changing epistemological circumstances due to the rise of computer-assisted research in film and media historiography. We explored an analytical framework that meaningfully brings together distant, hyper, and close reading as an iterative process within a broader context of platform hermeneutics. In so doing, we made a first step towards synthesizing, rather than pitting against each other, the variety of fruitful approaches that have emerged in digital history and digital humanities over the last decade and a half.

With N. Katherine Hayles, we believe that it can be a source of experimentation for researchers and students to work with programs that are easy to use: "machine analysis opens the door to new kinds of discoveries that were not possible before and that can surprise and intrigue scholars accustomed to the delights of close reading."[75] However, where Hayles refers to the textual, we mainly dealt with audiovisual data, consisting of movement, cinematography, sound, dialogue, music, editing, narratives, etc. To make the data work for computational methods, the inherent multimodal complexity requires simplification, which commonly results in either a focus on the visual or the auditory. It is therefore important to carefully draft a research design that takes into consideration the ideal unit of analysis for each stage of the research.

We would like to conclude this chapter by reflecting on a methodological issue concerning units of analysis, interest, and observation. In our study, we identified the collection of videos uploaded on 5ninthavenueproject as our primary top-level "unit of observation." Although we were aware that this collection

75 Hayles, "How We Read," 78.

was a modified version of Sullivan's original work, it represented the current community's interest in his work as reflected in the available video corpus. Future research could expand this collection to include other units of observation, such as the reuse of Sullivan's footage in DIY YouTube documentaries, TikTok shorts, or Instagram posts.

For our distant reading approach, we chose to focus on a selection of frames as the most effective way to utilize computational tools. Reducing the analysis to these smaller units is a standard practice when deploying computer vision. However, this reduction comes at the cost of disregarding the richness of the materials in terms of narrative, temporal, spatial, deictic, and other significant aspects. Determining the unit of analysis becomes even more complex when applying tools that enable automated speech recognition, especially when combined with visual components. For example, how should we define various forms of speech as units of analysis? How do they determine a scene, or a long take? In Sullivan's work, dialogue can accompany long takes or short scenes. Communication and multilingualism studies, which similarly involve video analysis, have explored various solutions, suggesting some flexibility by considering "mini-scenes."[76] Alternatively they deploy conversations as a unit of analysis – an intermediate unit smaller than a scene, but more meaningful than an individual utterance or instance of speech.[77] This is where a clear conceptual choice about what to focus on is essential, as in our case where "deixis" became helpful to keep our unit of analysis true to the multi-layered, multimodal, and multi-f/vocal complexity of our object of study.

In summary, audiovisual materials present a complexity that necessitates theoretical, practical, and functional solutions. We suggest that digital hermeneutics encourage researchers to delve into the specificities of the object under study to identify the appropriate units of analysis or to redefine them as "units of interest" in a multi-layered process, considering the most suitable instances of multimodal expressions. As Nelson Sullivan himself so eloquently observed in one of his last videos, the multifaceted nature of his footage is "pure information, not the real thing."[78]

76 Anthony Baldry, "Multisemiotic Transcriptions as Film Referencing Systems," *InTRAlinea: Online Translation Journal* (2016), accessed April 20, 2023, https://www.intralinea.org/specials/article/2195.

77 See, for example, Patrick Zabalbeascoa and Montse Corrius, "Conversation as a Unit of Film Analysis: Databases of L3 Translation and Audiovisual Samples of Multilingualism," *MonTI: Monografías de Traducción e Interpretación* 4 (2019): 57–85, https://doi.org/10.6035/MonTI.2019.ne4.2.

78 5ninthavenueproject, "Nelson Sullivan Explains his Camcorder Technique to Lahoma," uploaded on YouTube February 4, 2009, 2:04.

References

Aasman, Susan. "Unlocking Multiple Histories of Amateur Media: From Micro- to Macro-Histories." *Screen* 61, no. 1 (2020): 157–166. https://doi.org/10.1093/screen/hjaa011.

Aasman, Susan, Liliana Melgar Estrada, Tom Slootweg, and Rob Wegter. "Tales of a Tool Encounter: Exploring Video Annotation for Doing Media History." *VIEW Journal for European Television History and Culture* 7, no. 14 (2018): 73–87. https://doi.org/10.18146/2213-0969.2018.jethc154.

Aasman, Susan, Tim van der Heijden, and Tom Slootweg. "Amateurism: Exploring Its Meaning in the Age of Film, Video and Digital Media." In *Digital Roots: Historicizing Media and Communication Concepts of the Digital Age*, edited by Gabriele Balbi, Nelson Ribeiro, Valérie Schafer, and Christian Schwarzenegger, 245–265. Berlin: De Gruyter, 2021.

Adami, Elisabetta, and Gunther Kress. "Introduction: Multimodality, Meaning Making, and the Issue of the 'Text'." *Text & Talk* 34, no. 3 (2014): 231–237. https://doi.org/10.1515/text-2014-0007.

Armes, Roy. *On Video*. London: Routledge, 1988.

Arnold, Taylor, and Lauren Tilton. "Distant Viewing: Analyzing Large Visual Corpora." *Digital Scholarship in the Humanities* 34, issue supplement 1 (2019): i3–i16. https://doi.org/10.1093/llc/fqz013.

Baldry, Anthony. "Multisemiotic Transcriptions as Film Referencing Systems." *InTRAlinea: Online Translation Journal* (2016). Accessed April 20, 2023, https://www.intralinea.org/specials/article/2195.

Bateman, John A. "The Decomposability of Semiotic Modes." In *Multimodal Studies: Multiple Approaches and Domains*, edited by Kay L. O'Halloran and Bradley A. Smith, 12–38. London: Routledge, 2011.

Belton, John. "Looking Through Video: The Psychology of Video and Film." In *Resolutions: Contemporary Video Practices*, edited by Michael Renov and Erika Suderberg, 61–72. Minneapolis: University of Minnesota Press, 1996.

Bender, K. "Distant Viewing in Art History: A Case Study of Artistic Productivity." *International Journal for Digital Art History* 1 (2015): 101–110. https://doi.org/10.11588/dah.2015.1.21639.

Boyle, Deidre. "From Portapak to Camcorder: A Brief History of Guerrilla Television." *Journal of Film and Video* 44 (1992): 67–79.

Buckingham, David, Rebekah Willet, and Maria Pini. *Home Truths? Video Production and Domestic Life*. Ann Arbor: University of Michigan Press and University of Michigan Library, 2011.

Burgess, Jean, and Joshua Green. "The Entrepreneurial Vlogger: Participatory Culture Beyond the Professional-Amateur Divide." In *The YouTube Reader*, edited by Pelle Snickars and Patrick Vonderau, 89–107. Stockholm: National Library of Sweden, 2009.

Casetti, Francesco. *Inside the Gaze: The Fiction Film and its Spectator*. Bloomington: Indiana University Press, 1999.

Cubitt, Sean. *Timeshift: On Video Culture*. London: Routledge, 1991.

De Kosnik, Abigail. *Rogue Archives: Digital Cultural Memory and Media Fandom*. Cambridge, MA: MIT Press, 2016.

Deleon, Joseph. "Nelson Sullivan's Video Memories: YouTube Nostalgia and the Queer Archive Effect." *The Velvet Light Trap* 86 (2020): 16–26. https://doi.org/10.7560/VLT8603.

Dovey, Jon. *Freakshow: First Person Media and Factual Television*. London: Pluto Press, 2000.

Drucker, Johanna. *Visualization and Interpretation: Humanistic Approaches to Display*. Cambridge, MA: MIT Press, 2020.

Eichhorn, Katie. *The Archival Turn in Feminism: Outrage in Order*. Philadelphia: Temple University Press, 2013.

Fickers, Andreas, Juliane Tatarinov, and Tim van der Heijden. "Digital History and Hermeneutics. Between Theory and Practice: An Introduction." In *Digital History and Hermeneutics: Between Theory and Practice*, edited by Andreas Fickers and Juliane Tatarinov, 1–19. Berlin: De Gruyter Oldenbourg, 2022.

Fosch-Villaronga, Eduard, and Adam Poulsen. "Diversity and Inclusion in Artificial Intelligence." In *Law and Artificial Intelligence*, edited by Bart Custers and Eduard Fosch-Villaronga, 109–134. The Hague: T. M. C. Asser Press, 2022.

Gerbaudo, Paolo. "From Data Analytics to Data Hermeneutics: Online Political Discussions, Digital Methods and the Continuing Relevance of Interpretive Approaches." *Digital Culture and Society* 2 (2016): 95–111. https://doi.org/10.14361/dcs-2016-0207.

Gitelman, Lisa. *Paper Knowledge: Toward a Media History of Documents*. Durham, NC: Duke University Press, 2014.

Gitelman, Lisa, and Virgina Jackson. "Introduction." In *Raw Data is an Oxymoron*, edited by Lisa Gitelman, 1–14. Cambridge, MA: MIT Press, 2013.

Godec, Primož, Matjaž Pančur, Nejc Ilenič, Andrej Čopar, Martin Stražar, Aleš Erjavec, Ajda Pretnar, Janez Demšar, Anže Starič, Marko Toplak, Lan Žagar, Jan Hartman, Hamilton Wang, Riccardo Bellazzi, Uroš Petrovič, Silvia Garagna, Maurizio Zuccotti, Dongsu Park, Gad Shaulsky, and Blaž Zupan. "Democratized Image Analytics by Visual Programming Through Integration of Deep Models and Small-Scale Machine Learning." *Nature Communications* 10, no. 1 (2019): 1–7. https://doi.org/10.1038/s41467-019-12397-x.

Goodwin, John, and Laurie Parsons. "A Life in Motion: Exploring Auto/Biographical Exchanges by 'Walking With' Nelson Sullivan." *Sociological Research Online* 27, no. 3 (2022): 724–744. https://doi.org/10.1177/13607804211025541.

Hayles, N. Katherine. "How We Read: Close, Hyper, Machine." *ADE Bulletin* 150 (2010): 62–79. https://doi.org/10.1632/ade.150.62.

Hesselberth, Pepita. *Cinematic Chronotopes: Here, Me, Now*. Amsterdam: Amsterdam University Press, 2014.

Hiippala, Tuomo. "Distant Viewing and Multimodality Theory: Prospects and Challenges." *Multimodality & Society* 1, no. 2 (2021): 134–152. https://doi.org/10.1177/26349795211007094.

Hilderbrand, Lucas. *Inherent Vice: Bootleg Histories of Videotape and Copyright*. Durham, NC: Duke University Press, 2009.

Hoffmann, Anna Lauren. "Terms of Inclusion: Data, Discourse, Violence." *New Media & Society* 23 (2021), 3539–3556. https://doi.org/10.1177/1461444820958725.

Houwen, Janna. *Film and Video Intermediality: The Question of Media Specificity in Contemporary Moving Images*. London: Bloomsbury, 2017.

Jackson, Ben, and Tim Hitchcock. "Old Bailey Voices." Accessed March 1, 2023, https://oldbaileyvoices.org.

Lebow, Lisa. "Introduction." In *The Cinema of Me: The Self and Subjectivity in First-Person Documentary Film*, edited by Lisa Lebow and Michael Renov, 1–12. New York: Columbia University Press, 2012.

Manovich, Lev. *Cultural Analytics*. Cambridge, MA: MIT Press, 2020.

Montez, Ricardo. "Virtuosic Distortion: Nelson Sullivan's Queer Hand." *ASAP/Journal* 2 (2017): 395–421. https://doi.org/10.1353/asa.2017.0030.

Moran, James. *There's No Place Like Home Video*. Minneapolis: University of Minnesota Press, 2002.

Moretti, Franco. *Distant Reading*. London: Verso, 2013.

Morse, Nicole Erin. *Selfie Aesthetics: Seeing Trans Feminist Futures through Self-Representational Art*. Durham, NC: Duke University Press, 2022.

Motrescu-Mayes, Annamaria, and Susan Aasman. *Amateur Media and Participatory Culture: Film, Video and Digital Media*. London: Routledge, 2019.

Prelinger, Rick. “The Appearance of Archives.” In *The YouTube Reader*, edited by Pelle Snickars and Patrick Vonderau, 268–274. Stockholm: National Library of Sweden, 2009.
Renov, Michael. *The Subject of Documentary*. Minneapolis: University of Minnesota Press, 2004.
Romele, Alberto, Marta Severo, and Paolo Furia. “Digital Hermeneutics: From Interpreting with Machines to Interpretational Machines.” *AI & Society* 35 (2020): 73–86. https://doi.org/10.1007/s00146-018-0856-2.
Ruby, Jay. “Speaking For, Speaking About, Speaking With, or Speaking Alongside: An Anthropological and Documentary Dilemma.” *Visual Anthropology Review* 7 (1991): 50–67. https://doi.org/10.1525/var.1991.7.2.50.
Ruoff, Jeffrey K. “Home Movies of the Avant-Garde: Jonas Mekas and the New York Art World.” *Cinema Journal* 30, no. 3 (1991): 6–28. https://doi.org/10.2307/1224927.
Sayers, Jentery. “Bringing Trouvé to Light: Speculative Computer Vision and Media History.” In *Seeing the Past with Computers: Experiments with Augmented Reality and Computer Vision for History*, edited by Kevin Kee and Timothy Compeau, 32–49. Ann Arbor: University of Michigan Press, 2019.
Schwandt, Silke. “Opening the Black Box of Interpretation: Digital History Practices as Models of Knowledge.” *History and Theory* 61 (2022): 77–85. https://doi.org/10.1111/hith.12281.
Slootweg, Tom. “Resistance, Disruption, Belonging: Electronic Video in Three Amateur Modes.” PhD diss., University of Groningen, 2018.
Stuckenbrock, Anja. *Deixis in der face-to-face-Interaktion*. Berlin: De Gruyter, 2015.
Terrell, Matthew. “About Last Night: Nelson Sullivan’s Camcorder Videos of 1980s Queer New York are a Rare Cultural Treasure.” *Slate*, June 19, 2019. Accessed April 2, 2023, https://slate.com/human-interest/2019/06/nelson-sullivan-videos-queer-new-york.html.
Van der Ven, Inge, and Tom van Nuenen. “Digital Hermeneutics: Scaled Readings of Online Depression Discourses.” *Medical Humanities* 48 (2022): 335–346. https://doi.org/10.1136/medhum-2020-012104.
Wevers, Melvin, and Thomas Smits. “The Visual Digital Turn: Using Neural Networks to Study Historical Images.” *Digital Scholarship in the Humanities* 35, no. 1 (2020): 194–207. https://doi.org/10.1093/llc/fqy085.
Zabalbeascoa, Patrick, and Montse Corrius. “Conversation as a Unit of Film Analysis: Databases of L3 Translation and Audiovisual Samples of Multilingualism.” *MonTI: Monografías de Traducción e Interpretación* 4 (2019): 57–85. https://doi.org/10.6035/MonTI.2019.ne4.2.
Zimmermann, Patricia R. “Speculations on Home Movies: Thirty Axioms for Navigating Historiography and Psychic Vectors.” In *Private Eyes and the Public Gaze: The Manipulation and Valorisation of Amateur Images*, edited by Sonja Kmec and Viviane Thill, 12–13. Trier: Kliomedia, 2009.
Zimmermann, Patricia R. *Reel Families: A Social History of Amateur Film*. Bloomington: Indiana University Press, 1995.

Videography

5ninthavenueproject. “After the 1983 Gay Pride Parade.” Uploaded on YouTube November 26, 2015, 5:31. https://youtu.be/pK03EtGcBVs.
5ninthavenueproject. “Nelson Sullivan’s First Video Blog in 1983.” Uploaded on YouTube December 6, 2011, 7:23. https://youtu.be/8Qc9sISu5yc.

5ninthavenueproject. “Nelson Sullivan and his New Camcorder – Part 1.” Uploaded on YouTube March 23, 2016, 8:06. https://www.youtube.com/watch?v=Gp9e1YkWqAA.
5ninthavenueproject. “Nelson Sullivan Explains His Camcorder Technique to Lahoma.” Uploaded on YouTube February 4, 2009, 2:04. https://www.youtube.com/watch?v=CFCUL9HyJZI.
5ninthavenueproject. “Nelson Sullivan’s 39th Birthday – Waiting for Eric and Liz.” Uploaded on YouTube August 15, 2018, 13:33. https://youtu.be/AjL6goBP0Kg.
5ninthavenueproject. “Stopping at McDonald’s in 1989.” Uploaded on YouTube March 24, 2016, 10:24. https://www.youtube.com/watch?v=EZJAtkqiV8U&t.

Epilogue

Eric Hoyt

Narration, Agency, and the Digital Film Historiography Group Chat

In the first chapter of his classic 1961 work, *What is History?*, E. H. Carr offered his first of several answers to the question posed in the book's title: "[history] is a continuous process of interaction between the historian and his facts, an unending dialogue between the present and the past."[1] To be sure, much has changed over the 65 years since Carr's Oxford lectures were presented in book form and, today, with the open access online distribution of this edited volume. Both the objects of our study (cinema and its contexts) and the gender of the researchers who pursue it (male, female, and nonbinary) extend beyond what is implicitly assumed to be disciplinary boundaries in Carr's chapter titled, "The Historian and His Facts." Yet many of the interpretive and iterative approaches shared in this edited volume echo Carr's dynamic conception of history as "a continuous process of interaction." History is not a static set of facts in the past, waiting to be unearthed, but a process for generating knowledge through putting the evidence of the past into conversation with our present frameworks of inquiry, thought, concern, and language.

Rather than "an unending dialogue," though, the range of critical methods and literacies presented in this edited volume may leave us feeling immersed in an unending group conversation, or, to use a digital metaphor, a group chat or text message chain. No sooner do we sit down at our computers to pursue a research question, utilizing the CLARIAH Media Suite, Distant Viewing Toolkit, or Media History Digital Library,[2] then we see a message reminding us to think critically about the search algorithms at play. This is immediately followed by three gray dots indicating that source criticism, tool criticism, and data visualization are all typing. A cascade of messages soon pours out, providing valuable prompts for reflection about the databases and interfaces that we are utilizing. As the chapters in this volume make clear, computer programming skills, while helpful, are not required for the critical reflections and interpretations that follow. Curiosity, empathy, and patience are ultimately more valuable for the unending conversation than proficiency in Python or SQL. The digital film historiography group chat is alive, offering strategies, insights, and words of caution as we proceed.

1 E. H. Carr, *What is History?* (New York: Random House, 1961).

2 See https://mediasuite.clariah.nl/; https://www.distantviewing.org/; https://mediahistoryproject.org/, accessed July 11, 2024.

https://doi.org/10.1515/9783111082486-019

Yet such a dynamic and multi-faceted conversation requires its own note of caution. There is a risk that we get so overwhelmed in the self-reflective analysis and criticism of sources, tools, and algorithms that we lose sight of the original research questions that animated our interests. It's that feeling of scrolling up on one's phone, struggling to locate the message of the greatest importance among so many other thoughts. This matters because it may leave less space in the conversation for the lives and stories that make history compelling to study. While it is by no means impossible to utilize digital methods and present exciting and compelling histories – and, indeed, this is precisely what I aspire toward in my own work – I have also had to confront the fact that my attention and energy are finite. So too are the word counts and length limitations of editors. In practice, lengthy discussions of computational methods and data structures can come at the expense of concentrating on the lives, films, and contexts that most excite our readers and ourselves. To put the question differently, whose stories are we telling – historical actors from the distant past, the database developers of the more recent past, and/or the journeys of ourselves and our students in the present navigating, in the words of John Lewis Gaddis, "the landscape of history?"[3] The reflexive turn runs the risk of turning our unending conversations too far inward.

If *Doing Digital Film History* were to add a fifth section of the book (sequel, please!), I would propose that this new section concentrate upon historical narration across different modes of expression – including interactive websites, videos, podcasts, and, of course, writing. Narration and storytelling are tremendously important for reaching audiences inside and outside of academia. And, through the process of narrating history, we explicitly or implicitly assign agency – to historical actors and causal forces, the decisions of archivists and software developers, and our own priorities and choices as historians. Elsewhere, I have commented at length about the challenges of writing histories that emerge from computationally intensive and highly collaborative research projects.[4] When there are multiple authors, it can be a struggle to write history in a unified and coherent voice. And, even more challenging to resolve, there is the question of the amount of depth and detail in which digital methods and algorithms should be explained to readers. Too much technical detail halts the momentum of the argument and narrative; too little makes the process opaque and creates the much maligned "black box" that other scholars don't understand and view with suspicion; and almost

3 John Lewis Gaddis, *The Landscape of History: How Historians Map the Past* (New York: Oxford University Press, 2004).

4 Eric Hoyt, "Curating, Coding, Writing: Expanded Forms of Scholarly Production," in *The Arclight Guidebook to Media History and the Digital Humanities*, ed. Charles R. Acland and Eric Hoyt (Falmer: REFRAME Books, 2016), 347–373.

any amount of methodological and technical discussion at all turns attention away from the underlying historical actors. Writing is a craft, and there is no one-size-fits-all solution. When working on a book chapter or essay, I am always wondering how much or how little of myself and my methods to foreground in my writing (which, I am quickly realizing, is a lot in this particular essay!).

Perhaps because I have still not resolved these conundrums, narration and agency have been on my mind while reading and re-reading the excellent essays in this edited volume, from Janes Gaines' reflections on counterfactual histories,[5] to Marcus Burkhardt and Skadi Loist on what data visualizations highlight vs. obfuscate,[6] to Christian Gosvig Olesen's invocation of Hayden White's book, *Metahistory*, and Edward Branigan's article, "Color and Cinema: Problems in the Writing of History" (analytical writings that, notably, use diagrams and visualizations to help clarify their points).[7] Christian Gosvig Olesen cites White and Branigan in the context of calling attention to the CLARIAH Media Suite's lineage with hypermedia and database-driven film scholarship of the 1990s and 2000s (much of which is sadly no longer accessible). Still, and driven by my own self-interests, I couldn't stop wondering about what applying White's and Branigan's emphasis on narration and writing would reveal about film histories published since the launch of the first Media History Digital Library website in 2011.[8] In their published 2018 study, Julia Noordegraaf, Kathleen Lotze, and Jaap Boter found that scholars were far more likely to incorporate the Cinema Contexts database of Dutch film exhibition into traditional research programs and publications, rather than harnessing it for big data and computationally intensive research projects.[9] My sense is that the same patterns hold for the way that most scholars utilize the Media History Digital Library and Lantern[10] – that is, chiefly as a free gateway to locate sources that can be cited and quoted as evidence (and extracted as illustra-

5 See Jane M. Gaines, "The DH Dilemma: Knowing More & Knowing for Sure vs. Never Knowing At All," this volume.

6 See Marcus Burkhardt and Skadi Loist, "Visualization In/As Digital Media Studies," this volume.

7 See Christian Gosvig Olesen, "Timelines of Scholarly Video Annotation: For a Tool Critical History of Digital Film Historical Scholarship," this volume.

8 For more on the development of Lantern and the Media History Digital Library, see Eric Hoyt, "Building a Lantern and Keeping it Burning," in *Applied Media Studies*, ed. Kirsten Ostherr (New York: Routledge, 2017), 238–250.

9 Julia Noordegraaf, Kathleen Lotze, and Jaap Boter, "Writing Cinema Histories with Digital Databases: The Case of Cinema Context," *TMG Journal for Media History* 21, no. 2 (2018): 106–126, https://doi.org/10.18146/2213-7653.2018.369.

10 See https://lantern.mediahist.org/, accessed July 11, 2024.

tions), rather than as a big data repository that can be queried via API.[11] While some of this may be attributed to technical knowledge, we should not underestimate the degree to which computationally intensive and big data methods challenge the norms of Humanities writing, particularly when it comes to book-length projects. In other words, it's not simply that the inputs of computational analysis can be difficult to implement; the outputs can be unwieldly and hard to embed within historical narratives and scholarship.

As I wrestle with these questions, my mind keeps returning to the Women Film Pioneers Project (WFPP).[12] The WFPP features prominently in the book, and for good reason. It is an exemplary model for "doing digital film history," and doing it in a way that attends to narrative, agency, big data opportunities, and thoughtful self-reflection. The WFPP includes engagements and references to primary sources. It allows for vast movements between modes and scales of presenting history, from the biographical information and narratives of individual women film workers, up to more abstract forms of presenting and analyzing information, such as the dendrogram graph of occupation classifications created for the Women Film Pioneers Explorer (WFPE), a digital data visualization project based on the data of the WFPP.[13] While the WFPP tracks the contributions of past film workers, the project also attends to the labor of twenty-first century scholars, curators, and archivists. As Sarah-Mai Dang and Kate Saccone highlight in their book chapters, the WFPP provides attribution to the writers of the project's entries, making visible the past and present labor that populates the website.[14] Moreover, the approaches to historical narration are expressive, heterogenous, and dynamic. Curators bring their own writing styles to the project, yet the narration is never set in stone; Saccone describes the method as one of "'(re)visioning,' which draws on the open-ended processes of creating visibility at the heart of (feminist) film historiography and the practice of versioning at the heart of digital humanities."[15]

Like all of the best digital film history resources, the WFPP utilizes database structures without being defined by their logic. The WFPP presents historical narratives in ways that are open to creative expression, collaboration, and revision,

11 Documentation on how to utilize the Media History Digital Library's API is available at https://lantern.mediahist.org/api, accessed July 11, 2024.

12 See https://wfpp.columbia.edu/, accessed July 11, 2024.

13 See https://www.informatik.uni-marburg.de/women-film-pioneers-explorer/, accessed July 11, 2024.

14 See Sarah-Mai Dang, "Managing the Past: Research Data and Film History," this volume; Kate Saccone, "(Re)Visioning Women's Film History: The Women Film Pioneers Project and Digital Curatorial-Editorial Labor," this volume.

15 Kate Saccone, "(Re)Visioning Women's Film History," this volume.

providing attribution to the individuals responsible in the production of new knowledge. It is a resource that recognizes the agency and contributions of both historic women film workers and contemporary archivists and scholars. And the WFPP explicitly situates itself within a larger network of institutions and resources, ranging from searchable catalogs and collections (e.g., the AFI Catalog and Media History Digital Library), to conferences and festivals that foster community and collaboration (e.g., Women and the Silent Screen Conference, Il Cinema Ritrovato).

Collectively, these interconnected initiatives can transform the film historiography group chat from metaphor into reality. The messages come from our collaborators, colleagues, and friends, and not simply the critical inner voices that bounce around inside our own heads. We leave space for others to speak, allowing for new perspectives and opportunities. Sometimes, the messages come from those currently working in the field who are actively pursuing research and curation. Sometimes, they are the voices of those long gone who have the power to speak to us again.

References

Carr, E. H. *What is History?* New York: Random House, 1961.

Gaddis, John Lewis. *The Landscape of History: How Historians Map the Past*. New York: Oxford University Press, 2004.

Hoyt, Eric. "Building a Lantern and Keeping it Burning." In *Applied Media Studies*, edited by Kirsten Ostherr, 238–250. New York: Routledge, 2017.

Hoyt, Eric. "Curating, Coding, Writing: Expanded Forms of Scholarly Production." In *The Arclight Guidebook to Media History and the Digital Humanities*, edited by Charles R. Acland and Eric Hoyt, 347–373. Falmer: REFRAME Books, 2016.

Noordegraaf, Julia, Kathleen Lotze, and Jaap Boter. "Writing Cinema Histories with Digital Databases: The Case of Cinema Context." *TMG Journal for Media History* 21, no. 2 (2018): 106–126. https://doi.org/10.18146/2213-7653.2018.369.

List of Authors

Susan Aasman is Professor in Digital Humanities at the Centre for Media and Journalism Studies, University of Groningen, Netherlands, and Director of the Centre for Digital Humanities. Her field of expertise is in media history, especially in amateur media, documentaries, web history, web archaeology, and (born-)digital archives. She co-edited *Materializing Memories: Dispositifs, Generations, Amateurs* (Bloomsbury, 2018) and is the co-author of *Amateur Media and Participatory Culture: Film, Video and Digital Media* (Routledge, 2019). ORCID: 0000-0003-1675-2998

Taylor Arnold is Professor of Data Science & Statistics at the University of Richmond, United States, in the interdisciplinary programs in Data Science and Linguistics. A recipient of grants from the NEH and ACLS, Arnold's research focuses on the interdisciplinary study of how messages are communicated through texts and visual media. He has published four books. With co-author Lauren Tilton, he runs the Distant Viewing lab at the University of Richmond.
ORCID: 0000-0003-0576-0669

Jonatan Bruun Borring is a freelance filmmaker and graduate student at the University of Copenhagen, Denmark. He is interested in the intersection of emerging technologies and film analysis, hoping to offer innovative perspectives on the evolution of cinema.

Nicole Braida completed her international Master's degree in Film and Audiovisual Studies (IMACS) at the University of Udine, Italy, with semesters abroad in Frankfurt and Liège. In 2021 she received her doctorate as part of the graduate school "Configurations of Film" at the Johannes Gutenberg University Mainz, and in 2022 her dissertation was published by Transcript with the title *Migrating through the Web: Interactive Practices about Migration, Flight and Exile*. Since April 2021 she has been a research assistant and postdoc for the DiCi-Hub (A Digital Hub for Film Studies) project. She teaches Digital Methods for Film Historiography and Data Visualization in Germany and Italy at the University of Udine. ORCID: 0000-0002-9446-6762

Marcus Burkhardt is Professor for Media, Algorithms and Society at Paderborn University, Germany. He has a background in media studies, philosophy and computer science and received his doctorate at Justus-Liebig-University Gießen with a thesis on the media theory of digital databases. His research interests include the history and theory of digital media, especially the logi(sti)cs of database technologies, big data, and algorithmic media, as well as digital methods. Marcus is Co-PI of the research projects "Agentic Media: Formations of Semi-Autonomy" (DFG 2020–2027, with Karin Knorr Cetina) and "Industry of Personal Data" (DFG 2024–2027, with Maximilian Becker) in the Collaborative Research Center 1187 "Media of Cooperation." ORCID: 0000-0003-0382-3128

Sarah-Mai Dang is Principal Investigator of the BMBF research group "Aesthetics of Access: Visualizing Research Data on Women in Film History" (DAVIF) (2021–2025) at the Institute of Media Studies at Philipps-Universität Marburg, Germany. Drawing on her current research and teaching focus on digital film historiography, data visualization, feminist theory, open science, and media aesthetics, she also founded the international DFG research network "New Directions in Film Historiography: Digital Tools and Methods in Film and Media Studies" (2019–2023). In addition, she initiated the collaborative *Open Media Studies Blog* and the hybrid publishing project oa books, where she published her dissertation and blogged about academic publishing. She received a doctoral

https://doi.org/10.1515/9783111082486-020

degree in Film Studies from Freie Universität Berlin and has published widely on the impact of datafication in film and media studies. ORCID: 0000-0002-1960-247X

Josephine Diecke is Assistant Professor of Film Studies at the University of Zurich, Switzerland. She has conducted extensive research on the history of film color technologies, moving image preservation, and digital methods for computer-assisted text and video analysis. Her expertise stems from her work as a research associate on the "Filmcolors" project (University of Zurich), as the academic coordinator of the "Digital Cinema-Hub" project (Philipps-Universität Marburg), and as a film lab technician for various service providers. She holds a PhD from the University of Zurich with a thesis on the color film processes Agfacolor and Orwocolor, and is co-editor of the *Open Media Studies Blog*. ORCID: 0000-0002-9342-0631

Paul Duchesne is a developer and archivist currently based at the National Film and Sound Archive of Australia, where he manages their collection systems and operations. He has previously worked extensively in film preservation and spent time at ACMI and the Technische Informationsbibliothek Open Science Lab in Hannover, specializing in Wikidata and Linked Open Data for cultural data. He is an active member of the FIAF Cataloguing and Documentation Commission, including authoring the FIAFcore ontology and contributing substantially to the forthcoming FIAF Knowledge Graph.

Jane M. Gaines is Professor of Film, Columbia University, and Professor Emerita of Literature and English, Duke University, in the United States. In 2018 she received the Society for Cinema and Media Studies Distinguished Career Award, and in 2022 an Honorary Doctorate from the University of Stockholm. She is the author of three award-winning books: *Contested Culture: The Image, the Voice and the Law* (North Carolina, 1991), *Fire and Desire: Mixed Race Movies in the Silent Era* (Chicago, 2001), and *Pink-Slipped: What Happened to Women in the Silent Film Industries?* (Illinois, 2018). She also works on philosophy of history, intellectual property and documentary radicalism.

Matthias Grotkopp is Assistant Professor for Digital Film Studies at "Cinepoetics – Center for Advanced Film Studies" and the Seminar for Film Studies at Freie Universität Berlin, Germany. He is the author of *Cinematic Poetics of Guilt: Audiovisual Accusation as a Mode of Commonality* (Berlin/ Boston: De Gruyter, 2021). He supervises the project "Intervening World Projections: Audiovisuality of Climate Change," which is part of the CRC 1512 "Intervening Arts" funded by the German Research Council. His research interests include the audiovisuality of climate change and ecological disaster, genre theory, and the relation of politics and poetics, as well as digital methods of film analysis. He is the managing editor of the open access online journal medi*a*esthetics based at Cinepoetics. ORCID: 0000-0002-0954-3111

Malte Hagener is Professor of Media and Film Studies at Philipps-Universität Marburg, Germany. He studied in Hamburg, Norwich, and Amsterdam where he received his PhD with a dissertation on the European film avant-garde (*Moving Forward, Looking Back: The European Avant-garde and the Invention of Film Culture, 1919–1939*, 2007). He is director of the DFG-funded project "media/rep/ – Open Access-repository for media studies" (https://mediarep.org/); principal investigator of the DFG-funded graduate research group "Configurations of film" (https://konfigurationen-des-films.de/en/); co-speaker of the consortium "NFDI4Culture. Consortium for Research Data on Material and Immaterial Cultural Heritage" (2020–2025); and received the Swedish-German Research Award 2016 by Riksbanken Jubileumsfond and Alexander-von-Humboldt-Stiftung (2016/2017 in Karlstad). He was co-author (with Thomas Elsaesser) of *Film Theory: An Introduction through the Senses* (2015, many

translations); editor of *The Emergence of Film Culture: Knowledge Production, Institution Building and the Fate of the Avant-garde in Europe, 1919–1945* (2014); and co-editor of *Handbuch Filmanalyse* (2020; with Volker Pantenburg) and *How Film Histories Were Made: Materials, Methods, Discourses* (2024; with Yvonne Zimmermann). ORCID: 0000-0003-4664-5536

Adelheid Heftberger is deputy head of the Film Department in the Bundesarchiv, Berlin, Germany. In the past, she worked as a researcher, curator, and archivist at the Brandenburg Center for Media Studies in Potsdam and at the Austrian Filmmuseum in Vienna. She holds a PhD in Slavic Studies and obtained a Master's degree in Comparative Literature from the Universities of Innsbruck and Vienna. In 2016, she completed her MA in Library and Information Sciences at Humboldt-Universität zu Berlin. She is head of the FIAF Cataloguing and Documentation Commission and actively supports the Open Science movement. ORCID: 0000-0001-7909-6858

Franziska Heller is Professor for Media Studies at Martin Luther University Halle-Wittenberg, Germany. After obtaining a PhD from the Ruhr University Bochum in 2009 with a thesis on phenomenology and filmic narration focusing on film aesthetics and fluidity, she spent a number of years in research and teaching at the University of Zurich (Habilitation in 2018). She has worked on various research projects investigating the digitization of archival films. From 2019–2021, she held a position as FONTE Visiting Professor at the Filmuniversität Babelsberg KONRAD WOLF in Potsdam and as Senior Fellow of the Center for Advanced Film Studies *Cinepoetics* at the FU Berlin. In addition to having published numerous articles on moving images and digital cultures, she is also the author of three monographs, the most recent dating from 2020: *Update! Film and Media History in the Age of Digital Reproducibility* (Brill/Fink), available in print and open access. ORCID: 0000-0002-4443-1238

Kerstin Herlt is a project manager at the digital department of DFF – Deutsches Filminstitut & Filmmuseum in Frankfurt am Main. Working there since 2001, she coordinated "f_films," a database and filmography of female film pioneers in Germany and Europe. From 2008 to 2017, she was head of office and project manager at the Association des Cinémathèques Européennes. She has worked on projects that digitize, preserve, curate, and enrich Europe's film heritage, such as EFG – European Film Gateway, the WWI digitization project EFG1914, the orphan works film project FORWARD, and the Visual History of the Holocaust (VHH). She rejoined the DFF in 2018 and is currently the coordinator of the European project "DE-BIAS," which takes an AI-driven and integrative approach to identify contentious language in the collection descriptions of cultural heritage institutions. She holds a Master's degree in Romance Studies, Sociology, and European Media Studies.

Eric Hoyt is the Kahl Family Professor of Media Production in the Department of Communication Arts at the University of Wisconsin-Madison. He is the Director of the Wisconsin Center for Film and Theater Research and Media History Digital Library, which has digitized over three million pages of historic books and magazines for open access. His books include *Hollywood Vault: Film Libraries before Home Video* (2014) and *Ink-Stained Hollywood: The Triumph of American Cinema's Trade Press* (2022), as well as the co-edited anthologies *The Arclight Guidebook to Media History and the Digital Humanities* (2016), *Saving New Sounds: Podcast Preservation and Historiography* (2021), and *Global Movie Magazine Networks* (2025). ORCID: 0000-0001-7753-5343

Imme Klages is a film historian with expertise in the field of German film exiles. Her PhD about the films of Fred Zinnemann was published in 2018 with Schüren (Marburg) under the title *I do not get rid of the ghosts. Zur Exilerfahrung in den Filmen Fred Zinnemanns: THE SEARCH (1948), THE NUN'S STORY*

(1959) und JULIA (1977). She is currently a postdoctoral researcher in the DFG-funded project "Mapping German Film Migration 1930–1950: Eine datengraphbasierte Perspektive auf die Filmemigration aus NS-Deutschland" (2021–2024), approaching film-historical research with digital methodology. She is a member of the network "Cinéma et 'rémigration' vers l'Allemagne et l'Europe après 1945" supported by CIERA 2023–2025, University of Lyon, Film University Babelsberg KONRAD WOLF, and Johannes Gutenberg University Mainz, and was a member of the research-network "German-Jewish Film History of the Federal Republic" (DFG, 2019–2023). ORCID: 0000-0003-0830-9283

Fabian Kling is a PhD candidate and research associate at the Johannes Gutenberg University Mainz, Germany. He studied Media Design (BA) until 2016 and worked in advertising after graduating. In 2017 he began studying Film Studies (BA) and Media Dramaturgy (MA) at the Johannes Gutenberg University Mainz, completing both in 2022, the latter with the thesis "Digital Cartography as a Method of Film Historiography – Conception and Design for a Biographical Source Index based on Digital Mapping." His work focuses on digital methods and dramaturgies, data visualization, live broadcasting, and media sports. His dissertation focuses on ZDF/Das kleine Fernsehspiel and the interface between film and television history. ORCID: 0009-0009-5084-8572

Skadi Loist is Assistant Professor (*Juniorprofessor*) for Production Cultures in Audiovisual Media Industries at the Film University Babelsberg KONRAD WOLF in Potsdam, Germany. Their research focuses on film festivals and circulation, queer film culture, and screen industries, diversity and sustainability. Skadi was Principle Investigator (PI) in projects that used data visualization as part of the research process, such as "Film Circulation on the International Film Festival Network and the Impact on Global Film Culture" (BMBF 2017–2022) and "GEP Analysis: Assessing, Understanding, and Modeling the Impact of Gender Equity Policies (GEP) in the Film Industry" (DFG/ESRC/SSHRC 2021–2024). The current project within the QUADRIGA Datenkompetenzzentrum is dedicated to making digital methods accessible to early career researchers. ORCID: 0000-0002-8299-4103

Julia Noordegraaf is Professor of Digital Heritage in the Faculty of Humanities at the University of Amsterdam, Netherlands, where she founded the research program and lab Creative Amsterdam (CREATE), which studies the history of urban creativity using digital data and methods. Her research focuses on the reuse of digital cultural heritage for media historical research. In her current book project, *Digital Truths: Practicing Media Historiography in the Digital Age*, she investigates the methodological and epistemological implications of digital media historiography. She is Editor-in-Chief of the Cinema Context database on Dutch cinema culture, a member of the Royal Netherlands Academy of Arts and Sciences, and has been involved as a board member in two Dutch national digital humanities infrastructure projects (CLARIAH, SSHOC-NL). In the context of the European Time Machine project, Noordegraaf coordinates the realization of the Amsterdam Time Machine, a spatial and temporal platform for location-based, personalised access to historical information on the city. ORCID: 0000-0003-0146-642X

Christian Gosvig Olesen is Assistant Professor of Digital Media and Cultural Heritage at the University of Amsterdam, where he teaches courses in film and media studies and in moving image archiving in the MA "Preservation and Presentation of the Moving Image." His research interests include moving image archiving theory and history, film and media historiography, digital methods, found footage, and remixing. For the Dutch CLARIAH Media Suite and SSHOC-NL infrastructure projects, he currently coordinates and develops the Learn teaching and training initiatives. ORCID: 0000-0002-4860-1951

Frauke Pirk has been research assistant at Johannes Gutenberg University Mainz, Germany, since July 2021, with a focus on Digital Humanities in the project "DiCi-Hub: A Research Hub for Digital Film Studies" (funded by the VW Foundation). She obtained her Bachelor of Arts in German Studies at the Technical University Chemnitz in 2018. In 2021 she completed her Master's degree in "Digitale Methodik in den Geistes- und Kulturwissenschaften" (Digital Methodology in the Humanities and Cultural Studies) at the Johannes Gutenberg University and the University of Applied Sciences Mainz. From October 2021 to September 2023, she was also research assistant at the Studium Generale at the University of Mainz. ORCID: 0000-0001-8640-0213

Ulrich Ruedel holds a doctorate in Analytical Chemistry from the University of Muenster, Germany, and worked on optical biochemical sensors and intellectual property rights before turning to the practice and science of film preservation. A 2005 graduate of the L. Jeffrey Selznick School of Film Preservation, he explored heritage color systems, such as Technicolor, as archivist at the George Eastman House. Subsequently, he has worked as Research and Development Manager at Haghefilm Conservation in the Netherlands and, concurrently, as Project Manager for the non-profit Haghefilm Foundation. Following employment as Conservation Technology Manager at the British Film Institute's J. Paul Getty Jr Conservation Centre, he has accepted a position as Professor of Conservation and Restoration at HTW – University of Applied Sciences, Berlin, where he focuses on teaching and researching the preservation of audiovisual media.

Kate Saccone is a PhD researcher at the University of Amsterdam, the Netherlands, where her practice-led project engages with silent cinema, moving-image exhibition practice and theory, and critical-feminist approaches to curatorial labor. She is also the project manager and an editor of Columbia University's Women Film Pioneers Project (https://wfpp.columbia.edu/) and currently serves on the Women and Film History International steering committee.
ORCID: 0009-0002-0007-6315

Alexandra Schneider is a Professor of Film and Media Studies at the Johannes Gutenberg University Mainz, Germany. Her fields of expertise include amateur media studies, media archaeology, and decentered film histories. She is the co-founder of "necs – European Network for Cinema and Media Studies" and the co-curator of the Data Paper Section of *Necsus*. She is PI in the Research Training Group "Configurations of Film" and in "Digital Cinema Hub (DiCi-Hub): A Research Hub for Digital Film Studies." In 2020 she co-edited *Format Matters: Standards, Practices and Politics in Media Cultures* (Lüneburg: Meson Press). ORCID: 0000-0002-8914-3790

Tom Slootweg is Assistant Professor of Contemporary History at the University of Groningen, Netherlands. He wrote his dissertation on the (often-utopian) hopes and expectations, but also the everyday amateur uses, of electronic video in the Netherlands and beyond, between the 1960s and 1990s. He is currently exploring various digital tools for doing media historical research and is the parting editor-in-chief of *TMG Journal for Media History*. ORCID: 0000-0002-5181-8094

Lauren Tilton is the E. Claiborne Robins Professor of Liberal Arts and Digital Humanities in the Department of Rhetoric and Communication Studies at the University of Richmond, United States. Her research combines American studies and digital humanities in the study of modern visual culture. She is the author of four books, including *Layered Lives*. With co-author Taylor Arnold, she runs the Distant Viewing lab at the University of Richmond. ORCID: 0000-0003-4629-8888

Casper Tybjerg is Associate Professor of Film Studies at the Department of Communication, University of Copenhagen, Denmark. He is a scholar of Danish and international silent film. He has recently completed a book on the theory and method of film historiography, focusing on the work of Carl Theodor Dreyer, for which he has been awarded the Dr. Phil. (Habilitation) degree. He has also assisted in the DFI's efforts to restore Danish silent films and has co-organized several retrospectives, including at the Pordenone Silent Film Festival. ORCID: 0000-0001-6741-9090

Tim van der Heijden is an Assistant Professor of Media Studies at the Open University of the Netherlands. He holds a PhD in Media History from Maastricht University. His dissertation "Hybrid Histories: Technologies of Memory and the Cultural Dynamics of Home Movies, 1895–2005" investigates the relationship between changing media technologies (film, video, digital media) and home movie practices from a long-term historical perspective. His book *Doing Experimental Media Archaeology: Practice*, co-authored with Aleksander Kolkowski, is part of a dual-set publication that explores the theory and practice of experimental media archaeology as a practice-based and sensorial approach to media historiography. ORCID: 0000-0002-3053-5140

Luan Nhu Vu is a graduate student at the University of Copenhagen, Denmark, with a strong interest in the field of film and media studies, particularly in the realm of cinema. He is dedicated to exploring the intricacies of cinema and making meaningful contributions to this academic discipline.

Julia Welter has been working with the DFF – Deutsches Filminstitut & Filmmuseum as project manager for the EFG – European Film Gateway since 2009. She was also project manager for "EFG1914 – Digitising Film from and about the First World War" (2012–2014) and has been involved as WP lead in several other European projects dealing with the publication of archival collections. She is currently co-leading the EU project "DE-BIAS – Detecting and cur(at)ing harmful language in cultural heritage collections." For the Goethe University Frankfurt, she oversaw the technical development of the online exhibition "Frames of Reconstruction" developed in the EU-funded international research project "ViCTOR-E – Visual Culture of Trauma, Obliteration and Reconstruction in Post-WWII Europe" (2019–2022) and handled all archival contacts. She studied at the University of Marburg, Germany, and the University of Portsmouth, England, and holds a Master's degree in German and English Literature and Film Studies.

Yvonne Zimmermann is a Professor of Media Studies at Philipps University Marburg, Germany. Her recent books include the co-authored *Advertising and the Transformation of Screen Cultures* (AUP, 2021) and the co-edited volumes *How Film Histories Were Made: Materials, Methods, Discourses* (AUP, 2024) and *Films That Work Harder: The Circulation of Industrial Film* (AUP, 2024). Zimmermann is also the editor of a special issue on Asta Nielsen and the introduction of the film star system before the First World War in *Early Popular Visual Culture* (2021) and is co-responsible for the editorial content of the Importing Asta Nielsen Database (IANDb, https://importing-asta-nielsen.online.uni-marburg.de). She is a PI in "Digital Cinema Hub (DiCi-Hub): A Research Hub for Digital Film Studies," a project funded by Volkswagen Stiftung that tests and develops digital tools and methods to be implemented in film and media studies. ORCID: 0009-0006-3571-6414

Index

https://doi.org/10.1515/9783111082486-021