

Bridging the gap: Capturing UK trans health discourse in the Archive of Tomorrow

Alice Austin

Abstract: The barriers trans and non-binary people in the UK face when accessing healthcare have been well documented in recent years, and a proliferation of sites produced by and for trans communities have emerged to bridge the gaps left by suspended services and growing waiting times. Concurrently, a number of high-profile legislative cases and public debates have underscored the extent to which the provision of information about trans* health is defined and shaped by societal and political contexts. This chapter discusses the challenges of collecting online trans* health information in a rapidly changing and hotly contested environment, and explores the questions around representation and the ethical implications of collecting online health discourse.

Keywords: ethical collection, representativeness of collections, health information, contentious collecting.

In traditional conceptions ‘the archive’ functions from a position of neutrality, operating as a storehouse for the passive accumulation of information about the past that is maintained for the benefit and use of the future. More recently, however, it has been acknowledged that rather than reflecting our present reality, archival preservation recreates and reaffirms it, or, as Eric Ketelaar (2001) has argued, “the archive reflects realities as perceived by the ‘archivers’” (Ketelaar 2001, 133).

This position has been hugely informed by the ‘memory boom’ that characterized the late twentieth century, and the attendant rise in community archiving initiatives (Miztal 2010). As the archival profession has begun to attend more closely to the social, mnemonic, and affective aspects of archives, the concept of *representation* has become central to our understanding of the function that archives and archival collections play in society. Examining the impact of archival representation on communities who have traditionally been “ignored, misrepresented, or marginalized” by mainstream repositories, Michelle Caswell et al. (2016) have argued that feeling represented in an archives “has an ontological impact”:

...it changes [the viewer’s] sense of being in the world; she can ‘discover’ herself ‘existing’ in ways she did not before this record was created and made accessible. Representation in community archives catalyzes this ontological shift from not being/not existing/not being documented to being/existing/being documented, with profound personal implications.

(Caswell, Cifor, and Ramirez 2016, 61)

Being cognizant of these implications, then, and aware that if the archive attempts “to collect everything ... it will soon succumb to entropy and

Alice Austin, University of Edinburgh, Scotland, United Kingdom, alice.austin@ed.ac.uk, 0009-0007-5586-2571

Referee List (DOI 10.36253/fup_referee_list)

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

Alice Austin, *Bridging the gap: Capturing UK trans health discourse in the Archive of Tomorrow*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0413-2.07, in Sophie Gebeil, Jean-Christophe Peyssard (edited by), *Exploring the Archived Web during a Highly Transformative Age. Proceedings of the 5th international RESAW conference, Marseille, June 2024*, pp. 45-55, 2024, published by Firenze University Press, ISBN 979-12-215-0413-2, DOI 10.36253/979-12-215-0413-2

chaos” (Spieker 2017, xiii), archivists and curators have had to grapple with the question of which communities, which stories, which *realities*, to represent in their collecting. The difficulty of this is only magnified in the digital sphere, where digital records exist and operate within multiple networked realities simultaneously, and the short lifespan of web content means that archivists do not have the luxury of waiting for the flotsam and jetsam of documentary detritus to wash up upon archival shores.

These challenges were writ large as institutions in the heritage sector rushed to capture a picture of the Covid-19 pandemic from a mercurial and rapidly evolving digital landscape. The impacts of the pandemic were not evenly felt, and research has suggested that ethnic minorities and other already marginalized communities were more heavily impacted by the outbreak (Platt 2021): how can curators reflect and represent such a myriad of different experiences? Deep divisions emerged around what constituted appropriate medical, legal, and social responses to the pandemic: how can collecting respectfully and responsibly reflect the dissent and divisions in a moment without a single, unifying narrative? Key government and medical websites were updated on a daily basis as new information emerged, and social media reacted quickly to find, debate, and digest each new study or guideline: when moving at speed, what are the ethical implications of such ‘rapid response’ collecting?

Emerging from these observations, the Archive of Tomorrow project sought to explore these questions in more detail. After introducing the project in brief, this chapter will then focus on the trans*¹ health subcollection as an exemplary microcosm of the collection as a whole. It will detail how the subcollection evolved; shed insight on the ethical considerations that contributed to its development; and conclude by exploring what the experiences of this project can tell us about creating ‘representative’ collections.

1. Project background

The Archive of Tomorrow (AoT) project sought to build a collection of archived websites to reflect how online spaces were used to share, discuss, and debate issues around health in the aftermath of the pandemic. A Wellcome Trust-funded initiative that was led by the National Library of Scotland with extensive support from the British Library, AoT sought to explore best practices in preserving, describing, and enabling access to information captured from the web. The project team comprised three web

¹ Trans* is an umbrella term referring to a number of identities within the gender identity spectrum. The use of an asterisk expands the definition beyond binary trans identities (i.e. transmen/transwomen) to include non-binary and gender-fluid identities.

archivists (one based with each academic partner), a metadata analyst and a rights officer, and was led by a project manager (all of whom were based at the National Library of Scotland). As well as the tangible objective of preserving a collection of 10,000 targets relating to health, the project also had exploratory aims around how to ethically collect from the web; how to republish responsibly; and what is needed to increase research usage of web archives.

Collection was performed within the context of the UK Web Archive (UKWA), a partnership of the six Legal Deposit Libraries (LDLs) that performs the web function of the LDL's legislative responsibility to collect and preserve a copy of all material published in the UK and Ireland. The UKWA has been systematically collecting non-print material since 2013, with the majority of material being captured through an annual domain crawl that attempts to make a copy of any content published to a website with a recognizable UK top-level domain (e.g., .uk, .scot), or hosted on a server physically located in the UK (identified via a GeoIP lookup). The yearly crawl is supplemented by curated collecting which is achieved by manually adding targets to the Annotation and Curation Tool (W3ACT), a web-based interface that allows a user to create an entry for a specified URL, establish parameters such as depth or frequency of a crawl, and record metadata for description and rights-management purposes.

Curated collections are made available via the UKWA's public interface, where there are over 100 thematic collections available to browse. The collection resulting from the AoT project (since named Talking About Health) comprises around 3,500 targets, and has been further subdivided along various lines (such as source, form, focus, etc) to allow for navigation and discovery. The Regulations that govern legal deposit impose some constraints on collection and access: they only allow for the collection of material that has been made publicly available and do not cover material made available to a 'restricted group' (i.e. requiring an individual to provide credentials to access), nor do they cover material that is predominantly audio-visual in format. Additionally, access to archived material is restricted by default to users at computer terminals onsite in LDLs, unless permission for access has been explicitly granted by the website owner.

2. Trans* health in the UK

The subject of trans* health was selected for focused collecting as it exemplifies both how the digital sphere has transformed contemporary approaches to health information, and how the collecting of such information is complicated by the social and legal contexts in which it exists. Recent years have seen a sustained increase in the media coverage and commentary on the provision of gender-related care and treatment in

the UK. The treatment of gender dysphoria in children has become a particular area of debate, with a number of high-profile legal cases and inquiries being conducted into questions such as the competency of minors to consent to medical care and the long-term impacts of medical treatments such as puberty blockers. The barriers that trans* and non-binary people face when accessing healthcare in the UK were already in dire straits in 2018. An uneven geographical distribution of gender services both across and within the four nations results in many people being required to travel long distances to access healthcare services, and one study released in that year concluded that long wait times “exacerbate gender dysphoria and mental health problems, and increase risks of suicide and self-harm” (TransActual 2022). This was made significantly worse by reduced access to medication and transition-related care as a result of the Covid-19 pandemic, with the average waiting time for a first appointment at an adult gender dysphoria clinic rising to around 38 months in recent years.

Furthermore, there is an increasingly toxic culture of debate surrounding the issue of trans* health and rights, and indeed, on the question of whether trans* identities are or should be considered valid: a recent court ruling concluded that ‘gender critical’ beliefs—broadly put, that a trans* person’s internal feelings about their gender identity has no basis in material reality—constitute a philosophical belief that is protected under the Equality Act and the European Convention on Human Rights (Forstater v CGD Europe 2021). In a 2021 report the Council of Europe observed a “baseless and concerning” level of transphobia, and noted that “rhetoric ... which denies trans identities ... is being used to roll back the rights of trans and non-binary people and is contributing to growing human rights problems” in the UK (Council of Europe 2021).

3. Producing the trans* health subcollection

It was against this backdrop that collecting took place. The trans* health subcollection is comprised of 76 URLs and includes information published by providers both within and external to the publicly funded healthcare systems; gray literature and guidance on the delivery of healthcare; campaign sites relating to the provision of trans* healthcare; peer-to-peer information sharing sites; and social media discussion. A ‘top-down, center-out’ approach to identifying material was employed, with initial efforts focused on material published by service providers operating at the national or top level (NHS, private providers), followed by the regional or local instances of those services. Next, material which addressed the delivery of those services was targeted: this included best practice guidance for individual providers ‘on the ground’, as well as monitoring and advocacy regarding service provision at a national and international level. From these

targets, key areas of discussion and debate emerged which could then be used as access points for the identification of peripheral or ‘bottom-up’ discourse on social media.

During collecting, a decision had to be made regarding the extent to which ‘transceptical’ or gender-critical sites would be included in the collection, and if/how these would be described to users. The project had initially adopted a framing of ‘information vs. misinformation’, but as collecting progressed and the complexities of the documentary landscape emerged it became clear that such a binary distinction was unhelpful: not only were the project team unqualified to make judgments about the veracity, reliability, or appropriateness of a source’s content, it was also felt that attempting to distinguish between information and misinformation in this way would lead to a misrepresentation of the context in which health information is located, accessed, and understood. Instead of trying to determine information from misinformation, then, the project team instead sought to collect all relevant material that could be found on a subject in order to better reflect the documentary landscape at the time of collection.

A vibrant culture of ‘information activism’ has emerged around the subject of trans* health that the project team felt it was important to capture. Sites offer commentary, provide peer-to-peer support for trans* people, and generally seek to bridge the gaps in trans* healthcare provision by collating information on specific medicines (such as guidance on safely acquiring and self-administering hormones in the absence of a prescription) or by sharing first-hand accounts and experiences of treatments, procedures, and providers. The sharing of information about trans* healthcare therefore serves to counter perceived social and systemic barriers to medical treatment and support. One such site, Trans Healthcare Intelligence, sums up their mission thusly:

Accurate and useful information about trans healthcare in the UK is difficult to come by, limited by a transphobic medical system as well as targeted harassment from hate campaigners... This resource aims to collect information about transgender healthcare and our community experiences of a system not designed to cope with our existence, ensuring it's as accessible as possible to our community.

(Trans Healthcare Intelligence)

While there is a reasonable expectation that information published by official or authorized sources such as the NHS will be preserved through other channels, many of these peer-to-peer initiatives exist only in their web-based form with no supporting infrastructure and no regulatory record-keeping duties: they exist only as long as interested individuals have the means and motivation to maintain a website. It is in the capacity to capture and preserve these traces of a community and a documentary landscape that has routinely been excluded from the historical record that the value of web

archiving—as a route towards a more representative and diverse documentary record—can be most clearly observed.

4. Ethical considerations

In addition to the grassroots peer-to-peer sites that were targeted, the subcollection also includes captures of many UK-specific threads, accounts, and forums on social media platforms such as Twitter/X, Reddit, and Tumblr. When approaching this material the project partners had to carefully consider the need to balance the research value of capturing social media discourse against the risk of bringing harm or distress to individuals. There is a growing body of literature exploring the ethical challenges of collecting and using social media posts for academic research that the project team was able to draw on when conceptualizing these challenges and how to address them, and the issue of implied vs. informed consent required particular attention. As Hunter et al. (2018) have noted, although “consent for usage and collection of data are usually implied via [a] platform’s terms of service” the extent to which this can be considered ‘informed consent’ is questionable, and social media users “may not necessarily expect their personal data to be used for research purposes” (Hunter et al. 2018, 345) Furthermore, there was a concern that collecting these social media sources may undermine the social logic by which such platforms and spaces operate. As Nicholas Norman Adams (2022) observes, “many Reddit forums position themselves as ‘safe spaces’ where users can discuss various struggles. Users posting on these forums do so in the knowledge that postings are contextualized within a wider, local topic board conversation: i.e. the ‘safe space’, which is policed by local online moderators” (Adams 2022, 52). To remove this content from this safe space, then, significantly changes the context in which any implied consent is given. This becomes an even more pressing concern when considering the potentially sensitive nature of the topics under discussion here. The project team recognized that the long-term archival implications of posting online may not be at the forefront of an individual’s mind when turning to the internet for information on health-related topics, and particularly considering the possibility that posts were made at a time of crisis or distress.

Similarly, the nature of the topics under discussion within this subcollection required the project team to be mindful of the potential risk that capture and preservation might pose to creators. In their efforts to develop a framework and toolkit for the ethical collection and use of social media content, the *Documenting the Now* project team recognized that “while the benefits of social media to the democratization of information access are clear, the abundance of and access to social media content and

data by countless third parties also presents opportunities for some to ‘weaponize’ the platforms and the data they generate in ways that can cause harm to marginalized and already vulnerable communities” (Jules, Summers, and Mitchell 2018, 3). In October 2022 draft guidance was published indicating that young people in England who access medication or treatment for gender dysphoria without the support of NHS clinicians may be referred to safeguarding agencies, including the police (Topping 2022). This presents questions as to the extent to which the information being shared on these personal blogs, message boards, and other sites could conceivably be construed as promoting the use of controlled substances or—at the extreme—encouraging child endangerment. Adams’ exploration of the ethical challenges of using social media content in scholarly research noted that “replication of Reddit user postings—verbatim—in scholarly publications can often lead to internet reverse-searching. In some cases, this could allow the original Reddit threads to be easily and rapidly located online, therefore risking invalidation of any assumed ‘participant’ anonymity and allowing the linking of specific isolated comments used in publications to specific user accounts and postings” (2022, 7). It is not inconceivable that a user may remove content they have shared in an attempt to protect themselves from potential legal action, but then “discover that their comments now exist in a permanent archive, for which they have no control over the ways in which such comments are used; no autonomy and decision over the deletion of these materials, nor access to any procedure from which to de-associate these comments with their Reddit username” (Adams 2022, 9). While there is a clear argument for the historic and social value of preserving such material, then, it is also important for curators and collecting initiatives to be aware of these issues and to consider what responsibilities the archive has in relation to content creators.

As noted above, a minimal amount of description has been applied to sites within the Talking About Health collection. Sites were assigned to the main collection using W3ACT’s tagging function, and from there, could be further assigned to one or more sub-categories that had been chosen to aid navigation and discoverability. These low-level descriptors were largely intended to describe the publisher rather than the content—denoting a target as being NHS-published, or a social media resource, or a charity website, for example—and the tagging function was also used to group sites along a theme, producing subcategories like the trans* health subcollection.

However, even minimal levels of description and arrangement can influence how a resource is understood by a future researcher. In relation to this subcollection, the risks of descriptive choices exoticizing and ‘othering’ an already marginalized community were apparent. Historian Jules Gill-Peterson (2022) has argued that the “material difference between transgender healthcare and non-transgender healthcare...is transphobia”.

The medical resources needed to transition are not of a different species than the equally numerous ways that non-trans people's sex and gender are routinely medicalized. Yet they are treated fundamentally differently. Although they share the same clinical and scientific history, one is treated as new, experimental, and potentially dangerous, while the other is rarely the subject of sustained news coverage at all. One is treated as always arriving too quickly while the other is treated as so unremarkable it is as if it has always existed.

(Gill-Peterson 2022)

This siloing of trans* health concerns has very real consequences, with Wall et al. exploring how 'trans broken arm syndrome' (a form of medical discrimination faced by transgender and gender diverse patients wherein healthcare providers "conceptualize patients through their transgender identity first, and chief complaint second") can adversely affect the level of care that trans* people receive (2023, 18). Recognition of this required the project team to consider whether by isolating 'trans* health' from 'cis health' our collecting practices might be compounding the othering and exclusion of a marginalized group of people and how our descriptive practices can better reflect the ways that these communities view, understand, and describe themselves.

5. Conclusions

In many respects, the trans* health subcollection can be understood as a microcosm of the Talking About Health collection as a whole. As the UK's legal gender recognition processes require a clinical diagnosis of gender dysphoria, this topic is particularly illustrative of how questions of health are entwined with debates in other areas such as politics, science, or law; and the tangled questions of authority and representation that arise as a result can be clearly observed in the subcollection.

The subcollection is also exemplary of the way that information ecosystems emerge around communities with particular health issues or concerns, and it is in the potential to capture this 'information activism' that the value of web archiving tools for producing a more representative and inclusive historical record can be observed. As Andrew Flinn (2007) has noted, where the conventional archive does document historically marginalized or excluded communities "it ... rarely allows them to speak with their voice, through their own records". Instead, "traces are generally one-dimensional, often reducing individuals to statistics, appearing as problems, occupations, rigid ethnic or faith-based identities which minimize or ignore complexity and deny them their own voice" (2007, 152; 160). Web archiving can therefore be seen to offer a corrective to this, but it is important that we recognize that "the internet affords the luxury of a certain amount of distance to be able to observe people, consume information

generated by and about them, and collect their data without having to participate in equitable engagement as a way to understand their lives, communities, or concerns” (Jules, Summers, and Mitchell 2018, 3). Proponents of participatory archiving practices that invite communities to create or describe archives in their own ways have suggested that such approaches can “have an impact in diversifying and democratizing heritage” (Flinn 2007, 165) and while many mainstream organizations have experimented with inviting communities into the archival process through crowdsourced description projects or by soliciting contributions of material, it has been argued that such approaches reinforce the claim of the archive to ‘speak for’ communities: that is, in their control of the terms on which the community can engage, power over final decisions regarding appraisal, arrangement, and description still rests with the ‘experts’ (Eveleigh 2015). If web archiving programs are to engage such methods in search of a more representative record, then, we need to work with communities to find sustainable, respectful, and equitable avenues for participatory collection building.

Furthermore, the highly politicized atmosphere around the topic of trans* health made collecting this topic particularly challenging, and Eira Tansey’s observation that the historical record should not be “a high priority while people are trying to keep their shit together and attempt to not die” is particularly pertinent in the context of building a collection like the trans* health subcollection (Tansey 2020). When the subject under debate is kidney stone treatment, very few would question the right of someone experiencing symptoms to access healthcare, and even fewer would express doubt about the existence of kidney stones or kidney stone pain in the first place. In contrast, discussion surrounding trans* health issues can (and regularly does) include questions over the legitimacy of trans* identities, and the extent to which they should be recognized and respected by law. Such questions can be distressing to witness even for those outside of the trans* community. Before we ask individuals to frame and examine their personal lived experiences in this way, it is crucial that we ensure they can be adequately supported in this work. Returning to the observation on the power of archival representation that opened this chapter, we must remain aware that this is a power that must be wielded responsibly—and consider what it means to ‘discover yourself existing’ in a context that constitutes your existence as deviation from ‘the norm’. As Tansey cautions, archive and heritage professionals must recognize “that respecting people’s privacy and right to forget their own past means accepting that we will lose parts of the historical record that others may wish we had gone to great lengths to get” (Tansey 2020).

References

- Adams, Nicholas Norman. 2002. "Scraping' Reddit posts for academic research? Addressing some blurred lines of consent in growing internet-based research trend during the time of Covid-19." *International Journal of Social Research Methodology* 27, no. 1: 47–62. <https://doi.org/10.1080/13645579.2022.2111816>.
- Caswell, Michelle, Marika Cifor, and Mario H. Ramirez. 2016. "To Suddenly Discover Yourself Existing: Uncovering the Impact of Community Archives." *The American Archivist* 79, no. 1: 56–81.
- Council of Europe Committee on Equality and Non-Discrimination. 2021. "Combating rising hate against LGBTI people in Europe Doc. 15425." Belgium: Council of Europe. Retrieved 7 March 2024. <https://web.archive.org/web/20230228040752/https://pace.coe.int/en/files/29418/html>
- Eveleigh, Alexandra Margaret Mary. 2015. "Crowding out the Archivist? Implications of Online User Participation for Archival Theory and Practice." PhD diss. University College London. <https://discovery.ucl.ac.uk/id/eprint/1464116>
- Flinn, Andrew. 2007. "Community Histories, Community Archives: Some Opportunities and Challenges." *Journal of the Society of Archivists* 28, no. 2: 151–76. <https://doi.org/10.1080/00379810701611936>.
- Gill-Peterson, Jules. 2022. "Three Questions for Every Paper of Record That Publishes a Story on Trans Healthcare." *Sad Brown Girl* (blog) *Substack*. 15 June 2022. Retrieved 8 February 2024. <http://web.archive.org/web/20220630001614/https://sadbrowngirl.substack.com/p/three-questions-for-every-paper-of>
- Hunter, Ruth F., Aisling Gough, Niamh O'Kane, Gary McKeown, Aine Fitzpatrick, Bergis Jules, Ed Summers, and Vernon Mitchell. 2018. "Documenting the now-white paper: ethical considerations for archiving social media content generated by contemporary social movements: challenges, opportunities, and recommendations." *Documenting the Now*. Retrieved 23 January 2024. <https://web.archive.org/web/20240202062719/https://www.docnow.io/docs/docnow-whitepaper-2018.pdf>
- Ketelaar, Eric. 2001. "Tacit Narratives: The Meanings of Archives." *Archival Science* 1: 131–41. <https://doi.org/10.1007/BF02435644>.
- Platt, Lucinda. 2021. "Why ethnic minorities are bearing the brunt of COVID-19." *Research for the World* (blog) *London School of Economics and Political Science*. 9 November 2021. Retrieved 17 Jan 2024.

- <https://web.archive.org/web/20230925103916/https://www.lse.ac.uk/research/research-for-the-world/race-equity/why-ethnic-minorities-are-bearing-the-brunt-of-covid-19>
- Spieker, Sven. 2008. *The Big Archive: Art from Bureaucracy*. London: MIT Press.
- Tansey, Eira. 2020. "No One Owes Their Trauma to Archivists, or, the Commodification of Contemporaneous Collecting" (blog). 5 June 2020. Retrieved 15 Jan 2024. <https://web.archive.org/web/20200627194117/http://eiratansey.com/2020/06/05/no-one-owes-their-trauma-to-archivists-or-the-commodification-of-contemporaneous-collecting/>
- Topping, Alexandra. 2022. "Young trans people accessing treatment outside NHS may get safeguarding referral." *The Guardian*. 14 October 2022. Retrieved 6 February 2024. <https://web.archive.org/web/20221102130414/https://www.theguardian.com/society/2022/oct/14/young-trans-people-accessing-treatment-outside-nhs-may-get-safeguarding-referral>
- Trans Healthcare Intelligence. 2023, "Trans Healthcare Information." Retrieved 12 February 2024. <http://web.archive.org/web/20230815183354/https://www.transhealthcareintel.com/>
- TransActual. "Bridging Prescriptions." Retrieved 9 February 2024. <https://web.archive.org/web/20240121233504/https://transactual.org.uk/bridging-prescriptions/>
- Hunter, Ruth F., Aisling Gough, Niamh O’Kane, Gary McKeown, Aine Fitzpatrick, Tom Walker, Michelle McKinley, Mandy Lee, and Frank Kee. 2018. "Ethical Issues in Social Media Research for Public Health." *American Journal of Public Health* 108: 343–348. <https://doi.org/10.2105/AJPH.2017.304249>.
- Wall, Catherine S.J., Alison J. Patev, Eric G. Benotsch. 2023. "Trans broken arm syndrome: A mixed-methods exploration of gender-related medical misattribution and invasive questioning." *Social Science & Medicine* 320.