

Come cambia la professionalità della catalogazione

Stefano Bargioni

Abstract: Questo scritto riporta sostanzialmente il contributo dallo stesso titolo presentato all'inaugurazione di Parsifal l'11 maggio 2023. Si incentra sul rinnovamento del lavoro delle colleghe e dei colleghi della Rete URBE in seguito all'adozione di RDA, all'avvio e alla successiva manutenzione del progetto Parsifal, alle novità introdotte dall'intelligenza artificiale nell'ambito della ricerca di informazione in rete e al loro possibile utilizzo per la catalogazione.

Parole chiave: catalogazione per entità; controllo di autorità; entity management; intelligenza artificiale; URBE; Parsifal.

Abstract: *This paper essentially reproduces the paper of the same title presented at the inauguration of Parsifal on 11 May 2023. It focuses on the remodeling of the work of colleagues in the URBE Network following the adoption of RDA, the launch and subsequent maintenance of the Parsifal project, the innovations introduced by artificial intelligence in the field of information retrieval on the web and their possible use in cataloguing.*

Keywords: *entity-based cataloguing; authority control; entity management; artificial intelligence; URBE; Parsifal.*

1. Introduzione

La catalogazione si sta ormai orientando verso un lavoro improntato alla metadattazione, cioè all'aggiunta di informazione ai dati, per qualificarli sufficientemente e renderli disponibili in rete in modo che possano confrontarsi, interagire e integrare tra loro e con altri dati culturali (Schreur 2023, 43, *passim*). Talvolta si accusano i dati culturali di essere come nascosti, di fare parte del *deep web* invece che del *surface web*, cioè di non essere raggiungibili dai motori di ricerca e quindi di rimanere scarsamente visibili e utilizzabili¹. I silos della cultura raccolgono in

¹ Per approfondimenti su *surface*, *deep* e *dark web*, si veda <<https://www.comparitech.com/blog/vpn-privacy/using-deep-web-search-engines-for-academic-research>>.

realtà quanto di più prezioso abbia prodotto lo scibile umano, la cui fruizione diventa ogni giorno più importante. Per superare questo limite, è necessario che i dati culturali si affaccino al web semantico, che si presentino come scomposti in entità, ognuna costituente informazione autonoma, coerente e auto-esplicativa.

2. L'innovativo Parsifal e RDA

Parsifal presenta quattro modalità di ricerca: due classiche, ricerca semplice ed avanzata, e due particolari, ricerca sui record di autorità e ricerca per entità. Quest'ultima è quasi audacemente, si potrebbe dire, offerta come homepage. Le entità sono infatti l'anima non solo del server Parsifal, ma anche e soprattutto il cuore del nuovo modo in cui URBE sta lavorando dal 2017, anno in cui URBE ha adottato ufficialmente le linee guida RDA; anzi dal 2014, da quando si tenne – proprio in questa sede dell'Università Urbaniana – il primo corso su RDA per personale della Rete URBE, tenuto dal prof. Mauro Guerrini² e dalla dott.ssa Tiziana Possemato.

RDA, tra l'altro, è un metodo per valorizzare i *punti di accesso*, i possibili cammini di ricerca, percorsi – a volte inconsciamente – da studenti, ricercatori e docenti quando si avvicinano a un catalogo di dimensioni sufficientemente ampie, come nel caso di Parsifal, dove le risposte alle ricerche semplici o avanzate possono in effetti essere composte più da dati inutili che dai dati attesi dall'utente. La razionalizzazione dei dati realizzata attraverso raggruppamenti (*cluster*) e legami interni permette all'utente di percorrere cammini sicuri di approfondimento, e di giungere a quanto cercato (o anche di dedurre che quanto cerca non si trova in catalogo) in modo sicuro, non per tentativi.

Questi raggruppamenti permettono di trattare le entità in modo unitario, benché possano differire per come vengono denominate nelle diverse biblioteche: nomi di autori, di opere, di luoghi, e – si spera più avanti – di soggetti. Per ottenere i raggruppamenti, le entità prodotte dalle biblioteche con il lavoro di metadateazione dovranno essere sufficientemente identificate, non solo in base ad elementi che le rendano tali all'interno della Rete URBE, ma con criteri di valore universale, in modo che il loro reperimento e riuso fuori dalla Rete URBE sia robusto nel tempo e non solo nello spazio. Un ambiente che produce e trasmette cultura non si può permettere di descrivere più volte lo stesso autore in voci di autorità diverse³, o di attribuire opere di autori diversi a un autore

² Si ringrazia particolarmente Mauro Guerrini per aver aiutato a introdurre RDA in URBE, avere proposto e favorito la partecipazione di URBE in EURIG, European RDA Interest Group, e di averne così valorizzato il lavoro, ponendo URBE all'altezza delle biblioteche nazionali europee, come sottolineato dall'intervento precedente di Carlo Bianchini, le quali sono la maggioranza di EURIG stesso.

³ Si pensi ai casi di autori antichi denominati in modi molti vari. Ma anche ad autori contemporanei, come succede con pseudonimi, o con doppi cognomi utilizzati singolarmente o con combinazioni diverse da una pubblicazione all'altra.

solo⁴. L'attività di identificazione, per esempio nel caso dei luoghi, non può essere lasciata alla sola denominazione, spesso fortemente diversa da lingua a lingua, ma non poche volte comune a luoghi distinti tra loro, quali Houston o Rome. Ma risulta ancor più utile se si intende offrire all'utente del catalogo la possibilità di approfondire dati geografici o storici. Un semplice legame a un'altra entità presente in un repertorio autorevole di luoghi non solo soddisfa alle necessità dell'utente, ma solleva l'agenzia di catalogazione dal dover riportare, nella propria voce di autorità, informazione di cui non ha la competenza e che non manterrà.

In chimica si distingue tra emulsione, soluzione e composto. Nella storia dei cataloghi collettivi, non poche volte si è assistito a sistemi che offrono un amalgama dei dati, o se curati in qualche modo sono come una soluzione; nel caso di sistemi che lavorano per entità, si possono ottenere composti, oggetti più ricchi, con proprietà superiori alla somma delle parti, come il sodio e il cloro formano il sale da cucina, così diverso dal primo e dal secondo elemento. URBE produce e produrrà entità tramite Parsifal grazie a RDA e a norme che sono state approntate da tutte le biblioteche sotto la guida della dott.ssa Paola Manoni e più recentemente della dott.ssa Agnese Galeffi, denominate «Modello per dati di autorità per la Rete URBE»⁵.

Nell'intervento precedente, il prof. Bianchini ha usato la metafora del naufrago e della sua isola per mostrare proprio come la catalogazione abbia nuovi confini, se non addirittura li abbia del tutto persi. La catalogazione e ancor più la metadattazione avviene come se stessimo tutti in una sola piazza, o intorno a un solo tavolo, non più solo alla propria scrivania, perché la rete porta informazione a noi, così come porta i nostri dati ovunque nel web – nel web semantico, quello dove anche le macchine, nella loro ottusità, *capiscono* a cosa si riferiscono i dati.

3. Repertori e metodi di identificazione

Un fronte nuovo della metadattazione è quello delle fonti dell'informazione. Il loro moltiplicarsi presenta aspetti sui quali occorre soffermarsi e prendere delle decisioni, a livello di agenzia di catalogazione o di rete di biblioteche. Da un lato la quantità di fonti a disposizione può apparire già adesso eccessiva, così come può impensierire la difformità nella qualità. Vanno quindi prese alcune decisioni, a salvaguardia della produttività, soprattutto, e dell'omogeneità del lavoro.

Per procedere all'identificazione di un autore, ho osservato con interesse una *scaletta* logica seguita da diversi colleghi, che si interrompe al raggiungimento dello scopo e dove quasi ogni passo comporta una ricerca in rete:

⁴ Questo errore, denominato conflazione, è particolarmente difficile da individuare e correggere. È causato da omonimie non risolte dall'aggiunta di data di nascita o altro dato sufficientemente identificante. La propagazione in rete delle conflazioni va evitata il più possibile, favorendo la comunicazione tra biblioteche per la segnalazione delle correzioni.

⁵ Documento interno alla Rete URBE, 2023.

- Date biografiche
- Opere
- Autori collegati
- Case editrici
- Ambienti di lavoro
- Pubblicazioni non monografiche
- curriculum vitae, sito web personale, social web
- ... email all'autore ...

Non si tratta certo di una novità, dato che per esempio in un ambiente partecipato come SBN anche le stesse REICAT normano questa attività. Ma, a quanto affermato nel documento «Norme per il trattamento di informazioni e dati comuni a tutte le tipologie di materiale»⁶, lo fanno al solo scopo di disambiguazione interna a SBN stesso, non di identificazione tra due John Smith noti o con un altro John Smith non ancora noto⁷.

È quindi lecito chiedersi se il repertorio cartaceo tramonti. Non direi che debba essere allontanato dalle scrivanie, ma pare difficile che ne possano nascere di nuovi in questo formato. Forse la cooperazione può entrare anche qui, nel rivedere come produrli e mantenerli. Un repertorio su carta soffre dell'obsolescenza dal momento della pubblicazione, un repertorio in rete -- purché condiviso anche in modifica da un sufficiente numero di esperti -- è ora possibile e ha ben altre prospettive di utilizzo e aggiornamento continuo.

4. L'intelligenza artificiale (IA) in biblioteca

Assurta quasi improvvisamente agli onori della cronaca, benché oggetto di studio e ricerca da molti anni⁸, l'IA sta variando il modo di reperire informazioni in rete: i motori di ricerca degli ultimi venti anni hanno richiesto l'uso di parole chiave per restituire collegamenti a pagine web, mentre i nuovi motori -- ora denominabili motori di risposta o motori di dialogo -- offrono risposte a fronte di domande, significato a fronte di significato. Diventa quindi imminente,

⁶ <https://norme.iccu.sbn.it/index.php?title=Norme_comuni/Authority_file/Nomi/Registrazione_di_authority/Qualificazioni>, consultato il 13.9.2023 e attualmente vigente, afferma: «Se due o più nomi (compresi i rinvii) risulterebbero identici, pur riferendosi a persone diverse (o a una persona e a un ente), per distinguerli si aggiungono al nome una o più qualificazioni secondo il seguente ordine di preferenza: le date di nascita e/o di morte o altre indicazioni cronologiche; la forma completa del nome, il nome reale di una persona, se registrata sotto uno pseudonimo, o altro nome che aiuti a identificarla; una espressione che indichi la professione, lo stato o l'ambito principale di attività della persona».

⁷ Può aiutare anche tenere presente che le date e i luoghi di nascita e lavoro di un autore aiutano molto la ricerca, perché offrono sinteticamente un quadro della cultura a cui appartiene l'autore in studio.

⁸ Tra le tante possibili fonti sulla storia dell'IA, si veda per esempio il primo capitolo di Floridi 2022.

anche per l'utenza delle biblioteche, confrontarsi con questi nuovi strumenti⁹ e saperne trarre vantaggio tramite un uso mirato. Disporremo di sistemi generici, quali Bing¹⁰, Chat-GPT¹¹, Gemini¹², Perplexity¹³, ecc., e di sistemi più specializzati, anche orientati sui dati culturali, alcuni gratuiti, altri a pagamento. Di questi strumenti, dovranno essere valutate le modalità di interrogazione, la qualità delle risposte, l'integrazione con gli strumenti di metadateazione.

Per la metadateazione, gli ambiti di possibile applicazione sono certamente le descrizioni bibliografiche e le voci di autorità. Alcune prove effettuate nei giorni scorsi¹⁴ portano in diversi casi a discreti risultati, significativi non solo in quanto a risposte ricevute, ma forse ancora di più per i vari modi di interazione. Ecco alcuni esempi di uso di Perplexity.

Primo caso, ricerca della data di morte di un autore con nome simile a un altro:

Q: data di morte di Luigi Elvio Accattatis

A: Luigi Elvio Accattatis was an Italian historian [...] born on Nov 2, 1838, died on Jun 8, 1916.

Q: no, si tratta di un'altra persona. ripeto: "Luigi Elvio Accattatis"

A: I apologize for the confusion. According to my search results, there is no information available on Luigi Elvio Accattatis. All the search results I found refer to Luigi Accattatis (1838-1916)

Q: Luigi Elvio Accattatis è descritto in Wikidata e forse altrove

A: Thank you for providing more context. According to my search results, Luigi Elvio Accattatis was an Italian poet, journalist and writer born in 1914 [...].

Secondo caso, costruzione di un record bibliografico a partire da una pagina web di un editore:

Q: crea un record bibliografico in formato MARC 21 per il libro descritto in <https://www.libreriacoletti.it/libros-espanol/salvador-canals-una-biografia-1920---1975-mendiz-alfredo/9788432151248>

A: Ecco un esempio di record bibliografico in formato MARC 21 per il libro "Salvador Canals: una biografia (1920-1975)" di Josep Maria Solé i Sabaté:

```
LDR 00900nam a2200277 i 4500
001 123456789
005 20230413
008 040630s2004 sp d 000 0 spa d
```

⁹ Per una bibliografia sull'uso dell'IA in biblioteca, vedasi <<https://digital-scholarship.org/ai/ai-libraries.htm>>.

¹⁰ <<https://www.bing.com>>.

¹¹ <<https://chat.openai.com>>.

¹² <<https://gemini.google.com>>.

¹³ <<https://perplexity.ai>>.

¹⁴ I test si riferiscono al mese di aprile 2023.

020 \\\$a9788432151248
 100 1\\\$aSolé i Sabaté, Josep Maria
 245 10\\\$aSalvador Canals: una biografía (1920-1975) /\$cJosep Maria Solé
 260 \\\$aMadrid :\$bEdiciones Rialp,\$c2004.
 300 \\\$a300 p. ;\$c 24 cm.
 650 \\0\$aCanals, Salvador,\$d1920-1975.
 650 \\0\$aBiografie.
 651 \\0\$aSpagna\$xStoria\$y1923-1975.
 907 \\\$a.b123456789\$b123456789\$c20230413
 998 \\\$a123456789\$dm\$e123456789\$f20230413\$g0

Q: l'autore è sbagliato

A: Chiedo scusa per l'errore nella mia risposta precedente. Dopo aver effettuato una nuova ricerca, ecco il record bibliografico in formato MARC 21 per il libro "Salvador Canals: una biografía (1920-1975)":

LDR 00900nam a2200277 i 4500

001 123456789

005 20230413

008 040630s2004 sp d 000 0 spa d

020 \\\$a9788432151248

100 1\\\$aMéndiz, Alfredo

245 10\\\$aSalvador Canals: una biografía (1920-1975) /\$cAlfredo Méndiz.

260 \\\$aMadrid :\$bEdiciones Rialp,\$c2004.

300 \\\$a300 p. ;\$c 24 cm.

650 \\0\$aCanals, Salvador,\$d1920-1975.

650 \\0\$aBiografie.

651 \\0\$aSpagna\$xStoria\$y1923-1975.

907 \\\$a.b123456789\$b123456789\$c20230413

998 \\\$a123456789\$dm\$e123456789\$f20230413\$g0

Il dialogo è proseguito, facendo correggere l'anno di edizione in 2019, chiedendo di usare il tag 264 al posto del tag 260, di togliere i tag 9xx e di introdurre i tag 33x tipici di RDA. Il risultato è stato soddisfacente, ma forse la risposta finale è costata troppi passi di interazione.

È stato anche possibile ricevere i record in formato BIBFRAME, o in MARCXML, e ciò indica che il motore di IA è in grado di parlare un linguaggio standard adatto per l'import in un catalogo e quindi di potervi essere integrato.

È stata fatta anche una prova analoga per la costruzione di un record di autorità in MARC 21, con risultati analoghi.

È opportuno anche valutare l'uso dell'IA nel *reference*, settore nel quale è possibile che l'utente stesso si cimenti autonomamente, ma dove conterà molto di più l'esperienza professionale del *reference librarian* nell'uso dei motori di risposta.

5. Conclusioni

Parsifal costituisce per la Rete URBE un terreno fertile per proseguire nell'applicazione delle linee guida RDA, con la finalità di portare a pieno titolo i propri dati nel web semantico. Implica un rinnovamento di diverse logiche di catalogazione e di cooperazione, dando così una dimensione pratica molto concreta alla formazione tante volte condivisa tra le biblioteche della Rete.

Al tempo stesso, l'avvento dell'IA apporterà modifiche a tecniche e strumenti di catalogazione, così come alla consultazione del catalogo. Come ausilio alla catalogazione, se ne può apprezzare la velocità della composizione del record, ma anche notare una certa sommarietà e quindi un basso grado di affidabilità delle risposte, e la necessità di dover costringere il sistema correggendolo e orientandolo.

Probabilmente, così come i motori di ricerca non hanno spaventato le biblioteche, benché a lungo si siano confrontate con essi (basti ricordare il Convegno Stelline 2016 «Bibliotecari al tempo di Google»¹⁵), anche i motori di risposta potranno avere un ruolo nel lavoro di *back-office* e *front-office* in biblioteca.

Riferimenti bibliografici

- Schreur, Philip E. 2023. *Homo Deus - Techno-Humanism vs Dataism and the place of Information in Society*. Fiesole (Firenze): Casalini Libri.
- Floridi, Luciano. 2022. *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*. Milano: Raffaello Cortina.

¹⁵ <<https://www.convegnostelline.it/stelline2016>>.