

Nuove prospettive di evoluzione per i cataloghi

Francesca Tomasi

INTRODUZIONE

I cataloghi nel MAB (Musei, Archivi e Biblioteche) svolgono tradizionalmente il ruolo di sistemi di organizzazione del sapere, ovvero strumenti che descrivono e ordinano le risorse documentarie per supportare la ricerca dell'informazione e, auspicabilmente, il raggiungimento della conoscenza.

Nell'era digitale i cataloghi sono database/datasets dotati di metadati, di variabile grado di complessità e dettaglio, interconnessi. Il loro obiettivo rimane quello di offrire all'utente percorsi di scoperta del sapere veicolato dalle fonti e di garantire l'accesso alla conoscenza in modo coerente con bisogni sempre più diversificati. Oggi, infatti, mentre aumenta la complessità, e anche la consistenza, dei dati, emergono nuove esigenze di utenti eterogenei, ma al contempo consapevoli, con bisogni di natura spesso interdisciplinare; utenti che desiderano risposte trasversali a dati del dominio MAB diversi per provenienza, formato, supporto e tipologia documentaria.

In questo contesto si inserisce il paradigma dei LOD (Linked Open Data), che interpreta i cataloghi come nodi di un più ampio 'Web dei dati' per la costruzione di *knowledge graphs* nel cosiddetto *LOD cloud*.

La prospettiva LOD trasforma ogni record in un insieme di entità (autori, opere, soggetti, eventi temporali o spaziali) collegate attraverso relazioni semanticamente esplicite, e connesse ad altri dataset aperti (ad es. authority file internazionali come VIAF, thesauri come i vocabolari del Getty¹ o la *Subject Headings* delle Library of Congress, LCSH², e in generale soggettari³ e sistemi di classificazione, da DDC (Dewey Decimal Classification) a Iconclass⁴, ma anche fonti dati come Wikidata). Questo approccio

1 <https://www.getty.edu/research/tools/vocabularies/>

2 Accesso alla versione LOD <https://id.loc.gov/authorities/subjects.html>

3 Dal soggettario iconografico di ICCD, <https://iccd.beniculturali.it/getFile.php?id=185> al nuovo Thesaurus della BNCF, <https://thes.bnfc.firenze.sbn.it/ricerca.php>

4 <https://iconclass.org/>

dovrebbe abilitare l'interoperabilità tra sistemi informativi di biblioteche, archivi e musei, favorendo ricerche e integrazioni di risorse su scala globale.

La conversione in LOD, portata avanti in Italia dagli istituti del MiC (Ministero della Cultura)⁵ – ovvero ICCU, ICCD e ICAR –, ha imposto sfide notevoli, come ad esempio la presa di coscienza della quantità enorme di record da gestire, la variabilità delle regole di descrizione fra domini, la mancanza di codifiche standard per certe tipologie documentarie o per certi valori di attributi, le esigenze di pulizia e riconciliazione dei dati. E siamo solo all'inizio di un percorso lungo e complesso.

Ciò evidenzia che la ricchezza semantica garantita dai LOD, attraverso l'uso di schemi e ontologie o più in generale di vocabolari RDF, dipende dalla coerenza, correttezza e qualità dei metadati di base, anche ragionando in ottica di reperibilità, accessibilità, interoperabilità e riuso (ovvero anche in ottica FAIR).

Comunque sia, la diffusione dei dati aperti rende il catalogo capace di rispondere a interrogativi sempre più articolati, a query SPARQL sempre più sofisticate, e di proporre percorsi di lettura molteplici, differenziati e orientati all'utente, ma anche di consentire la costruzione di narrative, o storie, trasversali rispetto ai domini.

In questa ottica LOD, emergono due traiettorie di evoluzione complementari per i cataloghi che gli istituti di conservazione del MiC potrebbero perseguire (e che in parte stanno già realizzando).

VERSO L'ATTRATTIVITÀ: LAYOUT, STRUMENTI E SERVIZI

La prima traiettoria punta alla creazione di ambienti di ricerca innovativi, basati su sistemi di visualizzazione capaci di svelare informazioni latente (Tomasì 2022).

In prima battuta è innegabile che l'attrattività visuale di un catalogo digitale non è un dettaglio estetico, ma una componente centrale della sua efficacia (Schmidt 2009). L'interfaccia utente (UI) e l'esperienza utente (UX) devono essere progettate con attenzione: un layout disordinato, poco

5 Mi riferisco in particolare a: ICAR, SAN Open Data, <https://www.san.beniculturali.it/web/san/dati-san-lod>; ICCD, Catalogo generale dei beni culturali, <https://catalogo.beniculturali.it/>; ICCD, Dati aperti dei beni culturali, <https://dati.beniculturali.it/>; ICCU, Alfabetica, <https://alfabetica.it/web/alfabetica/>.

intuitivo o sovraccarico può ostacolare la navigazione e frustrare l'utente. Un buon catalogo deve presentare una gerarchia chiara, ovvero un'architettura logica di interfaccia efficace ed efficiente; filtri, faccette e opzioni di ricerca ben posizionati rispetto alle griglie di layout; un uso appropriato di colori e tipografia, e un design responsivo che si adatta a dispositivi mobili. L'uso di miniature, icone, sliders, cards e altri elementi visivi o componenti di interfaccia (*widgets*) può migliorare significativamente l'attrattiva del catalogo, facilitando l'interazione anche per utenti meno esperti. La grafica non deve oscurare l'informazione, ma guidare l'utente nella sua esplorazione. E l'usabilità diventa un concetto chiave per la creazione di strumenti di corredo, ovvero è un mezzo per la valutazione qualitativa delle applicazioni (Feliciati 2016).

Ma visualizzazione significa anche selezione di strumenti di accesso al contenuto veicolato dai dati. La *data visualization* rappresenta una delle innovazioni più affascinanti verso la quale i cataloghi digitali contemporanei in LOD si stanno già muovendo perché è prima di tutto strumento di trasferimento del sapere e quindi di comunicazione (Archambault 2015). Lontani dai classici elenchi testuali, i dati possono oggi essere rappresentati in modo visivo, rendendo evidenti *pattern*, connessioni e sviluppi spaziali o temporali. Giusto per fare qualche esempio: mappe geografiche consentono di localizzare eventi, pubblicazioni o autori, facilitando l'analisi spaziale di un fenomeno; linee del tempo sono molto utili per visualizzare la distribuzione cronologica di documenti, autori, eventi storici; *tag cloud* rappresentano visivamente la frequenza di parole chiave o soggetti, aiutando a cogliere rapidamente i temi prevalenti. Questi strumenti non solo arricchiscono l'interazione, ma migliorano anche la comprensione: un utente può farsi un'idea globale del contenuto prima ancora di leggerne i dettagli (Po *et al.* 2020). Inoltre, permettono una navigazione esplorativa che stimola la curiosità e favorisce la scoperta non sequenziale o non lineare. La visualizzazione può essere poi pensata anche come *visual analytics*, quindi strumento per analizzare tendenze e correlazioni nei dati (Windhager *et al.* 2019). Questi sistemi si basano su rappresentazioni grafiche interattive; ogni resa grafica è il risultato di specifiche domande di ricerca poste ai dati, sulla base tanto del 'tipo di dati' quanto del tipo di risposta che si ambisce a ricevere, secondo, per esempio, il modello di *extreme presentation method*, dove il layout è fondamentale (Abela 2008). Istogrammi e grafici a torta sono ad esempio ideali per rappresentare frequenze e proporzioni; *treemaps* visualizzano gerarchie e volumi informativi con l'uso di rettangoli proporzionali; grafi rappresentano relazioni tra entità (es.

co-autorship, citazioni, legami tematici). Vari sono gli strumenti che permetterebbero di visualizzare e analizzare metadati estratti da cataloghi (ad esempio Gephi⁶, VOSviewer⁷, RAWGraphs⁸ o Palladio⁹ ma anche librerie diverse come D3.js¹⁰), offrendo un approccio visuale a diverse tipologie di relazioni e interazioni fra dati diversi, come, per fare qualche esempio, nel campo della bibliometria (e.g. relazioni citazionali, collegamenti bibliografici, co-citazioni o relazioni di paternità), della storia delle idee o della genealogia intellettuale (ovvero anche a relazioni e influenze fra studiosi).

Ma una delle trasformazioni più radicali a cui i cataloghi digitali potrebbero aderire è la dimensione della partecipazione attiva. Questa evoluzione è legata all'emergere di servizi che permettono all'utente di non essere solo un passivo fruitore di dati, ma un co-creatore di contenuti, ovvero un produttore di conoscenza (Bonora *et al.* 2024). Gli utenti potrebbero selezionare risorse, ordinarle, commentarle e condividerle, realizzando narrazioni personalizzate. Si tratta di costruire percorsi tematici, selezionando materiali di interesse su un certo argomento, magari corredate da commenti e testi introduttivi; creare collezioni digitali personalizzate, dove gli utenti potrebbero creare raccolte visuali; esplorare forme di *storytelling* interattivo in cui documenti, immagini e video sono collegati per formare una narrazione coerente, utile per fini didattici o divulgativi. In questo modo, il catalogo diventerebbe una piattaforma narrativa e creativa, che incoraggia l'utente a 'raccontare con i dati', attraverso la creazione di diverse storie (si vedano ad esempio le funzionalità di Melody Stories¹¹). Si tratta di una modalità innovativa, che valorizza sia il contenuto, ovvero la capacità dei dati di trasmettere conoscenza, sia l'esperienza attiva dell'utente.

Un altro aspetto importante su cui si potrebbe riflettere è l'introduzione di sistemi di feedback, mutuati dalle logiche dei social network e dell'e-commerce. Gli utenti potrebbero lasciare commenti su una risorsa, dare valutazioni (*rating*), segnalare contenuti con 'like' o 'thumbs up'. Questi sistemi non servono solo a coinvolgere l'utenza, ma possono alimentare 'motori di raccomandazione', che suggeriscono materiali in base a interessi, comportamenti e preferenze. In questo modo il catalogo diventerebbe più intelligente e personalizzato (Coyle 2011). L'interazione sociale, inoltre, con-

6 <https://gephi.org/>

7 <https://www.vosviewer.com/>

8 <https://www.rawgraphs.io/>

9 <https://hdlab.stanford.edu/palladio/>

10 <https://d3js.org/>

11 <https://melody-data.github.io/stories/>

tribuisce, in generale, a creare una comunità attorno al contenuto, in cui l'utente sente di avere un ruolo attivo e significativo. Questo aiuterebbe ad aumentare la fidelizzazione e la frequenza d'uso della piattaforma.

I cataloghi digitali, infine, non si dovrebbero più limitare al solo testo: oggi si possono integrare e fare interagire immagini, audio, video, modelli 3D, fino a elementi di realtà aumentata. L'obiettivo è offrire una esperienza multimediale immersiva, capace di rispondere a bisogni di diversa natura, ovvero educativi, divulgativi o anche ludici. Questa transmedialità espanderebbe le possibilità narrative, rendendo il catalogo un crocevia di linguaggi e supporti, adatto a pubblici differenti per età, competenze e interessi.

Insomma, questi strumenti e servizi possono trasformare il catalogo in un ecosistema funzionale, capace di accompagnare l'utente lungo l'intero ciclo della ricerca e della fruizione. La presenza di numerose API (Application Programming Interface) pubbliche consentirebbe inoltre l'integrazione dei cataloghi con altri software e piattaforme, favorendo la connessione tra sistemi bibliotecari, archivistici e museali, arricchendoli con applicativi di diversa origine e provenienza, capaci di erogare servizi differenziati, come quelli sopra descritti.

ARRICCHIMENTO DEI METADATI PER UTENZE SPECIALIZZATE

La seconda traiettoria pone al centro l'arricchimento qualitativo di dati e metadati, finalizzato alle esigenze di un'utenza altamente specializzata, come ricercatori e studiosi. Questo pubblico richiede informazioni attendibili e documentate, nonché un accesso critico alla conoscenza veicolata dai dati; gli studiosi hanno bisogno di ambienti che vadano oltre la mera restituzione di dati fattuali, ovvero anche delle canoniche informazioni di paratesto.

Il paradigma LOD, come noto, già favorisce un primo grado di arricchimento perché consente di collegare ogni entità al rispettivo authority file e a vocabolari controllati per certi attributi (come ad esempio accade per i sistemi di soggettazione o per le tipologie documentarie), garantendo così identificatori univoci e contesti verificati. In questo modo un autore viene automaticamente collegato ad un'identità globale, un luogo storico è georeferenziato, una fonte archivistica è contestualizzata e le entità quindi, anche se vengono da domini diversi e sono descritte attraverso altrettanto diverse ontologie, sono ricondotte ad unità e la ricerca diventa trasversale. Raggiungere questo obiettivo in un'ottica MAB significa operare un

lungo lavoro di normalizzazione, omogeneizzazione e di *data cleaning*. La copiosa quantità di record provenienti da epoche diverse e l'assenza di codici uniformi impongono un paziente lavoro di standardizzazione dei dati: ogni elemento deve essere rappresentato in modo che abbia significato univoco nel nuovo sistema LOD. E già questa è una sfida per gli istituti di conservazione del patrimonio culturale.

Ma la riconciliazione ai sistemi di controllo di autorità è solo il primo passo per la realizzazione autentica di open data che possano essere davvero *linked*. In pratica, l'obiettivo è ambire ad una catalogazione di nuova generazione: più analitica, con più metadati, e quindi con più dati descrittivi, più connessa con strumenti esterni capaci di estendere le informazioni di base e più aderente alla complessità intrinseca delle risorse culturali.

Questo significa, ad esempio, che i dati del catalogo dovrebbero essere arricchiti con i dati esito dell'esperienza di ricerca condotta sulle risorse culturali; dovrebbero quindi interagire, ad esempio, con *repositories* accademici istituzionali come IRIS, con le riviste open access e con gli aggregatori come SCOPUS o WOS (Web of science), per far dialogare gli oggetti culturali con le fonti secondarie.

Ma questo vuol dire anche estendere le potenzialità degli attuali standard di descrizione di origine strettamente archivistica (dallo schema Encoded Archival Description, EAD¹² all'ontologia Record in Contexts, RiCO¹³), bibliografica (dal modello astratto dei Functional Requirements for Bibliographic Records, FRBR¹⁴ alla concettualizzazione con IFLA LRMoo¹⁵, o ancora Resource Description and Access, RDA¹⁶ assieme a Bibframe¹⁷ della Library of Congress) e museale (dallo standard metodologico Cataloging Cultural Objects, CCO¹⁸ alla famosa ontologia CIDOC CRM¹⁹) non solo collegandoli fra di loro – come già accade grazie all'impiego di sistemi di *mapping* o alle tecniche di allineamento, che sempre più numerosi sono disponibili in rete –, ma facendoli dialogare con altri modelli capaci di do-

12 <https://www.loc.gov/ead/ead3schema.html>

13 <https://www.ica.org/standards/RiC/ontology>

14 Per cui si veda IFLA Study Group on the Functional Requirements of Bibliographic Records. 1998. "Functional Requirements of Bibliographic Records: final report." München: K. G. Saur. Available online at <http://www.ifla.org/VII/s13/frbr/frbr.pdf>.

15 <https://repository.ifla.org/handle/20.500.14598/3677>

16 <https://www.loc.gov/aba/rda/>

17 <https://www.loc.gov/bibframe/>

18 <https://www.vraweb.org/cco>

19 <https://cidoc-crm.org/>

cumentare, ad esempio, la *provenance* degli asserti (come PROV-O²⁰) a garanzia dell'attendibilità, descrivere il sistema citazionale (come CITO²¹, del gruppo delle SPAR ontologies²²) per far dialogare le fonti fra di loro, lavorare sul testo pieno delle opere (come Text Encoding Initiative, TEI²³) per valorizzare la dimensione filologica della rappresentazione dei documenti, documentare l'atto interpretativo dello studioso (come HiCO²⁴) per estendere la conoscenza sugli oggetti culturali (Daquino Tomasi 2015).

A tutto ciò si aggiunge che i dati del catalogo sono spesso 'neutrali' e non documentano il potenziale dibattito critico sulle fonti (Barabucci *et al.* 2022) che per molte risorse culturali è fondamentale (si pensi a casi illustri, che vanno dalla Donazione di Costantino all'identità di Elena Ferrante).

Insomma, varie sono le azioni che è possibile intraprendere per realizzare questa seconda traiettoria: documentare l'interpretazione dello studioso sugli oggetti culturali; mettere a disposizione informazioni sulle fonti secondarie nella forma di citazioni bibliografiche; creare strumenti di raccolta delle opinioni degli *scholars* sulle risorse; disporre di piattaforme per l'annotazione, anche interdisciplinare, di testi e immagini o anche per l'aggiunta di commenti; realizzare ambienti per la gestione di spazi individuali distribuiti.

Questi sono i temi che è necessario discutere per trasformare il catalogo da sistema di ricognizione di risorse a spazio culturale per accogliere il dibattito critico. In tal senso, la disponibilità di interfacce studiate per questo tipo di fruizione (es. visualizzatori digitali IIF, account personali per salvare e annotare ricerche) supporta l'utenza specialistica nell'usare i dati con cognizione di causa.

In tal modo, il catalogo riflette la complessità del sapere critico; lo rende immediatamente interrogabile, ma allo stesso tempo documentato in modo esauriente per gli studiosi (Daquino *et al.* 2020).

CONCLUSIONI

Le nuove prospettive per i cataloghi si fondano dunque su un duplice principio: da un lato, ambienti di ricerca interattivi che sfruttano visualizzazioni

20 <https://www.w3.org/TR/prov-o/>

21 <http://purl.org/spar/cito>

22 <http://www.sparontologies.net/>

23 <https://www.tei-c.org/>

24 <http://purl.org/emmedi/hico>

dinamiche e sistemi di analisi semantica per soddisfare bisogni informativi ampi e complessi; dall'altro, metadati arricchiti che rispondono ai requisiti di accuratezza e attendibilità, capaci di garantirsi la fiducia (o anche *trust*) da parte di utenti specialistici. Entrambe le traiettorie sono sostenute dal paradigma Linked Open Data: i dati aperti e collegati offrono la flessibilità necessaria per implementare interfacce innovative e, allo stesso tempo, forniscono il contesto necessario per una conoscenza affidabile. In altri termini, quella che potremmo chiamare una 'visualizzazione informativa' rende il catalogo un potente mezzo per far emergere *insight* nascosti nei dati, mentre l'adozione di modelli descrittivi avanzati assicura che l'utente possa lavorare su informazioni certe e documentate.

L'obiettivo finale è bilanciare queste due dimensioni: progettare cataloghi LOD come portali di conoscenza integrata in cui l'utente possa esplorare visivamente la ricchezza dei dati e, al contempo, avere a disposizione descrizioni complete e attendibili delle fonti. Così facendo, i cataloghi moderni diventeranno autentici sistemi di organizzazione del sapere, capaci di rispondere alle esigenze più diversificate e di stimolare nuove forme di ricerca interdisciplinare.

Guardando al futuro, che è ormai il presente, possiamo immaginare cataloghi potenziati dall'intelligenza artificiale, in grado di offrire sistemi di classificazione automatica, HWR/HTR (Hand Written Recognition / Hand Text Recognition) sempre più attendibile, esperienze predittive, realtà aumentata per l'accesso immersivo e una sempre maggiore personalizzazione del contenuto. In questo scenario, che mette comunque al centro l'utente e i suoi bisogni informativi, il catalogo diventa uno spazio culturale partecipativo, in cui la conoscenza non solo si conserva e si descrive, ma si vive e si reinventa ad ogni accesso.

BIBLIOGRAFIA

- Abela A. 2008, *Advanced Presentations by Design: Creating Communication that Drives Action*, John Wiley & Sons.
- Archambault, S.G. et al. 2015, *Data visualization as a communication tool*, in «Library Hi Tech News», 32 (2), pp. 1-9.
- Barabucci G. et al. 2022, *Modeling data complexity in public history and cultural heritage*, in Serge Noiret et al. (edited by) *Handbook of Digital Public History*, De Gruyter, Oldenbourg, pp. 459–474.

- Bonora Paolo *et al.* 2023, *Department Digital Library: from a collector of digital resources to a tool for research, teaching and the third mission*, in «Bibliothecae.It», 12 (2), pp. 443–463.
- Coyle K. 2011, *Visualizing linked data*, in «Coyle's InFormation», 24 April 2011, <https://kcoyle.blogspot.com/2011/04/visualizing-linked-data.html>.
- Daquino M. *et al.* 2020, *Knowledge Representation of digital Hermeneutics of archival and literary Sources*, in «JLIS.it», 11 (3), pp. 59-76.
- Daquino Marilena; Tomasi Francesca, *Historical Context (HiCo): a conceptual model for describing context information of cultural heritage objects*. In: *Metadata and Semantics Research. (MTSR 2015)*, Manchester, UK, September 9-11, 2015, edited y Emmanouel Garoufallou, Richard J. Hartley, Panorea Gaitanou. (Communications in Computer and Information Science, vol 544). Berlin: Springer, 2015, p. 424-436. DOI:10.1007/978-3-319-24129-6_37.
- Feliciati Pierluigi, *Usability of library and archives' digital environments as a quality requirement. Context, models, and evaluation tools*, «JLIS.it», 7, (2016), n. 1. DOI: 10.4403/jlis.it-11512
- Po Laura [*et al.*], *Linked Data Visualization: Techniques, Tools, and Big Data*. Berlin: Springer, 2020. DOI:10.2200/S00967ED1V01Y201911WBE019.
- Schmidt, Aaron. "Catalog by Design: The User Experience." *Library Journal*, vol. 134, no. 2, 2009, pp. 26–28.
- Tomasi Francesca, *Organizzare la conoscenza: Digital Humanities e Web semantico*. Milano: Editrice Bibliografica, 2022.
- Windhager, Florian & Federico, Paolo & Schreder, Günther & Glinka, Katrin & Dörk, Marian & Miksch, Silvia & Mayr, Eva. (2019). Visualization of Cultural Heritage Collection Data: State of the Art and Future Challenges. *IEEE Transactions on Visualization and Computer Graphics*. 25. 2311 - 2330. 10.1109/TVCG.2018.2830759.

NOTE BIOGRAFICHE

Francesca Tomasi è professoressa ordinaria in Archivistica, Bibliografia e Biblioteconomia presso l'Università di Bologna, dove si occupa di informatica umanistica (digital humanities), con una speciale attenzione ai sistemi di organizzazione della conoscenza nel Web semantico. Coordina il Dottorato in Patrimonio Culturale nell'Ecosistema Digitale (CHeDE), ed è direttore del Centro di ricerca sulle Digital Humanities (/DH.arc). È sta-

ta presidente dell'Associazione per l'Informatica Umanistica e la Cultura Digitale (AIUCD), e infine ha fondato, ed è stata coordinatrice, del corso di laurea magistrale internazionale in Digital Humanities and Digital Knowledge (DHDK).