JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-714

JLIS.it

Agnese Galeffi: una di noi

Mauro Guerrini

Contact: Mauro Guerrini, mauro.guerrini@unifi.it

Agnese Galeffi (31 agosto 1974-24 settembre 2025) è stata una persona straordinaria per carisma, ironia, eleganza, cordialità, capacità di ascolto e di comunicazione, una donna luminosa e acuta, amante della convivialità e della buona tavola, dei viaggi e della moda, di grande curiosità culturale e di un'intelligenza straordinariamente mobile, capace di unire precisione, acume critico e visione d'insieme; e una professionista impeccabile e scrupolosa con una competenza che derivava dalla conoscenza profonda della filosofia, della storia e della pratica della catalogazione. È stata tra le poche bibliotecarie e tra i pochi bibliotecari italiani protagonisti sulla scena internazionale, in particolare nei congressi e negli incontri IFLA. Non partecipò al congresso IFLA 2009 di Milano e sentì l'assenza come una colpa, tant'è che promise di non perderne neppure uno in futuro. Agnese, fluent English (lingua che curava meticolosamente), negli anni ha creato una rete di relazioni con bibliotecari europei, americani e di altri paesi. Ha lavorato con Dorothy McGarry all'aggiornamento di ICP, *International cataloguing principles*, del 2016 (e con una precisazione del 2017)[1] e con Elena Escolano Rodríguez e Massimo Gentili-Tedeschi alla nuova versione di ISBD. Ricordo la partecipazione comune al Symposium on Managing the Cultural Heritage and Libraries in Local Governments di Bursa, in Turchia, nel 2017, selezionata tra le migliori professioniste e ricercatrici europee.

In Italia il suo nome è legato alla docenza decennale presso la Scuola vaticana di Biblioteconomia, con un costante riferimento scientifico al prof. Paul Gabriele Weston, suo "Maestro professionale e umano", come ha riconosciuto più volte, con cui ha collaborato anche all'Università di Pavia e a una ricerca condotta per conto della Regione Toscana, e con cui ha firmato vari contributi. Straordinaria era la sua qualità di docente nei contesti più diversi, come ha dimostrato nei corsi su MARC, ISBD, AACR2, RDA, ICP, IFLA LRM, ... per l'AIB, Regioni e altri enti, alcuni con l'amica Lucia Sardo, co-autrice di libri e saggi, fra cui uno su FRBR per la serie ET.[2]

Magistrale, fra i tanti suoi articoli, *Se il catalogo parlasse, lo capiremmo? Cinque assiomi della comunicazione catalografica* pubblicato in *AIB studi* del 2017, testo rielaborato di una relazione presentata al XVIII Workshop di Teche del Mediterraneo, promosso da Maria Abenante.[3]

[1] IFLA Cataloguing Section e IFLA Meetings of Experts on an International Cataloguing Code. 2017. *Statement of international cataloguing principles (ICP). 2016 Edition* 2016 *with minor revisions, 2017.* A cura di Agnese Galeffi, María Violeta Bertolini, Robert L. Bothmann, Elena Escolano Rodríguez, e Dorothy McGarry. https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/icp/icp_2016-en.pdf. Traduzione italiana: *Dichiarazione di principi internazionali di catalogazione. Edizione 2016 con piccole correzioni, 2017.* A cura di Massimo Gentili-Tedeschi, Carlo Bianchini, Marina Cennamo, Danilo Deana, Agnese Galeffi, Mauro Guerrini, Paola Manoni. https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/icp/icp_2016-it.pdf.

[2] Guerrini, Mauro. 2013. "Recensione di *Agnese Galeffi, Lucia Sardo, FRBR*." *Biblioteche oggi* 31 (8): 73-75.

[3] Galeffi, Agnese. 2017. "Se il catalogo parlasse, lo capiremmo? Cinque assiomi della comunicazione catalografica." *AIB*

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-714

Agnese è stata consulente per la costruzione della BEIC, Biblioteca europea di informazione e cultura (sempre con Lucia Sardo), e referente e docente per i bibliotecari di URBE, Unione romana biblioteche ecclesiastiche, una collaborazione fondamentale, proprio nel passaggio da AACR2 a RDA. L'11 maggio 2023, non casualmente, Agnese ha moderato e coordinato gli interventi presentati al convegno per l'inaugurazione del Catalogo Parsifal, alla cui redazione aveva collaborato, con gli atti editi a cura di Silvano Danieli da Firenze University Press nel 2024.

Ho avuto modo di conoscerla al concorso per assistente bibliotecario all'Università di Firenze del 1998; risultò vincitrice con altre bibliotecarie e altri bibliotecari poi divenuti bravissimi anche sul campo e non solo alla prova scritta e orale. Dal 2000 al 2007 ha lavorato alla biblioteca di Scienze sociali da cui si dimise per problemi di organizzazione personale, dato che il suo riferimento era Roma. Nel 2018 è entrata di nuovo in ruolo come bibliotecaria all'Università La Sapienza di Roma, con l'attribuzione della responsabilità del catalogo e del Polo RMS-SBN; la sua competenza tecnica è stata riconosciuta dall'ICCU che un anno fa l'ha chiamata a far parte della Commissione Reicat; sarebbe stata un elemento fondamentale per il suo background internazionale e la sua conoscenza di RDA.

Nel 1999-2000 l'avevo incontrata anche alla SSAB, Scuola speciale per archivisti e bibliotecari, in cui insegnavo Biblioteconomia. Mi chiese la tesi e le proposi di studiare il pensiero e l'opera di Seymour Lubetzky. Prese un aereo e nell'agosto 2001 andò a incontrarlo a Los Angeles. Nell'occasione fece una scoperta sensazionale: rintracciò 15 fogli manoscritti inediti lasciati dal Maestro (con altra documentazione) all'UCLA, University of California, Los Angeles, che riguardavano commenti circa le posizioni di Eva Verona sui principi di catalogazione presentate all'ICCP, International Conference on Cataloguing Principles. Agnese proseguì gli studi all'Università di Udine, dove superò la prova di ammissione al dottorato in Biblioteconomia diretto da Attilio Mauro Caproni. Continuò a studiare Lubetzky; terminò nel 2007 discutendo la tesi dal titolo *L'eredità di Lubetzky nei sistemi bibliografici del XXI secolo*, io suo tutor e Maria Teresa Biagetti co-tutor. Lo studio su Lubetzky l'aveva portata a studiare le ragioni dell'interesse del Maestro per Panizzi, di cui Agnese aveva compreso il nucleo di fondo meglio di altri.[4] Le proposi di pubblicare un contributo su CCQ, ma ha scritto un documentato saggio nel volume *La trasmissione della conoscenza registrata*, curato da Carlo Bianchini e Lucia Sardo, pubblicato da Editrice Bibliografica nel 2021.

Agnese è stata membro dell'International Scientific Committee di *JLIS.it* fin dall'inizio delle pubblicazioni nel 2009. Ad agosto aveva dato ancora una volta la sua disponibilità come revisore di un saggio, segno evidente della sua forza d'animo nel combattere il suo male che la inseguiva da anni.

La sua memoria e la sua eredità scientifica e professionale resteranno vive. Ci sarà modo di parlare della varietà dei suoi interessi. Ne ricordo uno molto caro. In un invito a casa sua cucinò le mitiche polpette in vendita allo spaccio del Vaticano, accompagnate da un vino altrettanto eccellente.

Gli allievi sono figli intellettuali e la loro scomparsa colpisce così profondamente da provocare un dolore immenso.

studi 57 (2): [239]-252. Il testo della relazione è disponibile su YouTube all'indirizzo: https://www.youtube.com/watch?v=JZOOpwjHF1E.

[4] Galeffi, Agnese. 2009. "Biographical and cataloguing common ground: Panizzi and Lubetzky, kindred spirits separated by a century." *Library & information history* 25 (4): [227]-246.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-707

JLIS.it

A scientific journal's contribution to the debate on shared cataloging in Italy and the future of SBN, the National Library Service

Mauro Guerrini[(a)]

a) University of Florence, https://orcid.org/0000-0002-1941-4575

Contact: Mauro Guerrini, mauro.guerrini@unifi.it
Received: 11 September 2025; **Accepted:** 03 October 2025; **First Published:** 15 January 2026

ABSTRACT
Nearly 45 years after the birth of the National Library Service (SBN), a highly valuable bibliographic tool, *JLIS.it* resumes, from its proper academic journal perspective, the debate on cataloging cooperation in Italy, and the national bibliographic services promoted for the benefit of the research community and the broader reading public. The journal wishes to offer a proactive contribution at a time when many bibliographic agencies and libraries are experiencing a long transition period that is forcing them to thoroughly rethink their mission. Therefore, *JLIS.it* has chosen to focus this issue on a problematic axis, with a critical rather than celebratory dimension, and with an open eye to the innovative global context. It seeks to seize the opportunities offered by the language of the web and to highlight some Italian experiences that are useful for redefining cataloging cooperation in the contemporary context.

KEYWORDS
SBN; Shared cataloging.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-707

JLIS.it

To Diego Maltese
on the occasion of the publication of *Con metodo e con rigore*[1]

Nearly 45 years after the birth of the National Library Service (SBN), a highly valuable bibliographic tool, *JLIS.it* resumes, from its proper academic journal perspective, the debate on cataloging cooperation in Italy, and the national bibliographic services promoted for the benefit of the research community and the broader reading public. The journal wishes to offer a proactive contribution at a time when many bibliographic agencies and libraries are experiencing a long transition period that is forcing them to thoroughly rethink their mission. Therefore, *JLIS.it* has chosen to focus this issue on a problematic axis, with a critical rather than celebratory dimension, and with an open eye to the innovative global context. It seeks to seize the opportunities offered by the language of the web and to highlight some Italian experiences that are useful for redefining cataloging cooperation in the contemporary context.

This reflection is the second JLIS.it seminar, following the first, entitled "Modeling Knowledge: Comparing Archival and Bibliographic Standards", whose contributions are available in issue no. 3 of 2022.[2] Again, in keeping with the journal's style, the speakers invited young librarians who already hold important national roles, as well as experienced librarians and teachers with in-depth knowledge of the topic.

The contemporary bibliographic ecosystem — with the coexistence of print resources (still numerous) and digital ones — and the models that seek to represent it, require a renewal of the structural and terminological approach. Many questions and challenges arise; the primary goal is to understand the best direction to take, or, more realistically, the best path to take to ensure a high-quality future for Italian cataloging cooperation, which must always be integrated into the international context.

A first series of questions leads to answers that are anything but marginal.

How can we reconcile the original vision of SBN with the challenges of the digital age and the semantic web? What conceptual and technological perspectives can we hope for SBN to update and strengthen the effectiveness of its service? How can we include Italian stakeholders beyond institutional ones in the collaborative space, such as internationally recognized bibliographic agencies that have engaged libraries over the years with cutting-edge initiatives? How can we renew and strengthen cooperation with other communities, such as Ex Libris ITALE users, enriching it with content that enhances the role of the SBN, promotes the interoperability of bibliographic data, and fosters the shared development of projects based on Linked Open Data? Have the conceptual and technological innovations that characterize the digital era been matched, in Italy, by a vision capable of metabolizing the new international cooperation paradigm? Do the bibliographic models developed between the 20th and 21st centuries lead to a revival of the idea of an original SBN Index as a service hub, in favor of a distributed network of knowledge bases? How can we

[1] (Maltese 2025). The issue is a tribute to Diego Maltese who has worked hard to create the National Library Service (SBN), and is a wish to ICCU to renew it, in due time, in an increasingly collaborative perspective at the Italian and international level.

[2] https://www.jlis.it/index.php/jlis/issue/view/36.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-707

distinguish the international goals of the National Bibliography from those of building a library catalogue, albeit within a national and global perspective? How does the BNI fit into the current bibliographic and technological context, in light of the document issued by the IFLA Bibliography Section, "IFLA common practices for national bibliographies in the electronic age", translated and published in Italian by the AIB in 2022 under the title "Pratiche condivise per le bibliografie nazionali nell'età digitale?". How can we recover the fruitful collaboration between the BNI and *Casalini Libri* from 1968 to 1984 – from the BNI's XI (1968) to XXVII (1984) years – interrupted, writes Diego Maltese, "due to an absurd decision taken from above"? What is the relationship between Central National Library of Florence (BNCF), Central National Library of Rome (BNCR), National Library Service catalogue and network (ICCU), Italian National Bibliography (BNI), and SBN? How is digital legal deposit structured, and consequently, how is national bibliographic control evolving? What is the relationship between the renewed concepts of national book archives and national bibliography? What contribution has the collaborative relationship with Wikidata made, and will it make, both to the management of SBN's authority data with a view to their reuse and enrichment, and to the more general management of the catalog (or whatever it will be called in the future) by emphasizing the importance and role of identifiers? How can we ensure data interoperability across the web in heterogeneous domains? What implications might the new, rapid developments in artificial intelligence have for cataloging and metadata? Can machine learning algorithms help metadata creators recognize identical, similar, or novel entities represented in catalog records, given the importance of data quality control and the challenging clustering process? What awareness is there of the philosophy of metadata, that is, the shift — based on the IFLA LRM conceptual model — of the focus from resources to the bibliographic relationships that characterize them (responsibility, curation, translation, etc.)? What role can advanced research on entity modeling play, an approach that requires a paradigm shift in data recording characterized by the identification of entities through identifiers, tools typical of the semantic web? Is identifying entities functional to the goal of Universal Bibliographic Control, the most ambitious project ever conceived in the library sciences field, in which every country is called to participate. How can we facilitate a new level of sustainable cooperation for the creation, maintenance, and long-term protection of quality data? How should libraries, the BNI, and the SBN prepare for growing "hybrid threats," i.e., attacks that have multiple purposes and act on multiple levels of a system?
These are questions that need to be considered in order to propose solutions that authoritatively reposition the SBN in the contemporary international cooperative context. References to the international role assumed by non-institutional bibliographic agencies and to some experiences promoted by university libraries across the country clearly indicate the need for the SBN to evolve toward greater semantic interoperability, based on the adoption of open standards and shared vocabularies.
The topic is complex and broad; *JLIS.it* therefore declares its willingness to continue the discussion by hosting further contributions on the topic.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-707

JLIS.it

Il contributo di una rivista scientifica alla riflessione sulla catalogazione condivisa in Italia e sul futuro di SBN

Mauro Guerrini[(a)]

a) University of Florence, https://orcid.org/0000-0002-1941-4575

Contact: Mauro Guerrini, mauro.guerrini@unifi.it
Received: 11 September 2025; **Accepted:** 03 October 2025; **First Published:** 15 January 2026

ABSTRACT
A circa 45 anni dalla nascita del Servizio bibliotecario nazionale (SBN), uno strumento bibliografico di grande valore, *JLIS.it* riprende, dalla prospettiva che le è propria di rivista accademica, il dibattito sulla cooperazione catalografica in Italia e sui servizi bibliografici nazionali promossi a beneficio della comunità dei ricercatori e del più ampio pubblico dei lettori. La rivista desidera offrire un contributo propositivo, in un momento in cui molte agenzie bibliografiche e molte biblioteche stanno vivendo una fase di lunga transizione che le obbliga a ripensare a fondo la propria mission. Pertanto, *JLIS.it* ha scelto d'impostare il fascicolo su un asse problematico, con una dimensione critica e non celebrativa, e con lo sguardo aperto al contesto globale innovativo, cercando di cogliere le opportunità offerte dal linguaggio del web e di valorizzare alcune esperienze italiane, utili per ridefinire la cooperazione catalografica nella dimensione contemporanea.

PAROLE CHIAVE
SBN; Catalogazione condivisa.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-707

JLIS.it

A Diego Maltese
in occasione della pubblicazione di *Con metodo e con rigore*[3]

A circa 45 anni dalla nascita del Servizio bibliotecario nazionale (SBN), uno strumento bibliografico di grande valore, *JLIS.it* riprende, dalla prospettiva che le è propria di rivista accademica, il dibattito sulla cooperazione catalografica in Italia e sui servizi bibliografici nazionali promossi a beneficio della comunità dei ricercatori e del più ampio pubblico dei lettori. La rivista desidera offrire un contributo propositivo, in un momento in cui molte agenzie bibliografiche e molte biblioteche stanno vivendo una fase di lunga transizione che le obbliga a ripensare a fondo la propria mission. Pertanto, *JLIS.it* ha scelto d'impostare il fascicolo su un asse problematico, con una dimensione critica e non celebrativa, e con lo sguardo aperto al contesto globale innovativo, cercando di cogliere le opportunità offerte dal linguaggio del web e di valorizzare alcune esperienze italiane, utili per ridefinire la cooperazione catalografica nella dimensione contemporanea. La riflessione costituisce il secondo *Seminario JLIS.it*, dopo il primo, intitolato *Modellare la conoscenza, standard archivistici e bibliografici a confronto*, i cui contributi sono disponibili sul n. 3 del 2022.[4] Anche in questo caso, nello stile della rivista, sono stati invitati bibliotecari giovani che già ricoprono ruoli importanti a livello nazionale e bibliotecari esperti e docenti che conoscono la questione a fondo.

L'ecosistema bibliografico contemporaneo – con la compresenza di risorse a stampa (ancora numerose) e digitali – e i modelli che cercano di rappresentarlo, impongono un rinnovamento d'approccio strutturale e terminologico. Molte sono le domande e le sfide che si pongono; occorre primariamente cercare di capire quale sia la direzione migliore verso cui indirizzarsi o, più realisticamente, quale sia quella possibile per assicurare un futuro di qualità alla cooperazione catalografica italiana, che dev'essere sempre inserita nel contesto internazionale.

Una prima serie di interrogativi rinvia a risposte tutt'altro che marginali.

Come coniugare la visione originaria di SBN con le sfide dell'era digitale e del web semantico? Quali prospettive concettuali e tecnologiche auspicare per SBN affinché possa aggiornare e rafforzare l'efficacia del suo servizio? Come includere nello spazio di collaborazione attori italiani diversi da quelli istituzionali, quali agenzie bibliografiche riconosciute in ambito internazionale che negli anni hanno coinvolto le biblioteche con iniziative d'avanguardia? Come rinnovare e potenziare la cooperazione con altre comunità, come quella degli utenti di Ex Libris ITALE, arricchendola con contenuti che valorizzino il ruolo di SBN, promuovano l'interoperabilità dei dati bibliografici e favoriscano lo sviluppo condiviso di progetti basati sui Linked Open Data? Alle innovazioni concettuali e tecnologiche che caratterizzano l'era digitale ha corrisposto, in Italia, una *vision* capace di metabolizzare il nuovo paradigma internazionale della cooperazione? I modelli bibliografici emanati a cavallo fra XX e XXI portano a recuperare l'idea di un Indice SBN originario quale nodo al servizio, a favore di una rete distribuita di basi di conoscenza? Come distinguere gli scopi interna-

[3] (Maltese 2025). Il fascicolo è un omaggio a Diego Maltese che molto si è speso per la realizzazione del Servizio bibliotecario nazionale ed è un augurio all'ICCU perché lo rinnovi, coi tempi necessari, in una prospettiva sempre più collaborativa a livello italiano e internazionale.

[4] https://www.jlis.it/index.php/jlis/issue/view/36.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-707

zionali della Bibliografia nazionale da quelli della costruzione di un catalogo di biblioteca, seppur inserito in un'ottica nazionale e globale? Come si pone la BNI nell'attuale contesto bibliografico e tecnologico, alla luce del documento emanato dall'IFLA Bibliography Section, *IFLA common practices for national bibliographies in the electronic age*, tradotto e pubblicato in italiano dall'AIB nel 2022 col titolo *Pratiche condivise per le bibliografie nazionali nell'età digitale*? Come recuperare la fruttuosa collaborazione fra BNI e *Casalini Libri* degli anni 1968-1984 – dall'annata XI (1968) alla XXVII (1984) della BNI –, interrotta, scrive Diego Maltese, «per un'assurda decisione presa in alto»? Quale rapporto tra BNCF, BNCR, ICCU, BNI e SBN? Come si configura il deposito legale digitale e di conseguenza come si evolve il controllo bibliografico nazionale? Quale rapporto fra i rinnovati concetti di archivio nazionale del libro e di Bibliografia nazionale? Quale apporto ha dato e potrà dare la relazione collaborativa con *Wikidata* sia alla gestione dei dati d'autorità di SBN nell'ottica del loro riutilizzo e arricchimento, sia nella più generale gestione del catalogo (o di come si chiamerà in futuro) tramite il risalto dell'importanza e del ruolo degli identificatori? Come garantire l'interoperabilità dei dati nel web in domini eterogenei? Quali possono essere le conseguenze nel campo della catalogazione e della metadatazione dei nuovi, rapidissimi sviluppi dell'intelligenza artificiale? Possono gli algoritmi di *machine learning* aiutare i creatori di metadati a riconoscere entità uguali, simili o nuove rappresentate nelle registrazioni catalografiche in relazione all'importanza del controllo della qualità dei dati e dell'impegnativo processo di clusterizzazione? Quale consapevolezza vi è della filosofia della metadatazione, ovvero dello spostamento – basato sul modello concettuale IFLA LRM – del focus dalle risorse alle relazioni bibliografiche che le caratterizzano (responsabilità, curatela, traduzione ecc.)? Quale ruolo può avere lo studio avanzato sull'*entity modeling*, approccio che richiede un cambiamento di paradigma nella registrazione dei dati caratterizzato dall'identificazione delle entità, tramite gli identificatori, strumenti tipici del web semantico? Identificare le entità è funzionale all'obiettivo del Controllo bibliografico universale, il progetto più ambizioso mai ideato in ambito biblioteconomico, a cui ciascun Paese è chiamato a partecipare. Come facilitare un nuovo livello di cooperazione sostenibile per la creazione, il mantenimento e la protezione a lungo termine di dati di qualità? Come le biblioteche, la BNI e SBN debbono attrezzarsi difronte alle crescenti "minacce ibride", ovvero agli attacchi che hanno scopi molteplici e agiscono su più livelli di un sistema?

Domande su cui occorre riflettere per proporre soluzioni che riposizionino autorevolmente SBN nel contesto cooperativo internazionale contemporaneo. Il riferimento al ruolo internazionale assunto da agenzie bibliografiche non istituzionali e ad alcune esperienze promosse da biblioteche universitarie sparse sulla penisola indicano con chiarezza la necessità di un'evoluzione di SBN verso una maggiore interoperabilità semantica, basata sull'adozione di standard aperti e vocabolari condivisi.

Il tema è complesso e vasto; *JLIS.it*, pertanto, dichiara la disponibilità a proseguire la discussione ospitando ulteriori interventi sul tema.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-684

JLIS.it

The bibliographic transition and the challenges of cataloguing cooperation in the digital age: some reflections*

Mauro Guerrini[(a)]

a) University of Florence, https://orcid.org/0000-0002-1941-4575

Contact: Mauro Guerrini, mauro.guerrini@unifi.it
Received: 26 July 2025; **Accepted:** 14 October 2025; **First Published:** 15 January 2026

ABSTRACT
The context of cataloging has changed since 1984, the year SBN was launched. Its philosophy risks remaining unknown to young librarians, increasingly burdened by daily work and, in many cases, bureaucratic burdens. The context has also changed with respect to the numerous developments that have occurred over the years. Today, it is comforting to recognize that methodological and technical solutions developed at the turn of the 20th and 21st centuries (FRBR, IFLA LRM; linked data) offer unimaginable potential compared to what was possible at the beginning of SBN, when there was no internet and even a single bit needed to be saved. What can be done, therefore, in light of a renewed conceptual awareness and the reaffirmed need to operate with an eye to the international context and technological innovations. What is the relationship between BNI and Casalini Libri, the bibliographic agencies that collaborated on the BNI's issues XI (1968) to XXVII (1984)?

KEYWORDS
Bibliographic transition; Cooperative Cataloguing; Digital era; SBN.

La transizione bibliografica e le sfide della cooperazione catalografica nell'era digitale: alcune riflessioni

ABSTRACT
Il contesto del lavoro catalografico è cambiato a partire dal 1984, anno del suo inizio. La sua filosofia rischia di rimanere sconosciuta ai giovani bibliotecari, sempre più oberati dal lavoro quotidiano e, in molti casi, dagli oneri burocratici. Il contesto è cambiato anche per quanto riguarda i numerosi sviluppi che si sono verificati nel corso degli anni. Oggi è confortante riconoscere che le soluzioni metodologiche e tecniche sviluppate a cavallo tra il XX e il XXI secolo (FRBR, IFLA LRM, linked data) offrono un potenziale inimmaginabile rispetto a quanto era possibile all'inizio di SBN, quando non c'era Internet e bisognava risparmiare anche un solo bit. Cosa si può fare, quindi, alla luce di una rinnovata consapevolezza concettuale e della ribadita necessità di operare con uno sguardo al contesto internazionale e alle innovazioni tecnologiche? Qual è il rapporto tra BNI e Casalini Libri, le agenzie bibliografiche che hanno collaborato ai numeri XI (1968) e XXVII (1984) della BNI?

PAROLE CHIAVE
Transizione bibliografica; Cooperazione catalografica; Era digitale; SBN.

* Ringrazio Denise Biagiotti, collaboratrice preziosa, Giovanni Bergamin, Franco Neri, Valdo Pasqui e Tiziana Possemato, interlocutori costanti, insieme ad altre amiche e altri amici che hanno letto e commentato versioni intermedie del testo.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-684

Il contesto internazionale

La cooperazione catalografica a livello internazionale e nazionale si caratterizza oggi per svolgersi interamente nel web. La catalogazione si è evoluta in metadatazione e il cambio di paradigma costituisce un salto qualitativo rispetto al passato; il catalogo – termine adesso molto lontano dalla sua etimologia – assume la veste di uno strumento potente e 'scritto' nel linguaggio tipico della tecnologia contemporanea. L'orizzonte descrittivo della risorsa bibliografica si espande in un ventaglio di informazioni *cross domain*. La registrazione o *dataset* si arricchisce e si propone come nuova fonte informativa per altri contesti; essa è predisposta per assicurare il riutilizzo dei metadati e per garantire la loro interoperabilità fra le istituzioni della memoria registrata: archivi, biblioteche e musei. La realtà del web, dove domini differenti s'incontrano, obbliga al concepimento e all'impiego di standard flessibili di fronte alle innovazioni dell'universo bibliografico, mai prevedibili nella loro varietà, e alla molteplicità delle soluzioni tecnologiche. L'apertura dei dati e l'interoperabilità sono concetti e pratiche che caratterizzano molte biblioteche e agenzie bibliografiche dai primi anni del Duemila.
L'organizzazione dei dati bibliografici tramite linked data sta affiancando il *MARC ed è destinata a sostituirlo progressivamente.
Esperienze collaborative come *Wikidata* hanno sempre maggiore importanza rispetto a un nuovo modo di costruire e condividere le informazioni (Bianchini, Bargioni, e Pellizzari di San Girolamo 2025; Revelli 2025).[1]
La proliferazione su larga scala di progetti e iniziative di applicazione di AI prospetta nuovi modelli di assistenza alla metadatazione basati su fonti autorevoli e su strutture dei dati di qualità. Il nuovo paradigma evocato si basa su concetti come entità, identificatori persistenti, ontologie e interoperabilità semantica (Roncaglia 2025).

In Italia

Negli ultimi tre decenni del secolo scorso, contrassegnati da un contesto sociale e culturale molto attivo, la politica bibliotecaria in Italia si è distinta per la sua dimensione dinamica e propositiva. Importante è stata la discussione sulla costituzione di un sistema bibliotecario nazionale (chiamato giustamente *servizio*) che esaltasse la cooperazione, la catalogazione partecipata e la disponibilità delle pubblicazioni su scala nazionale. La relazione *Natura e funzioni della Biblioteca nazionale centrale* di Firenze del 1979 (Maltese 2025, 331-42) testimonia questa esplosione di idee, nonché la capacità progettuale dei bibliotecari di elaborare una visione così aperta e collaborativa da rappresentare tuttora un punto di riferimento nella definizione dei rapporti tra BNCF, BNCR, BNI, ICCU, SBN. Negli anni Settanta, inoltre, le biblioteche di ente locale e di università accrescono la loro presenza nel Paese tanto da creare nuovi equilibri nella comunità bibliotecaria fino ad allora rappresentata principalmente dai bibliotecari statali. I "nuovi" ingressi si distinguono per la loro tendenza innovativa e per il peso sempre più rilevante all'interno dell'AIB da modificarne la fisionomia. L'Associazione svolge l'importante funzione di connessione e confronto (talvolta molto ac-

[1] Una peculiarità nell'ambito italiano è il contributo che i volontari stanno offrendo a SBN relativamente l'authority dei nomi; vedi (Vrandečić, Pintscher, e Krötzsch 2023, 615-24) e (Pellizzari di San Girolamo 2024a, 33-44; 2024b, 147-62).

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-684

ceso) fra i membri più autorevoli di una comunità variegata e sempre più consapevole del proprio ruolo politico oltreché culturale.
Negli anni Ottanta la nascita di SBN rappresenta una rivoluzione che contribuisce enormemente a creare uno stile di lavoro condiviso e a uniformare i comportamenti catalografici, ricondotti, per la parte descrittiva, a ISBD, uno standard largamente accolto a livello internazionale.
Gli anni successivi si caratterizzano per luci e ombre; i problemi della gestione quotidiana prevalgono sulla riflessione e riducono il tempo per progettare nuove configurazioni bibliografiche e tecnologiche, nel confronto internazionale. Ciò avviene paradossalmente proprio nel momento in cui i nuovi paradigmi concettuali aprono opportunità significative e per certi aspetti rivoluzionarie: FRBR, FRAD, FRSAD, ICP, IFLA LRM emanati da IFLA, e RDA e BIBFRAME promossi da altre rilevanti associazioni e comunità bibliotecarie. Questi principi, modelli e standard sono alla base di iniziative di produzione e condivisione di metadati da parte di molte biblioteche statunitensi ed europee, esperienze che dimostrano la potenza dell'interoperabilità e la capacità di far dialogare culture diverse. Tuttavia, il lungo processo di elaborazione dei modelli, degli standard e delle applicazioni evidenzia la complessa e non lineare transizione bibliografica, in cui l'apertura di orizzonti nuovi richiede consapevolezza teorica, contestualizzazione storica e culturale, apertura alla contemporaneità, sguardo internazionale alle varie tradizioni; evidenzia, soprattutto, una nuova filosofia della cooperazione, secondo cui le priorità sono la massima strutturazione, visibilità e condivisione dei metadati descrittivi sul web, la possibilità di navigare tra i metadati delle risorse e di scoprirne di nuove, la capacità di riutilizzare i dati prodotti in domini differenti per ottimizzare i processi catalografici e renderli più veloci e sostenibili (Guerrini 2025, 23).
La transizione verso l'adozione dei linked open data riguarda principalmente la qualità del servizio reso alla comunità bibliotecaria e non va ridotta soltanto a una questione tecnologica (Guerrini e Possemato 2015). Le realizzazioni tecnologiche contemporanee in ambito internazionale sono fonte d'ispirazione per soluzioni concrete da predisporre in Italia. Le esperienze della Library of Congress, della Bibliothèque nationale de France, della British Library, della Deutsche Nationalbibliothek e delle biblioteche nazionali del Nord Europa sono punti di riferimento eccellenti se si vuole ripensare la cooperazione catalografica nel nostro Paese.
In Italia, a cavallo degli ultimi due secoli, importanti sono le traduzioni promosse da Luigi Crocetti: ISBD e AACR2 (entrambe con Rossella Dini o solo a cura di Rossella) e DDC, con la costituzione del gruppo EIDE (Edizione Italiana DEwey) e l'organizzazione di alcuni seminari di studio, fra cui quello indimenticabile con John Comaromi, e di altri sostenuti dall'AIB, fra cui *Dewey: da 20 a 21*, con atti editi nel 2001. Un dibattito fecondo avviene con la riflessione su FRBR, modello concettuale tempestivamente tradotto dall'ICCU nel 2000 – l'Italia è il primo Paese al mondo a tenere un *Seminario FRBR* nel gennaio 2000 – e con l'impegnativa redazione delle nuove regole di catalogazione, motivo di importanti seminari precedenti e paralleli alla loro definizione. Dopo la pubblicazione delle Reicat nel 2009, il confronto istituzionale con la comunità professionale si offusca, mentre prosegue quello scientifico.
Sarebbe stato importante provvedere a monitorare sia l'applicazione del codice in SBN (a cui si è aggiunta nel 2016 *la Guida alla catalogazione*) sia la qualità dei metadati prodotti. La capacità di manutenere i dati e renderli adattabili all'evoluzione degli standard e dei servizi richiesti da una realtà sempre più aperta e fluida è elemento essenziale per rendere sostenibili a lungo termine le politiche che orientano le scelte di metadatazione e i progetti che su tali scelte vivono. Aggiorna-

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-684

menti periodici del sistema vi sono stati e vi sono, ma per lo più basati su aggiustamenti e rivisitazioni parziali, senza mai rivedere l'architettura rigida del sistema. La qualità dei dati è decisiva e si ottiene con un'impostazione metodologica chiara, con la formazione del personale e con l'indispensabile manutenzione, compito che riguarda tutti i partecipanti alla rete, ma che forse necessita di un gruppo di esperti *ad hoc* che operi con autorevolezza, delicatezza e condivisione.
Atteggiamento collaborativo ha caratterizzato la redazione e l'affermarsi del *Nuovo soggettario* della BNCF, la cui redazione ha coinvolto i migliori specialisti italiani e ha tratto vantaggio dal costante confronto internazionale; esso è uno strumento che garantisce uniformità e coerenza alla descrizione semantica.
Sul piano gestionale vi è stata l'apertura di SBN a software applicativi non nativi SBN e, quindi, a una rivisitazione della natura dei poli e dei sistemi interistituzionali di biblioteche, pur in un processo a velocità diverse. La nuova realtà, tuttavia, ha aumentato la produzione di metadati "non puliti" che rendono difficile le migrazioni. Occorre un maggiore investimento in formazione di base e avanzata.
A livello istituzionale ha prevalso un atteggiamento rivolto più al consolidamento interno che al confronto in ambito internazionale e nazionale, caratterizzato da diffidenza verso le sollecitazioni e le esperienze innovative sul piano bibliografico provenienti da Oltreoceano e dal Nord Europa.
Certamente non è facile ricreare quella sinergia tra tecnici e politici peculiare dell'inizio degli anni Ottanta che finanzi oggi una ristrutturazione di SBN, ma potrebbe essere avviato un confronto aperto e, pertanto, costruttivo per creare almeno le condizioni biblioteconomiche verso una reale transizione bibliografica anche in Italia.
Il contesto del lavoro catalografico è cambiato a partire dal 1984, anno d'inizio di SBN, ufficializzato con la firma del protocollo d'intesa tra il Ministero per i beni culturali e l'Università di Firenze, ricondotto alla leadership di Michel Boisset e Angela Vinay; due personaggi le cui straordinarie e originali elaborazioni (insieme a quelle di Luigi Crocetti e Diego Maltese, con Tommaso Giordano e Susanna Peruginelli) rischiano di rimanere sconosciute ai giovani bibliotecari, sempre più oberati dal lavoro quotidiano e, in molti casi, dalle incombenze burocratiche. Il contesto è cambiato anche rispetto ai numerosi sviluppi che si sono succeduti negli anni. Oggi conforta prendere atto che soluzioni metodologiche e tecniche elaborate a cavallo fra XX e XXI secolo (FRBR, IFLA LRM; tecnologia dei linked data) offrono potenzialità impensabili rispetto a quanto consentito all'inizio di SBN, quando non c'era internet e occorreva risparmiare anche un solo bit.
Che cosa si può fare, pertanto, alla luce di una rinnovata consapevolezza concettuale e alla ribadita necessità di operare tenendo conto del contesto internazionale e delle innovazioni tecnologiche?

Due proposte per nuove modalità di cooperazione basate sui linked data

Nel 2023 il Gruppo di lavoro AIB sulle biblioteche digitali pubblica il *Piano d'azione per l'infrastruttura nazionale della conoscenza* (AIB 2023). Il documento riprende l'obiettivo e la progettazione originari di SBN, ovvero la cooperazione fra le biblioteche, realizzata tramite la catalogazione partecipata, nel rispetto di standard internazionali, e tramite il prestito interbibliotecario;[2] obietti-

[2] ILL/DD di SBN è ancora rispondente alle esigenze contemporanee? Vedi: *Sondaggio tra le biblioteche partner*. 226 partecipanti (01/02/2022): https://www.iccu.sbn.it/export/sites/iccu/documenti/2022/Sondaggio_ILL_SBN_22.pdf.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-684

vi resi possibili negli anni Ottanta del secolo scorso grazie alla costituzione di una serie di servizi basata sulle migliori tecnologie allora disponibili. Come accrescere quella visione innovativa adeguandola alla realtà bibliografica e tecnologica contemporanea? Come garantire l'interoperabilità dei dati nel web in domini eterogenei? Come facilitare un nuovo livello di cooperazione sostenibile per la creazione e il mantenimento a lungo termine di dati di qualità? Il *Piano d'azione* ritiene che SBN debba

> presentarsi agli utenti come una confederazione nazionale di cataloghi e di risorse digitali che comprenda anche quelli di biblioteche e sistemi estranei alla rete SBN; il complesso di tali risorse deve risultare trasparente per i motori di ricerca e gli agenti software, a fini di riuso (AIB 2023, 16).

La confederazione di cataloghi e di risorse digitali implica un prodotto bibliografico condiviso e governato da una molteplicità di soggetti che si mettono paritariamente insieme per raggiungere obiettivi comuni. Ciò favorirebbe la convenienza di tutte le biblioteche, con ruoli di responsabilità differenti, a contribuire a SBN. La rete, infatti, non rappresenta l'intero panorama italiano, a cominciare dall'assenza di biblioteche importanti quali quelle della Camera e del Senato della Repubblica. Si tratta di un cambio di paradigma radicale rispetto all'attuale sistema fortemente centralizzato, in cui l'Indice funge da sorta di polo centrale, senza possibilità di dialogo con altri cataloghi.
L'impostazione centralizzata non è negativa, ma occorre essere consapevoli che è opposta rispetto a quella concepita dai fondatori di SBN. In alcuni Paesi vicini, come in Francia, Belgio e Olanda, funzionano uffici centralizzati che si concentrano sulla metadatazione delle risorse per tutte le biblioteche che non hanno e non possono avere personale qualificato per garantire buoni livelli qualitativi della descrizione. Il modello è largamente adottato anche in Scandinavia e, in Italia (ovviamente su scala territoriale minore), dalla rete provinciale di Trento.
Il Gruppo AIB, rifacendosi all'origine di SBN, ripropone una visione di politica catalografica che esalti la partecipazione, con modelli di *governance* basati sul consenso e sul contributo di ogni soggetto, privi di barriere tecnologiche (AIB 2023), verso una governance più informale e orizzontale. Nell'ipotesi della confederazione qual è il ruolo della BNI? Il documento emanato dall'IFLA Bibliography Section, *IFLA common practices for national bibliographies in the electronic age* (IFLA 2022), tradotto e pubblicato in italiano dall'AIB col titolo *Pratiche condivise per le bibliografie nazionali nell'età digitale* (AIB 2023) esordisce:

> La crescita esponenziale dei media digitali nel corso degli ultimi decenni ha messo in discussione molti assunti chiave su cui si basava la realizzazione e la diffusione della bibliografia nazionale (AIB 2023, 11).

La Commissione, al paragrafo 2.4, parla di cooperazione all'interno di uno stesso Paese, nel quale più istituzioni o agenzie possono condividere la responsabilità del deposito legale per diverse tipologie di risorsa:

> La responsabilità del controllo bibliografico nazionale è spesso distribuita tra agenzie incaricate di trattare diversi tipi di materiale, per esempio i testi possono essere di competenza della biblioteca nazionale, mentre i film e le trasmissioni televisive saranno di pertinenza dell'archivio cinematografico nazionale. Le responsabilità all'interno di questi diversi ambiti possono essere centralizzate, ulteriormente

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-684

> delegate o distribuite. In molti paesi numerose biblioteche ricevono copie per deposito legale e queste istituzioni possono anche ripartire fra loro la responsabilità di realizzare la bibliografia nazionale (AIB 2023, 16).[3]

I modelli collaborativi comportano delle sfide; la presa d'atto della difficoltà o dell'impossibilità per una sola istituzione (se non dispone di un numero di catalogatori elevato, come la BnF) a gestire tempestivamente tutte le risorse in autonomia obbliga ad adottare un sistema distribuito.

> Una struttura di tipo collaborativo o distribuito può mobilitare risorse sparse e indirizzarle verso lo scopo comune del controllo bibliografico. La condivisione della responsabilità distribuisce l'onere della gestione del sistema del deposito legale ma, in un sistema distribuito, può essere difficile mantenere coerenza e standardizzazione (AIB 2023, 16).

Si assiste oggi a una molteplicità di soggetti produttori di dati bibliografici, come ricorda anche il preambolo di FRBR. Tuttavia, le biblioteche, più di altre istituzioni, garantiscono un'organizzazione dell'informazione affidabile (Svenonius 2000). Pertanto, nessuna bibliografia nazionale ha il monopolio del controllo bibliografico del proprio territorio; per esempio, la descrizione analitica degli articoli su periodico o in volume miscellaneo è spesso opera di singole biblioteche, istituti scientifici e culturali, in taluni casi di singoli bibliotecari e studiosi.
Il futuro di SBN potrebbe prevedere un partenariato pubblico-privato e cioè interistituzionale. La cooperazione con editori e produttori di media è sempre più necessaria e rilevante, come ricorda *IFLA common practices*:

> Una cooperazione efficiente con gli editori, i produttori di media e i distributori è molto importante per una bibliografia nazionale di successo e sostenibile, per una serie di ragioni: gli editori e i produttori di media sono una fonte primaria di informazioni per la bibliografia nazionale; il buon funzionamento del deposito legale o volontario dipende dalla cooperazione con gli editori; gli editori, insieme ad altri contribuenti, hanno il diritto di aspettarsi che l'agenzia bibliografica nazionale tratti gli oggetti depositati in modo corretto ed efficiente. Gli editori dovrebbero trarre vantaggio da una bibliografia nazionale di qualità che possa conferire una maggiore visibilità alle loro pubblicazioni (AIB 2023, 16).[4]

Il controllo bibliografico nell'era digitale

Il controllo bibliografico è un tema che torna nei congressi IFLA, dimostrazione di una tematica *in progress*.[5] Fra le varie iniziative, la conferenza *The bibliographic control in the digital ecosystem* tenuta in remoto, causa Covid-19, dall'8 al 12 febbraio 2021, ha analizzato i mutamenti che sono avvenuti e che sono ancora in corso nel Controllo bibliografico universale nell'ecosistema digitale

[3] Sul deposito legale, vedi (Bergamin e Messina 2025).
[4] Vedi, inoltre, il paragrafo 2.5 del documento.
[5] Se ne è parlato anche al Satellite Meeting all'Astana IT University of the Republic of Kazakhstan il 14-15 agosto 2025, intitolato: *Artificial Intelligence, Bibliographic Control and Legal Matters: Navigating New Horizons*, promosso dall'IFLA Bibliography and Information Technology Sections e dall'IFLA Artificial Intelligence Special Interest Group.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-684

(Bergamin, Guerrini, e Manzoni 2022, [391]-393). Il controllo bibliografico è radicalmente cambiato negli ultimi decenni perché l'universo bibliografico è radicalmente cambiato: risorse, agenti, tecnologie, standard e pratiche. Nell'ambiente convenzionale il controllo bibliografico si basava principalmente sulle agenzie bibliografiche nazionali e il risultato derivava dall'insieme degli impegni nazionali.[6] Nell'ambiente digitale, invece, il controllo bibliografico è distribuito su una rete di nodi, ciascuno dei quali può condividere, valorizzare e riutilizzare i metadati. L'ecosistema digitale non ha confini nazionali e ogni nodo può contribuire al controllo bibliografico universale. Quali sono i nodi rilevanti nell'ecosistema digitale? Al primo posto vi sono ovviamente gli editori: le risorse pubblicate vengono create insieme ai loro metadati, agli identificatori, per permetterne la reperibilità e l'accesso. L'industria editoriale, ovvero la produzione e la distribuzione su larga scala di prodotti e servizi culturali, fa affidamento proprio sui metadati bibliografici per raggiungere obiettivi specifici. Alcuni editori, in particolare quelli del circuito delle University Press, pubblicano una parte crescente delle risorse seguendo il modello dell'accesso aperto.

Un secondo nodo fondamentale è costituito dalle biblioteche. Da oltre cinquant'anni esse sono consapevoli della "trasformazione digitale",[7] ovvero dell'adozione della tecnologia digitale per trasformare i servizi tradizionali e soprattutto per crearne e promuoverne di nuovi. Già col MARC, nato a metà degli anni Sessanta, i metadati per le risorse convenzionali e per quelle digitali sono in forma digitale. Dai primi anni del Duemila, le biblioteche sono state coinvolte in un percorso di transizione per «raccogliere i vantaggi della tecnologia più recente preservando un solido scambio di dati»[8] garantito dal MARC. Le tecnologie attuali si basano sulla visione del web semantico e sulla tecnologia dei linked data, ma ancora convivono con le precedenti,[9] poiché la transizione non è un guado facile né rettilineo. Il percorso tiene virtualmente conto dei problemi d'integrazione tra diverse tradizioni culturali e linguistiche e le ontologie accolgono e conciliano modi diversi di dire la stessa cosa.

La soluzione, da parte dell'ICCU, di MARC21 (anziché di UNIMARC[10]), utilizzata da circa l'85% delle biblioteche nel mondo, agevolerebbe il passaggio ai linked data e il riuso dei metadati, valutato il forte impegno in tal senso avviato dai primi anni del secolo dalla Library of Congress (LoC) e da altre importanti istituzioni di vari continenti che hanno recepito pienamente le evoluzioni di FRBR, di IFLA LRM e di RDA. Permane, come riferimento concettuale, *On the Record. Report of The Library of Congress Working Group on the Future of Bibliographic Control* del 9 gennaio 2008, il documento che delinea l'evoluzione del controllo bibliografico nell'era digitale, nella considerazione dell'impatto delle tecnologie emergenti e delle variegate esigenze degli utenti.[11] E punto di riferimento operativo, come naturale successore di MARC21, è BIBFRAME; ecco il motivo per cui l'adozione di MARC21 avrebbe facilitato la transizione. L'aggiornamento degli standard e dei software, infatti, è impegno che richiede sia ingenti risorse economiche sia un gruppo di esperti internazionali che studi e si confronti con altri professionisti prima di giungere al testo definitivo

[6] https://www.ifla.org/publications/node/7468.
[7] https://en.wikipedia.org/wiki/Digital_transformation.
[8] https://www.loc.gov/marc/transition/.
[9] https://www.w3.org/standards/semanticweb/data.
[10] Ultimo aggiornamento: 24 gennaio 2025: Just published: New version of UNIMARC/B and UNIMARC/A manuals – IFLA. https://www.ifla.org/news/just-published-new-version-of-unimarc-b-and-unimarc-a-manuals/.
[11] https://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-684

(standard) e a nuove soluzioni tecniche. MARC21 è una piattaforma che, per la granularità di registrazione dei metadati, è un'ottima base su cui costruire architetture del web semantico.
Un altro nodo del controllo bibliografico è rappresentato dai motori di ricerca, utilizzati anche come strumento per trovare i metadati; i creatori di metadati bibliografici utilizzano standard, come *Schema.org*, per migliorare la rappresentazione e la scoperta dei dati bibliografici sul web; in questo senso, i motori di ricerca possono essere annoverati tra le macchine che leggono cataloghi.
Un ultimo nodo (ma altri potrebbero essere citati) è *Wikidata*. Numerose iniziative bibliotecarie utilizzano *Wikidata* o *Wikibase*, basati sulle tecnologie del web semantico e dei linked data. L'IFLA, alla fine del 2019, ha creato un gruppo di lavoro per esplorare

> l'integrazione di *Wikidata* e *Wikibase* con i sistemi bibliotecari e l'allineamento dell'ontologia *Wikidata* con i formati dei metadati delle biblioteche come BIBFRAME, RDA e MARC.[12]

Wikidata può essere interpretato anche come una storia di successo per il riutilizzo e il potenziamento dei metadati prodotti dalle biblioteche nel campo del controllo d'autorità (soprattutto nella raccolta e nella correlazione degli identificatori). *Wikibase* è stata definita dal *Wikilibrary Manifesto* del 2020

> come una promettente infrastruttura tecnica per memorizzare, modificare e scambiare dati.[13]

ICCU collabora costantemente a *Wikidata* con l'arricchimento delle voci d'autorità.
Il controllo bibliografico universale in ambito digitale ha una forte interconnessione con la conservazione digitale nel lungo periodo; esso, infatti, non è di alcuna utilità se le risorse digitali a cui si fa riferimento non sono accessibili. Si tratta di una questione centrale, considerato che oggi una parte sostanziale della ricerca scientifica è pubblicata esclusivamente in formato digitale e si basa su precedenti ricerche pubblicate con le stesse modalità. Le risorse digitali sono costantemente minacciate principalmente dalle conseguenze dell'obsolescenza tecnologica. Se perdiamo le risorse digitali del passato, anche quelle di oggi perderanno il loro valore. L'obiettivo della normativa sul deposito legale è garantire nel lungo periodo

> un accesso universale ed equo alle informazioni registrate sia in forma digitale che in forma tradizionale.[14]

Le biblioteche nazionali hanno, nella maggior parte dei paesi, questo mandato stabilito per legge e il deposito legale è la base per il controllo bibliografico nazionale; esse sono consapevoli che la massa di informazioni da gestire è enorme e in costante crescita; ciò accentua la necessità della selezione e della cooperazione. Si tratta di un punto strategico.
In quanto forma di collaborazione nel controllo bibliografico e per migliorare la tempestività del servizio, importante è il programma CIP (Cataloging in Publication), completamente riformulato rispetto all'idea originaria, tramite il quale gli editori forniscono alle bibliografie nazionali (e ad

[12] https://www.ifla.org/node/92837.
[13] *Wikilibrary Manifesto* (Deutsche Bibliothek e Wikimedia Deutschland), https://www.wikimedia.de/the-wikilibrary-manifesto/.
[14] https://www.ifla.org/publications/guidelines-for-legal-deposit-legislation.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-684

altri soggetti) i metadati di un volume annunciato come di prossima pubblicazione, operazione che la Library of Congress compie da anni (AIB 2023, 27; Saccucci 2021, 11-32).
Il controllo bibliografico universale, dunque, si basa oggi su più nodi intesi in modo integrato; ed è per questo motivo che molte agenzie bibliografiche nazionali stanno ridefinendo la loro funzione. Le attività principali sono le stesse: controllo d'autorità, controllo di qualità, impegno sulla standardizzazione e sulla costruzione e manutenzione di vocabolari e thesauri, tempestività. Queste operazioni vengono svolte oggi direttamente sull'ecosistema digitale e possono avere un impatto nei cataloghi delle biblioteche e in altri domini culturali.
Le risorse culturali (in forma digitale o analogica) sono "oggetti sociali" tramite cui comprendiamo e interagiamo con il mondo.[15] Per questo motivo, il controllo bibliografico ha uno stretto collegamento con il contributo fondamentale offerto dalle biblioteche alla sostenibilità economica, sociale, sanitaria e ambientale della comunità.[16]

BNI e SBN

Nella riflessione degli anni Settanta e Ottanta del secolo scorso un motivo di discussione era la definizione del ruolo della *Bibliografia nazionale italiana* in quanto agenzia locale del controllo bibliografico universale (quindi con responsabilità nazionali e internazionali) e il catalogo SBN. La BNI avrebbe dovuto partecipare a SBN alla stregua di una biblioteca qualsiasi? Quale rapporto fra BNI e ICCU? E, più in generale, quale rapporto tra BNCF, BNCR, ICCU, BNI?
Diego Maltese, in *Servizi bibliotecari nazionali e articolazioni regionali* del 1978, cita l'intervento di Angela Vinay dell'anno precedente tenuto a Pisa al congresso dell'AICA dedicato alla cooperazione in Italia. Maltese commenta e offre la sua interpretazione dei compiti dei vari istituti:

> Il controllo bibliografico nazionale suppone, dunque, da un lato la raccolta e la conservazione delle pubblicazioni nazionali in un archivio, ma dall'altro, non meno essenziale, la documentazione dell'archivio, vale a dire la bibliografia nazionale. […] In questo senso, accanto alla Biblioteca nazionale centrale di Firenze, alla quale spetta la responsabilità di archivio nazionale del libro e di officina della bibliografia nazionale, con le connesse competenze nel campo della normalizzazione della notizia bibliografica e della sorveglianza sulla consegna obbligatoria degli stampati, la Nazionale di Roma ha la capacità e la vocazione per tutti quegli altri servizi, compresa la creazione di strumenti integrativi della bibliografia nazionale per il controllo delle pubblicazioni che ne sono escluse, in collaborazione anche con i sistemi decentrati via via che promoveranno la formazione di cataloghi collettivi regionali. L'Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche è da parte sua responsabile della diffusione della bibliografia nazionale nelle sue varie presentazioni (a stampa, a schede, in forma che può essere letta dalla macchina) e della gestione degli altri archivi di dati di rilevanza nazionale, che insieme vanno a costituire l'oggetto specifico della sua attività: appunto, il catalogo collettivo. (Maltese 2025, 482-85).

[15] "Oggetto sociale" è un concetto sviluppato dal filosofo italiano Maurizio Ferraris.
[16] https://www.ifla.org/libraries-development. È il contenuto della Tesi 12 del *Nuovo Manifesto per le biblioteche digitali* dell'AIB, https://tinyurl.com/ygmg9g5x.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-684

Una lunga citazione sia per ricordare la questione sia per porre di nuovo il problema, naturalmente nei termini attuali. Maltese prospetta percorsi separati tra BNI e SBN, come servizio diverso, seppur collegati. È ancora sostenibile questa visione, nella consapevolezza che la prima è una bibliografia e il secondo è un catalogo? La posizione di Vinay e di Maltese è superata nell'era digitale? Certamente la BNI non è l'Ufficio catalogazione della BNCF e i suoi compiti di agenzia locale del controllo bibliografico universale sono definiti dai documenti emanati dall'IFLA in più occasioni.
Il rapporto fra BNI e SBN è ripreso da Maltese in *Servizio bibliotecario nazionale e servizio locale: la realizzazione di Ferrara*, un intervento pronunciato dieci anni dopo rispetto al primo citato sopra, quando SBN era già una realtà.

> Un sistema «locale», tanto per intenderci, è da considerare anche quello della *Bibliografia nazionale italiana*, nel senso che essa fornisce un servizio specifico per situazioni specifiche, cioè quando si ha bisogno di un'informazione bibliografica autorevole sui prodotti dell'editoria nazionale in contesti competenti. Sostituire sistematicamente nel SBN quanto serve ed è sufficiente alla localizzazione di un documento con la sua registrazione nella *BNI* (che tra l'altro può riguardare solo una parte dei documenti localizzabili, quelli italiani a partire da una certa data) è un errore. [… Pertanto,] per il servizio nazionale, alla registrazione di un documento, introdotta nel sistema da qualsiasi biblioteca periferica con un minimo di dati sufficiente alla sua identificazione e localizzazione, va semplicemente associato, dalla biblioteca che ne ha la responsabilità, l'indirizzo della registrazione corrispondente nell'archivio della *BNI*, quando questa è disponibile (Maltese 2025, 511-13).

L'impostazione ricorda quella del modello francese attuale, in cui la BnF, utilizzando ontologie basate su RDA (RDA-FR) e contando su centinaia di bibliotecari, cataloga per l'intero Paese (compresi i libri che hanno la Francia come soggetto).[17] La BnF ha oggi un dinamismo notevolissimo nella transizione bibliografica, con un riesame costante, complessivo, partecipe e puntuale dei principi, dei modelli e degli standard condotto autorevolmente in sede IFLA.
In Italia, rimane sul tavolo il tema insoluto di enorme portata relativo alla separazione o integrazione fra bibliografia e catalogo, ovvero fra BNI e SBN.
Che il tema non sia affatto superato lo ricorda *Pratiche condivise per le bibliografie nazionali nell'era digitale*, che lo richiama e ne presenta le differenti soluzioni nei vari paesi. Il documento afferma che nel panorama internazionale,

> non c'è unanimità di vedute sul fatto che la bibliografia nazionale debba essere un database distinto oppure inserito nel catalogo della biblioteca nazionale. Soluzioni diverse saranno adatte alle diverse situazioni nazionali. Se la bibliografia nazionale è una parte del catalogo, gli utenti dovrebbero essere in grado di compiere ricerche su quel segmento. Indipendentemente dal modo in cui i dati bibliografici nazionali sono organizzati, si raccomanda che la bibliografia nazionale sia presentata a parte, distinta da qualsiasi catalogo istituzionale o collettivo (AIB 2023, 15).

[17] Sulla transizione bibliografica in Francia, un'esperienza molto interessante, vedi: https://www.transition-bibliographique.fr/rda-fr/.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-684

Il punto 3.4 *Le agenzie bibliografiche nazionali e il ruolo delle bibliografie nazionali*, recita:

> In molti paesi la responsabilità della produzione della bibliografia nazionale spetta alla biblioteca nazionale. Tuttavia, bibliografia nazionale non è sinonimo di catalogo della biblioteca nazionale in quanto ciascuno di essi ha uno scopo differente. In alcuni casi, il catalogo è una registrazione delle raccolte della biblioteca nazionale. […] La bibliografia nazionale documenta l'archivio della produzione editoriale del paese e quindi può incorporare materiale che non si trova nella raccolta della biblioteca nazionale (AIB 2023, 33-4).

È incredibile che la situazione attuale (e ciò è poco noto) derivi da casualità, come ricorda Gloria Cerbai, che conosce molto bene la questione in quanto protagonista del dibattito, da redattrice prima e responsabile poi della BNI. Complessa è la tematica che emerge dalla sua testimonianza:

> Riprendo il discorso dell'altro giorno, ricordo benissimo che Angela Vinay e Diego Maltese avevano convenuto che la BNI sarebbe rientrata in SBN per garantire il controllo bibliografico dei dati immessi da tutte le biblioteche in SBN. Dall'edizione mensile della BNI, alla fine degli interventi di coerenza delle descrizioni fatte in SBN da tutte le biblioteche, si era convenuto che al momento del licenziamento del fascicolo mensile sarebbe stato riversato in SBN il contenuto delle descrizioni BNI del mese. La descrizione BNI garantiva la coerenza del catalogo generale SBN; quindi collaborazione tra BNI e catalogo generale ma solo per le descrizioni comuni a tutte le biblioteche italiane: garanzia di coerenza delle descrizioni in SBN e la libertà BNI di rappresentare le descrizioni BNI di parte del materiale italiano. Perciò, dopo il 1975, veniva assunto il problema della coerenza della descrizione, salvo errori, tra SBN e BNI e dei cataloghi partecipanti al SBN. Al riguardo il rapporto IQ sarebbe stato garantito dal sistema generale dell'SBN dallo studio di un ingegnere dell'IBM. Veniva a Firenze tutte le settimane per un intero giorno, un'ingegnera che avrebbe seguito i consigli, per la BNI. Maltese, che avrebbe lasciato al termine dell'anno la direzione della BNI, assieme a quella della Biblioteca, garantito da Angela Vinay avanzava nelle procedure tecnologiche. Ma non ci fu nessun esito. Dopo il 1979 Angela Vinay probabilmente consigliata da Michel Boisset di riversare studi e risorse su SBN, decise che SBN si sarebbe occupato di tutte le biblioteche italiane e della BNI. Nel 1979 la Vinay, vessata dalle difficoltà finanziarie di SBN, pensò di provvedere alle spese con la partecipazione delle regioni, e quindi liberando il Ministero dalle spese e dagli esborsi della catalogazione italiana. Probabilmente questo è il momento in cui SBN dirotta il problema finanziario sul contributo delle regioni italiane. Ricordo che Michel Boisset, che Fulvia Farfara aveva conosciuto al convegno di Banbury. fu invitato alla riunione internazionale sui problemi della catalogazione in Italia. Le cose, presumo si siano orientate diversamente per i motivi finanziari del Ministero (direttore Sisinni). Susanna Peruginelli avrebbe lasciato la Biblioteca Nazionale. La Biblioteca ormai era schiacciata dai problemi e tutto precipitò in una prevedibile fine. […] Successivamente nulla si mosse e nessuna innovazione sarà accolta (tu lo sai bene!).[18]

[18] Gloria Cerbai, email del 27 giugno 2025.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-684

Cerbai accenna alla questione cruciale dei rapporti fra BNI e SBN anche nel suo intervento in occasione della presentazione del volume di Diego Maltese *Con metodo e rigore* (Maltese 2025) in BNCF:

> Diego Maltese ha costantemente insistito su problemi attuali, ma sempre disattesi: archivio nazionale del libro, *Bibliografia nazionale italiana*, SBN accanto a BNI, entrambi sempre al loro posto. Anche qui, progetti iniziati e poi interrotti, con la promessa di riprenderli nel giusto conto. La realtà attuale è chiara, però: ancora questi temi stanno lì ad aspettare il loro turno, forse ormai abbandonati o, se vivi, davvero in poca salute (Cerbai 2025, 61).

La discussione si è assopita nel tempo, ma «ancora questi temi stanno lì ad aspettare» una definizione chiara. Quali sono esattamente le finalità di ciascuno? La questione è tutta italiana, giacché nella maggior parte dei paesi esiste solo la Biblioteca nazionale – una sola e non due – e non esiste un istituto come l'ICCU. La tematica va ovviamente proposta nei termini di oggi, che sono diversi da quelli di quattro o cinque decadi fa, un arco di tempo che segna due contesti incommensurabili e un'esperienza di cui va tenuto conto. La differenziazione bibliografica, rispetto agli anni Settanta del secolo scorso, oggi è tale da pensare sempre più a uno schema cooperativo che esalti il coordinamento rispetto alla gerarchia di un solo ufficio.

BNI e Casalini Libri: una proficua collaborazione interrotta

Parallelamente alla realtà bibliotecaria istituzionale, dal 1958 opera in Italia Casalini Libri, una delle principali agenzie bibliografiche e fornitore di pubblicazioni europee a istituzioni di tutto il mondo.[19] La BNCF e Casalini Libri hanno collaborato per le annate dalla XI (1968) alla XXVII (1984) della BNI; poi, scrive Maltese,

> la collaborazione italiana al programma di *shared cataloging*, che tanto ci onorava, dovette cessare, per un'assurda decisione presa in alto: la bibliografia nazionale doveva entrare come un servizio qualsiasi di catalogazione in SBN e quindi poteva trattare solo libri "accessionati", cioè che avessero un numero d'ingresso e una collocazione in una biblioteca italiana (Maltese 2025, 580-81).

L'azienda, tramite una lunga e proficua collaborazione con le più importanti biblioteche americane ed europee, ha proseguito a produrre e diffondere una bibliografia nazionale italiana – *I Libri. Bibliografia italiana* – , che è divenuta un punto di riferimento a livello internazionale. Essa presenta mediamente un numero di registrazioni bibliografiche quattro volte maggiore rispetto a quelle prodotte dalla BNI; in realtà ancor più, dato che *I Libri* non descrive le opere straniere tradotte in italiano. Nei decenni successivi alla «assurda decisione» d'interrompere la collaborazione con Casalini ci sono stati alcuni tentativi (fra cui uno promosso dal sottoscritto nel 2013) per favorire il confronto fra *I Libri* e la BNI e, in prospettiva, arrivare a una fusione delle due iniziative.

[19] Si veda l'influenza e l'apporto alla fattiva collaborazione esercitati da Marion Schild (1907-2003), bibliotecaria alla Library of Congress, che lavorò per molti anni a Firenze in collaborazione con la Biblioteca nazionale centrale e con Casalini Libri al controllo bibliografico italiano, catalogando per la biblioteca statunitense; vedi: Maltese (2004, 445-52) ripubblicato in Guerrini (2017, [117]-26) e in Maltese (2025, 576-83).

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-684

Entrambe, infatti, svolgono le medesime funzioni di controllo bibliografico nazionale, pur con finalità diverse; peraltro, hanno sede a pochissimi chilometri di distanza l'una dall'altra, la prima a Firenze e la seconda a Fiesole.

Utilizzare i metadati prodotti da altre agenzie bibliografiche italiane rappresenta un tema storico della BNI, tema che, tuttavia, non ha mai trovato una soluzione. Come, pertanto, potrebbe essere integrato l'apporto di Casalini Libri (e di altre agenzie private) con SBN? Nella visione confederale proposta dal *Piano d'azione per l'infrastruttura nazionale della conoscenza* risulta naturale che BNI impieghi metadati di qualità redatti dai protagonisti della produzione, della distribuzione e del commercio del libro.

Il quadro dei servizi bibliografici italiani va ripensato: essi dovrebbero essere offerti in modo integrato fra pubblico e privato. Una sinergia da studiare a fondo, nel reciproco interesse. Il coinvolgimento di privati c'è stato ed è stato anche significativo, ma come rapporto tra committenti pubblici ed esecutori privati, non come partnership tra pari. Nel luglio 2024, Casalini Libri sottopone alla BNCF una proposta diversa, che non riguarda più la fusione fra *I Libri* e la BNI, bensì l'adesione della *Bibliografia nazionale italiana* all'iniziativa internazionale *National Bibliographies in Linked Open Data*,[20] cioè l'inclusione della BNI nel polo che ospita le bibliografie nazionali prodotte in linked open data – ambiente tecnologico che consente di cercare i metadati di più bibliografie nazionali su un'unica piattaforma –, in aggiunta a un portale dedicato ai soli titoli della BNI.

Casalini Libri, unendo le competenze con @Cult nel 2020, ha accentuato la funzione di centro di ricerca e sviluppo di tecnologie applicate al dominio bibliografico, collaborando con prestigiose biblioteche americane, europee e di altri continenti a progetti nel dominio dei linked open data. Tra le numerose iniziative, rilevante è la messa in produzione nel 2023 – insieme alla British Library – del nuovo portale della *British National Bibliography*. La BNB è divenuta la prima bibliografia ad aderire al *tenant* Share Family dedicato alle bibliografie nazionali.[21] I dati BNB sono stati arricchiti con URI, rimodellati e raggruppati per renderli compatibili con gli standard contemporanei, tra cui IFLA LRM, RDA e BIBFRAME.[22]

Il 26 marzo 2025 Casalini Libri ha stilato un protocollo d'intesa[23] fra RDA Steering Committee (RSC) e Share Family,[24] importante iniziativa di cooperazione internazionale che coinvolge numerose biblioteche americane (fra cui la Library of Congress), nordeuropee (fra cui la British Library e la Nasjonalbiblioteket norvegese) e italiane (fra cui la rete di biblioteche universitarie del Sud Italia, guidate dalla Federico II[25]). L'accordo garantisce un dialogo fra le comunità

[20] *Partecipazione alla Share Family secondo la tecnologia LOD Platform*, https://wiki.share-vde.org/w/images/archive/c/c0/20250212093145%21Partecipazione_alla_Share_Family_2025.pdf.

[21] https://www.share-family.org/. Per gli aggiornamenti, vedi: «Share Family Bulletin», https://wiki.share-vde.org/wiki/ShareFamily:NewsAndUpdates.

[22] La British Library ha migrato la *British National Bibliography* sul nuovo ambiente https://bl.natbib-lod.org.

[23] https://www.rdatoolkit.org/sites/default/files/uploads/RSC_Chair_2025_9_BIBFRAME_protocol_ShareVDE_2025_0.pdf.

[24] https://www.atcult.com/share-family. I diversi ambienti collaborativi sono: share-VDE; *SHARE-Catalogue*, la rete italiana delle biblioteche universitarie; PCC data pool – il Catalogo del Program for Cooperative Cataloging (PCC) in linked data; Bibliografie nazionali in linked data; Parsifal – il portale LOD del consorzio URBE (Unione Romana Biblioteche Ecclesiastiche); LILLIT: portale per i libri illustrati italiani 1501-1800.

[25] https://catalogo.share-cat.unina.it/sharecat/clusters. Vedi (Delle Donne 2025). Purtroppo Share non è compatibile con l'architettura SBN; anche SBN ha i LOD, ma l'interfaccia è al momento poco funzionale.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-684

bibliotecarie culturalmente legate al mondo IFLA LRM (il cui modello dati è concretizzato in RDA) e la comunità tradizionalmente legata all'ambito nordamericano di BIBFRAME. Il progetto evidenzia che ciascun dominio viene salvaguardato proprio con l'apertura dei suoi confini e non con la segregazione dei dati nei propri silos. La grande quantità di dati gestiti e di risorse conservate da biblioteche, archivi e musei che, finora, sono rimasti spesso nascosti in banche proprietarie, possono essere trasformate in linked open data, rendendoli accessibili a un pubblico ampio.

Share Family ha avuto come prima iniziativa l'entrata in produzione nel 2016 di *SHARE Catalogue*, un progetto di cooperazione e di condivisione di servizi tra le biblioteche di università campane, lucane e salentine. Una piattaforma per navigare i cataloghi delle biblioteche aderenti, organizzati secondo BIBFRAME. Il portale prevede l'accesso integrato alle risorse bibliografiche, analogiche e digitali (comprese quelle accessibili online sui siti dei fornitori) delle biblioteche che vi partecipano, con indirizzamento degli utenti ai *full text* dei saggi e delle monografie, nel rispetto delle autorizzazioni e delle licenze d'uso rilasciate dai titolari dei diritti. Il motore di ricerca *SHARE Discovery* attinge ai cataloghi e alle banche dati in rete e permette l'accesso unico ai vari OPAC delle biblioteche. *SHARE Catalogue* è un catalogo integrato in linked open data e, al contempo, un ambiente di sperimentazione concettuale e metodologica.

@Cult, nel 2023, ha implementato un'altra iniziativa appartenente alla Share Family: *Parsifal* di URBE (2024). *Parsifal* è una piattaforma tecnologica concepita nel rispetto di BIBFRAME, esteso per garantire la compatibilità con IFLA LRM, parte integrante delle linee guida RDA, adottate dalla rete bibliotecaria romana. Obiettivo primario della piattaforma è aiutare gli utenti a trovare, identificare, selezionare, ottenere e navigare i metadati sulle opere, i loro creatori (narratori, poeti, illustratori, enti ecc.) e le relazioni che intercorrono fra opere, creatori e soggetti. *Parsifal* consente agli utenti, interni ed esterni a URBE, di determinare la disponibilità di una specifica risorsa bibliografica tra le collezioni delle biblioteche aderenti e raffinare le modalità di ricerca, restituendo risultati arricchiti da fonti provenienti dai singoli cataloghi. La piattaforma applica una visione concettuale e una pratica descrittiva d'avanguardia che richiede, addirittura impone, un'accentuata collaborazione catalografica che segna un salto qualitativo sostanziale per i bibliotecari partecipanti al progetto. *Parsifal* prevede, inoltre, la creazione e la gestione, per la prima volta nella storia della collaborazione tra le biblioteche di URBE, di un *authority file* centralizzato per ora focalizzato sugli agenti e le loro opere; la sua progettazione consente ai catalogatori della rete di impiegare uno strumento partecipativo e di creare una registrazione condivisa da utilizzare, prima di tutto, per disambiguare le entità di tipo *agente* e di tipo *opera*, nonché di supportare l'attività catalografica delle diverse biblioteche.

Un nuovo tema di riflessione: l'*entity modeling*

Le idee nuove, nuove anche per il contesto internazionale, non mancano in Italia. Il concetto di *entity modeling* (Possemato 2024), elaborato da Tiziana Possemato, per esempio, è un tema di riflessione quanto mai stimolante. Esso rappresenta un nuovo approccio alla catalogazione, fondato sull'identificazione e descrizione dell'universo bibliografico secondo il modello entità-rel-

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-684

azioni, quindi un modello coerente con SBN. L'*entity modeling*, partendo da FRBR e da IFLA LRM, mira a strutturare il dominio della metadatazione tramite le entità (le unità fondamentali che popolano il mondo bibliografico, come opere, persone, luoghi ecc.), gli attributi che le qualificano e le relazioni che intercorrono tra di esse. Il metodo consente di descrivere e collegare in modo efficace le risorse, creando un'infrastruttura di dati aperta, riusabile e interoperabile, adatta al web semantico. L'*entity modeling*, pertanto, mira a superare la visione strettamente focalizzata sulla registrazione bibliografica come fotografia descrittiva della risorsa, spostandosi verso una rappresentazione che considera i dati della risorsa come entità indipendenti, identificabili e interrelate. È un approccio innovativo specifico per l'era del web semantico, ambiente in cui le relazioni fra entità e la loro rappresentazione interoperabile diventano cruciali per costruire ecosistemi di dati complessi e connessi. L'*entity modeling*, secondo l'autrice, costituisce la terza generazione della catalogazione, dopo la catalogazione e la metadatazione; il concetto accentua il passaggio dalla descrizione piatta, lineare, delle risorse alla loro modellazione come entità autonome e interconnesse, in un sistema in cui esse sono facilmente identificabili grazie all'uso di identificatori univoci e di relazioni semantiche. Questo approccio risponde alle esigenze di una realtà bibliografica sempre più complessa e digitalizzata, che esige la trasformazione dei cataloghi in strumenti più potenti, interoperabili e aperti all'innovazione tecnologica (Guerrini e Possemato 2025, 69-73).

Conclusione

Il modello di Indice "leggero" ipotizzato dalle madri e dai padri di SBN dovrebbe assorbire le evoluzioni bibliografiche e tecnologiche internazionali, nonché le esperienze catalografiche più innovative condotte sulla penisola italiana. Esse corroborano la necessità di disporre di un'infrastruttura che parli il linguaggio del web semantico e che sia realmente adeguata agli standard condivisi attualmente a livello internazionale. La rinuncia a logiche verticali e autoreferenziali è decisiva per favorire la sinergia con la comunità bibliotecaria globale e per offrire un servizio adeguato ai tempi. Questa è la direzione auspicata dal movimento Open Science/Open Data, che supporta i principi cardine dell'apertura dei dati, fra cui il loro libero e gratuito utilizzo, la trasparenza e il riuso. Il potere informativo dei dati sta proprio nella loro struttura aperta, comprensibile e condivisibile universalmente.

La transizione bibliografica in corso, il cui esito non è del tutto chiaro né scontato, implica l'adozione di logiche che esaltino i progressi tecnologici a supporto degli standard di metadatazione per garantire un accesso equo e inclusivo a informazioni e a dati realmente aperti.

SBN necessita di una nuova configurazione dopo la sua straordinaria concezione biblioteconomica negli anni Ottanta, il sistema di cooperazione catalografica che prefigurò, per primo al mondo, il modello di dati entità-relazioni che aveva le basi teoriche in Ákos Domanovszky (1975) e che sarà poi alla base di FRBR (Bergamin 2020, [82]-[102]; Leombroni 2023, 181-203). È ormai necessaria un'evoluzione profonda che faccia tesoro delle esperienze della comunità internazionale. I quesiti, le sfide e le problematiche da risolvere per garantire un dialogo che superi i confini nazionali, linguistici, culturali e di dominio, infatti, non possono essere risolti in solitaria. Evolvere in questa

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-684

direzione significa disporre di un nuovo paradigma teorico e tecnologico che valorizzi il lavoro svolto e dia forza a quello futuro. Qualora il processo di transizione non avvenisse il rischio della marginalizzazione internazionale di SBN diverrebbe realtà.
In Italia disponiamo di tutti gli strumenti e di tutte le competenze per avviare un percorso innovativo. La comunità italiana può raggiungere il traguardo di una nuova dimensione della cooperazione, offrendo ai giovani bibliotecari una prospettiva esaltante e coinvolgente come e ancor più di quella vissuta dalle madri e dai padri fondatori di SBN.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-684

Riferimenti bibliografici*

AIB (Associazione italiana biblioteche) Commissione nazionale biblioteche e servizi nazionali. 2023. *Pratiche condivise per le bibliografie nazionali nell'era digitale*, a cura di Rebecca L. Lubas, Mathilde Koskas, Pat Riva, Mauro Guerrini, Eva-Maria Häusner, Kazue Murakami, Anke Meyer-Heß, Miriam Nauri, Ylva Sommerland, Monika Szunejko, e Miyuki Tsuda, approvato da the IFLA Professional Council. L'Aja: IFLA. https://repository.ifla.org/rest/api/core/bitstreams/530ddb0b-80ba-42b0-952f-a285d6b02bfc/content.

AIB (Associazione italiana biblioteche) Gruppo di lavoro sulle biblioteche digitali. 2023. *Piano d'azione per l'infrastruttura nazionale della conoscenza.* Roma: Associazione italiana biblioteche.

Bergamin, Giovanni. 2020. "Il sistema di collegamenti bibliografici nell'archivio della cooperazione di T. Giordano, con la collaborazione di S. Peruginelli." *JLIS.it* 11 (2): [82]-[102].

Bergamin, Giovanni, Mauro Guerrini, e Laura Manzoni, *Closing remarks*, in: *The bibliographic control in the digital ecosystem*, edited by Giovanni Bergamin and Mauro Guerrini, with the assistance of Carlotta Alpigiano. Roma: Associazione italiana biblioteche; Macerata: Edizioni Università di Macerata; Firenze: Firenze University Press, 2022, p. [391]-393.

Bergamin, Giovanni, e Maurizio Messina. 2025. "Deposito legale digitale: cooperazione, conservazione, e nuove modalità di fruizione." *Jlis.it* 17 (1): 38-60.

Bianchini, Carlo, Stefano Bargioni, e Camillo Carlo Pellizzari di San Girolamo. 2025. "Wikidata e SBN. Un bilancio di due anni di lavoro (2023-2025)." *Jlis.it* 17 (1): 73-96.

Cerbai, Gloria. 2025. "Appunti sull'importanza del volume *Con metodo e con rigore* di Diego Maltese." *Biblioteche oggi*, XLIII (3): 60-1.

Delle Donne, Roberto. 2025. "SHARE Catalogue: cooperazione e innovazione semantica in un ecosistema bibliotecario condiviso." *Jlis.it* 17 (1): 141-153.

Domanovszky, Ákos. 1975. *Functions and objects of author and title cataloguing: a contribution to cataloguing theory,* edited by Anthony Thompson. München: Verlag Dokumentation. Traduzione italiana: *Funzioni e oggetti della catalogazione per autore e titolo: un contributo alla teoria della catalogazione*, a cura di Mauro Guerrini, traduzione di Barbara Patui, Carlo Bianchini e Pino Buizza. Udine: Forum, 2001.

Guerrini, Mauro. 2025 "La rappresentazione dell'universo bibliografico e delle collezioni in era digitale: l'entity modeling." *Biblioteche oggi trends* 11 (1): 23-7.

Guerrini, Mauro, e Tiziana Possemato. 2015. *Linked data per biblioteche, archivi e musei. Perché l'informazione sia del web e non solo nel web.* Milano: Editrice Bibliografica.

Guerrini, Mauro, e Tiziana Possemato. 2025. "Entity modeling: conversando con Tiziana Possemato sulla terza generazione della catalogazione." *Biblioteche oggi* 44 (2): 69-73. https://www.doi.org/10.3302/0392-8586-202502-69-1.

* Ultima data di consultazione dei siti web: 19 luglio 2025.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-684

IFLA (International Federation of Library Associations and Institutions) Bibliography Section Standing Committee. 2022. *Common practices for national bibliographies in the digital age*, a cura di Rebecca L. Lubas, Mathilde Koskas, Pat Riva, Mauro Guerrini, Eva-Maria Häusner, Kazue Murakami, Anke Meyer-Heß, Miriam Nauri, Ylva Sommerland, Monika Szunejko, e Miyuki Tsuda, approvato da the IFLA Professional Council. L'Aja: IFLA. https://www.ifla.org/wp-content/uploads/Common_Practices_for_national_bibliographies_2021-01.pdf.

Leombroni, Claudio. 2023. "«Come pezzi di un meccano»: alle origini del catalogo del SBN." In *Guardando oltre i confini: partire dalla tradizione per costruire il futuro delle biblioteche. Studi e testimonianze per i 70 anni di Mauro Guerrini*, a cura di Giovanni Bergamin e Tiziana Possemato, 181-203. Roma: Associazione italiana biblioteche.

Maltese, Diego. 2025. "Natura e funzioni della Biblioteca nazionale centrale." *Il ponte*. In *Con metodo e con rigore: scritti di bibliografia e biblioteconomia*. A cura di Mauro Guerrini, con l'assistenza di Patrizia Calò e con la collaborazione di Pino Buizza, 331-42. Roma: Associazione italiana biblioteche.

Maltese, Diego. 2025. "Gli anni di Firenze di Marion Schild." In *Con metodo e con rigore: scritti di bibliografia e biblioteconomia*. A cura di Mauro Guerrini, con l'assistenza di Patrizia Calò e con la collaborazione di Pino Buizza, 576-83. Roma: Associazione italiana biblioteche.

Maltese, Diego. 2025. *Con metodo e con rigore: scritti di bibliografia e biblioteconomia*. A cura di Mauro Guerrini, con l'assistenza di Patrizia Calò e con la collaborazione di Pino Buizza. Roma: Associazione italiana biblioteche.

Pellizzari di San Girolamo, Camillo Carlo. 2024a. "Riconciliare le voci di autorità in SBN con Wikidata. Progressi e prospettive dopo un decennio di lavoro (2013 2023)." *JLIS.it* 15 (1): 33-44. https://www.jlis.it/index.php/jlis/article/view/573.

Pellizzari di San Girolamo, Camillo Carlo. 2024b. "Storia della collaborazione tra Wikidata e le biblioteche della Rete URBE nel controllo di autorità." In URBE (Unione Romana Biblioteche Ecclesiastiche). 2024. *Parsifal: un modello di collaborazione bibliotecaria per condividere la conoscenza registrata*. A cura di Silvano Danieli, introduzione di Mauro Guerrini, 147-62. Firenze: Firenze University Press. https://books.fupress.com/catalogue/parsifal/14537.

Possemato, Tiziana. 2024. *Entity modeling: la terza generazione della catalogazione*. Premessa di Philip E. Schreur, prefazioni di Carlo Bianchini e Maurizio Vivarelli, introduzione di Mauro Guerrini. Firenze: Firenze University Press, https://books.fupress.it/catalogue/entity-modeling-la-terza-generazione-della-catalogazione/14374.

Roncaglia, Gino. 2025. "Catalogazione, metadati e IA generativa: Prime esperienze e prospettive future." *Jlis.it* 17 (1): 106-127.

Saccucci, Caroline. 2021. "Taking the Library of Congress CIP Program into the Future with Pre-Pub Book Link." *JLIS.it*, 12 (3): 11-32. https://www.jlis.it/index.php/jlis/article/view/371.

Svenonius, Elaine. 2000. *The intellectual foundation of Information organization*. Boston: MIT.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-684

URBE (Unione Romana Biblioteche Ecclesiastiche). 2024. *Parsifal: un modello di collaborazione bibliotecaria per condividere la conoscenza registrata*. A cura di Silvano Danieli, introduzione di Mauro Guerrini. Firenze: Firenze University Press.

Vrandečić, Denny, Lydia Pintscher, e Markus Krötzsch. 2023. *Wikidata: the making of.* In: *Companion proceedings of the ACM Web Conference 2023, Austin Texas (USA), April 30-May 4*, 615–24. https://doi.org/10.1145/3543873.3585579.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-685

JLIS.it

Cybersecurity of cultural heritage: a reflection on SBN and Magazzini Digitali as critical infrastructures of Italian libraries

Chiara Storti[(a)]

a) National Central Library of Florence, https://orcid.org/0009-0005-1767-3425

Contact: Chiara Storti, chiara.storti@cultura.gov.it
Received: 28 July 2025; **Accepted:** 14 October 2025; **First Published:** 15 January 2026

ABSTRACT
The contribution analyzes the escalating cyber risks for cultural heritage, driven by digitization and conflicts, highlighting severe incidents (e.g., British Library) causing economic and reputation damages, and data loss. It proposes considering these infrastructures as "critical" through the "C3E - *Cultural Cyber Critical Ecosystem*" model. Legal, managerial, and technological criticalities of SBN and Magazzini Digitali, as core components of this ecosystem, are then examined. The conclusion emphasizes the need to invest in cyber resilience and foster a cultural shift in risk management.

KEYWORDS
Cybersecurity implications; Cyberattack; Ransomware; SBN; Magazzini Digitali.

Cybersicurezza del patrimonio culturale: una riflessione su SBN e Magazzini Digitali come infrastrutture critiche delle biblioteche italiane

ABSTRACT
Il contributo analizza il crescente rischio cyber per il patrimonio culturale, determinato da fattori quali la forte accelerazione post-pandemica della digitalizzazione di processi e servizi e i conflitti in corso in Europa Orientale e nel Medio Oriente. A partire dal noto incidente occorso alla British Library, viene posto l'accento sui possibili danni economici e reputazionali, oltre che su conseguenze come la perdita irreversibile di dati. Alla luce del contesto delineato, si propone di considerare le infrastrutture del patrimonio culturale italiano, con particolare riferimento a quelle afferenti al dominio delle biblioteche, come "critiche", analizzandole in relazione al modello recentemente definito di "C3E - *Cultural Cyber Critical Ecosystem*". Vengono quindi esaminate le criticità legali, gestionali e tecnologiche di SBN e Magazzini Digitali, come componenti fondamentali dell'ecosistema. La conclusione sottolinea la necessità di investire nella resilienza informatica e di promuovere un cambiamento culturale nella gestione dei rischi.

PAROLE CHIAVE
Cybersicurezza; Cyberattacchi; Ransomware; SBN; Magazzini Digitali.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-685

JLIS.it

Perché dobbiamo occuparci di cybersicurezza delle infrastrutture del patrimonio culturale

> Se sino a qualche anno fa il problema sembrava limitato alle sole "applicazioni informatiche", in questi anni non esiste settore, dai sistemi industriali agli apparati medicali, che non sia stato intaccato da qualche forma di attacco informatico. [...]
> In questo contesto, lo strumento di cui maggiormente disponiamo per fronteggiare il rischio informatico è la consapevolezza; è infatti decisamente diverso l'impatto di eventi pianificati rispetto a quello di eventi inattesi.

Queste parole, che risuonano ancora di straordinaria attualità, si possono leggere nell'introduzione al primo Rapporto Clusit sulla sicurezza ICT in Italia (Clusit 2012) che, nel 2012, conteggiava in 470 gli attacchi cyber avvenuti a livello globale l'anno precedente. In (Clusit 2025) il dato è considerevolmente più alto: 3.541 attacchi, di cui ben 357 ai danni del nostro Paese[1]. Questa crescita molto consistente, concentrata negli anni 2020-2024, è dovuta in particolar modo a due opposte tendenze che si sono venute a delineare. Da una parte, la pandemia ha funto da acceleratore per la digitalizzazione dei processi e dei servizi a tutti i livelli, fenomeno che sta già subendo un'ulteriore crescita esponenziale a causa della diffusione degli strumenti di IA; dall'altra, i conflitti in corso, specialmente in Europa Orientale e nel Medio Oriente, hanno alzato il livello generale di tensione internazionale, e anche laddove tali conflitti non si trasformino in vere e proprie *cyberwar*, le infrastrutture informatiche sono obiettivi sensibili, alla stregua di altre infrastrutture strategiche. Non fanno eccezione nemmeno le infrastrutture del patrimonio culturale, pubbliche o private, come dimostrano i tanti incidenti registrati negli ultimi anni, di cui l'attacco ransomware[2] alla British Library dell'ottobre 2023 (The British Library 2024) è solo il più celebre.

Data	Istituzione	Tecnica di attacco	Conseguenze principali	Danni economici accertati
2018	Rijksmuseum Twenthe di Enschede (Paesi Bassi)	Man in the middle (Mitm)[3]	Il museo ha versato 2.4 milioni di sterline per l'acquisto di un quadro ad un conto bancario falso creato da un hacker.	2.4 milioni £

[1] La grandezza di questo dato, come evidenziato nel Rapporto, è dovuta sia alla mancanza o forte carenza di cultura digitale e di cultura della cybersicurezza, soprattutto nel comparto pubblico, sia dalla diffusa obsolescenza dei sistemi hardware e software in uso.

[2] Per una definizione di Ransomware si veda https://www.acn.gov.it/portale/csirt-italia/glossario.

[3] Per una definizione di Mitm si veda https://www.acn.gov.it/portale/csirt-italia/glossario.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-685

Data	Istituzione	Tecnica di attacco	Conseguenze principali	Danni economici accertati
Maggio 2019	Asian Art Museum di San Francisco (USA)[4]	Ransomware	Prolungata indisponibilità dei sistemi informatici del museo.	Costi di ripristino dei sistemi.
Dicembre 2021	Museum of Gloucester (Regno Unito)[5]	Malware[6]	Prolungata inaccessibilità del database contenente i dati relativi alle collezioni museali.	1 milione £ (la cifra complessiva spesa per il ripristino di tutti i servizi compromessi della Municipalità di Gloucester).
Novembre 2022	Archives départementales des Alpes-Maritimes (Francia)[7]	Ransomware	Prolungata inaccessibilità degli inventari online e dei servizi digitali degli archivi. Circa 292 GB di dati rubati.	Costi di implementazione di un nuovo portale.
Dicembre 2022	Metropolitan Opera di New York (USA)[8]	Non nota	Prolungata indisponibilità di tutti i sistemi di rete, compresi il sito web, la biglietteria, e il call center.	200.000 $ di biglietti invenduti a causa dell'indisponibilità del sistema di biglietteria per quasi 30 ore.
Agosto 2023	MiBAC (Italia)[9]	Ransomware	Violazione dei dati.	Non noti
Ottobre 2023	British Library (Regno Unito) (The British Library 2024)	Ransomware	600 GB di dati rubati; pubblicazione online di centinaia di migliaia di file sottratti, compresi i dati degli utenti e del personale. Prolungata indisponibilità di tutti i servizi e rallentamento di tutte le attività della biblioteca.	Tra i 6 e i 7 milioni £ per ricostruire la maggior parte dei servizi digitali (Uddin e Thomas 2024).

[4] https://abc7news.com/post/exclusive-cyberattack-launched-on-sf-museum-computers-/5399932/.
[5] https://www.bbc.com/news/uk-england-gloucestershire-64917275.
[6] Per una definizione di Malware si veda https://www.acn.gov.it/portale/csirt-italia/glossario.
[7] https://www.rfgenealogie.com/infos/alpes-maritimes-un-site-refondu-en-totalite-apres-la-cyberattaque.
[8] https://www.informationweek.com/cyber-resilience/the-metropolitan-opera-cyberattack-highlights-vulnerability-of-cultural-institutions.
[9] https://www.redhotcyber.com/post/attacco-informatico-al-ministero-dei-beni-culturali-italiani-rivendicato-da-lockbit-tra-9gg-la-pubblicazione-dei-dati/.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-685

Data	Istituzione	Tecnica di attacco	Conseguenze principali	Danni economici accertati
Ottobre 2023	Toronto Public Library (Canada)	Ransomware	Violazione dei dati personali di circa 8.000 tra dipendenti ed ex dipendenti, di circa 1.874 dei loro beneficiari e familiari a carico e di 4.100 tra utenti, donatori, appaltatori, volontari ecc.[10] e prolungata indisponibilità dei servizi[11].	Costi di ripristino di tutti i sistemi hardware e software, sostenuti dalla Municipalità di Toronto.
Ottobre 2023	Museum für Naturkunde Berlin (Germania)[12]	Non nota, probabile ransomware	Data breach di dati personali di oltre 37.000 visitatori del Museo	Non noti
Dicembre 2023	Gallery Systems (USA)[13]	Ransomware	Disservizi eMuseum.	Non noti
Maggio 2024	Seattle Public Library (USA)[14]	Ransomware	Violazione dei dati di 27.000 persone e prolungata indisponibilità dei servizi.	1 milione $ per coprire i costi dell'indagine forense, del ripristino dei sistemi e della consulenza legale per la gestione del data breach.
Maggio 2024	Internet Archive (USA)[15]	DDoS[16] e accessi non autorizzati	Prolungata indisponibilità dei servizi.	Non noti
Maggio 2024	MiC - Direzione generale Educazione ricerca e istituti culturali (Italia)[17]	Non nota	Compromissione di dati personali.	Non noti
Agosto 2024	Réunion des Musées Nationaux (Francia)[18]	Ransomware	Rete dei pagamenti dei bookshop del sistema museale compromessa.	Non noti

[10] https://www.ipc.on.ca/en/cases-of-note/toronto-public-library-cyberattack.
[11] https://www.cbc.ca/news/canada/toronto/toronto-public-library-ransomware-employee-data-1.7028982.
[12] https://www.museumfuernaturkunde.berlin/en/information-about-cyber-attack-museum-fur-naturkunde-berlin.
[13] https://apollo-magazine.com/art-news-cyber-attack-museums-ai/.
[14] https://www.seattletimes.com/seattle-news/seattle-library-ransomware-attack-affected-nearly-27k-people/.
[15] https://blog.archive.org/2024/05/28/internet-archive-and-the-wayback-machine-under-ddos-cyber-attack/.
[16] Per una definizione di Attacco DOS e DDOS si veda https://www.acn.gov.it/portale/csirt-italia/glossario.
[17] https://dgeric.cultura.gov.it/violazione-dei-sistemi-informatici-della-direzione-generale-educazione-ricerca-e-istituti-culturali/.
[18] https://www.lesechos.fr/industrie-services/services-conseils/la-cyberattaque-contre-les-musees-francais-na-touche-que-36-boutiques-2112529.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-685

Data	Istituzione	Tecnica di attacco	Conseguenze principali	Danni economici accertati
Ottobre 2024	Internet Archive (USA)[19]	DDoS e accessi non autorizzati	Violazione dei dati di 31 milioni di utenti e prolungata indisponibilità dei servizi.	Non noti
Novembre 2024	Library of Congress (USA)[20]	E-mail breach	Account mail compromessi.	Non noti
Dicembre 2024	National Museum of Royal Navy (Regno Unito)[21]	Ransomware	Prolungata indisponibilità dei servizi del museo.	Non noti
Gennaio 2025	British Museum (Regno Unito)[22]	Sabotaggio interno dei sistemi IT	Parziale chiusura dei servizi museali.	Non noti

Tabella 1. Prospetto dei principali attacchi cyber noti a istituzioni culturali pubbliche e private, nel mondo, a partire dal 2018.

Rispetto alle tipologie di attacco, possiamo rilevare come la maggior parte sia perpetrato con l'uso di ransomware o malware di altra natura, seguono gli attacchi DDoS e quelli causati da vulnerabilità[23] sia conosciute che sconosciute dei sistemi (i cosiddetti *zero-day attack*[24]), spesso in combinazione con tecniche di Phishing[25] e Social Engineering[26], che possono servire non solo a sottrarre dati alla vittima, ma anche ad introdurre malware nei sistemi attaccati[27]. Per quanto riguarda le motivazioni degli attacchi, queste possono essere riconducibili ad alcune macrocategorie:

[19] https://www.washingtonpost.com/nation/2024/10/18/internet-archive-hack-wayback/.
[20] https://securityaffairs.com/171138/data-breach/library-of-congress-email-communications-hacked.html.
[21] https://www.nmrn.org.uk/news/national-museum-royal-navy-statement.
[22] https://www.theguardian.com/culture/2025/jan/24/british-museum-forced-to-partly-close-after-alleged-it-attack-by-former-employee.
[23] Gli attacchi dovuti a vulnerabilità, soprattutto in Italia, non sono da sottovalutare. Nel 2020, ad esempio, il collettivo di hacktivisti *Anonymous* ha colpito diversi siti della PA italiana, tra cui quello di AgID, sfruttando una falla nel CMS sviluppato dall'azienda abruzzese ISWEB S.p.A.: https://www.dire.it/17-06-2020/474702-anorc-attacco-hacker-al-sito-agid-non-e-una-burla-servono-competenze/.
[24] Per una definizione di Attacco Zero Day si veda https://www.acn.gov.it/portale/csirt-italia/glossario.
[25] Per una definizione di Phishing si veda https://www.acn.gov.it/portale/csirt-italia/glossario.
[26] Nel contesto della sicurezza informatica, la Social Engineering o Ingegneria Sociale è l'uso dell'influenza psicologica sulle persone per indurle a compiere azioni o divulgare informazioni riservate. Per una definizione di Ingegneria Sociale si veda anche https://www.acn.gov.it/portale/csirt-italia/glossario.
[27] Il ransomware che ha colpito la British Library è stato introdotto, con grande probabilità, a seguito della compromissione delle credenziali di account con privilegi da amministratore, tramite attacchi di Phishing, Spear-Phishing (si veda https://www.acn.gov.it/portale/csirt-italia/glossario) o con un attacco di forza bruta (*brute force*), cioè con tentativi ripetuti di indovinare le password, facilitati da sistemi di autenticazioni poco solidi, vale a dire basati sull'utilizzo di password non forti o privi di Autenticazione Multi-Fattore (si veda https://www.acn.gov.it/portale/csirt-italia/glossario). Quest'ultima, sebbene, fosse stata introdotta nel 2020 per le applicazioni cloud adottate dalla Biblioteca, era stata esclusa per la connettività al dominio della British Library, inclusi l'accesso ai computer e ai server on-premise, per ragioni di praticità, costo e impatto sui programmi in corso (The British Library 2024).

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-685

- Motivazioni economiche (*cybercrime*): i cybercriminali possono ottenere guadagni diretti con strumenti quali i ransomware, anche se le istituzioni culturali, soprattutto pubbliche, tendono a non pagare i riscatti (o forse non ammettono pubblicamente di farlo). Guadagni indiretti possono, invece, derivare dai dati rubati sia utilizzando, ad esempio, le informazioni bancarie di utenti e donatori, sia rivendendo dati e informazioni a terzi, che potrebbero anche essere i reali committenti dell'azione criminale[28].
- Violazione della sicurezza informatica allo scopo di violare la sicurezza fisica: la compromissione dei sistemi di allarme, dei sistemi di monitoraggio ambientale o dei dati relativi alla dislocazione fisica di opere e documenti preziosi, nelle sale o nei magazzini di biblioteche, musei e archivi, possono dar luogo a danneggiamenti o furti di opere.
- Volontà di utilizzare una infrastruttura del patrimonio culturale come *backdoor*[29] *per attaccare infrastrutture collegate.*
- Volontà di aumentare il livello di tensione generale, come in caso di guerre e conflitti: l'obiettivo di questo di attacchi è, solitamente, quello di distogliere risorse destinate ad infrastrutture di altra natura e veri bersagli dell'azione criminale; le istituzioni culturali possono, quindi, essere vittime collaterali, come durante attacchi DDoS molto vasti.
- Volontà di colpire obiettivi simbolici: le infrastrutture del patrimonio culturale sono ormai anche obiettivi militari a tutti gli effetti nelle cosiddette *cyberwarfare*, come è stato evidente in particolare a partire dall'avvio del Conflitto in Ucraina nel febbraio 2022 (Neglia et al. 2024).
- Ricerca di visibilità: colpire istituzioni molto note, come nel caso della British Library, potrebbe far guadagnare notorietà e credibilità agli autori, singoli o gruppi, degli attacchi. È il caso dei *cyberactivists*, vale a dire di quei singoli o gruppi che utilizzano le tecnologie per promuovere cambiamenti a livello, politico, sociale e culturale.

Dai casi riportati nella Tabella 1 è possibile anche farsi un'idea delle principali conseguenze di un cyberattacco:

- Danni economici. In questa categoria, oltre alle perdite dirette causate dall'eventuale pagamento di un riscatto, rientrano anche le perdite indirette sia per mancati introiti (per tutti i luoghi ad accesso a pagamento, come musei, parchi archeologici o mostre temporanee), che derivanti dalla perdita di dati o informazioni, ovvero degli investimenti fatti nel corso del tempo. A ciò ci possono aggiungere i costi spesso molto alti per il ripristino dei sistemi e dei servizi.
- Perdita o fuga di dati. Considerata la specificità degli asset del patrimonio culturale, la perdita o fuga di dati può causare danni irreversibili. Si pensi, infatti, ai dati relativi al deposito legale digitale che potrebbero non essere più disponibili sul mercato editoriale o, semplicemente online, nel momento in cui i servizi vengono ripristinati[30].

[28] Nel 2024 i pagamenti provenienti da modelli RaaS - Ransomware as a Service hanno rappresentato oltre il 60% delle campagne ransomware: https://www.onlinehashcrack.com/guides/cybersecurity-trends/ransomware-as-a-service-2025-market-analysis.php.

[29] Per una definizione di *Backdoor* si veda https://www.acn.gov.it/portale/csirt-italia/glossario.

[30] Nel report pubblicato da Internet Archive, ad ottobre 2024, a seguito di due importanti attacchi DDoS che hanno reso

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-685

- Perdita reputazionale. L'interruzione di servizi per un tempo prolungato, la perdita o la fuga di dati, personali e non, sono un enorme danno di immagine per le istituzioni culturali, specialmente pubbliche.

Se gli attacchi al patrimonio culturale sono da considerarsi di estrema gravità non solo per il valore in sé dei beni che lo compongono ma, soprattutto, per quello che esso rappresenta per la società (Ovenden 2021), è necessario iniziare a ripensare le infrastrutture digitali del patrimonio culturale come infrastrutture critiche,[31] definendo il perimetro[32] da attenzionare. Negli ultimi anni, infatti, alle più tradizionali criticità legate alla sicurezza dei servizi informatici e di rete[33] e alla digital preservation[34], si sono affiancati i rischi derivanti da incidenti[35] che possono compromettere la cybersicurezza[36]. A questo scopo, si propone una rilettura dell'ecosistema delle biblioteche italiane alla luce del modello "C3E – *Cultural Cyber Critical Ecosystem*", recentemente definito in (Bellini 2025).

indisponibili, per alcuni giorni, non solo i servizi agli utenti, come la Wayback Machine, ma anche i servizi di archiviazione web veri e propri, si legge: "When a library goes offline, it doesn't just interrupt research; it stifles educational progress, halts public access to information, and, in a new and chilling way, creates gaps in the public memory" (Messarra, Freeland, e Ziskina 2024).

[31] La Direttiva UE 2022/2557 (c.d. NIS2), all'Art. 2, ne fornisce la seguente definizione: «infrastruttura critica»: un elemento, un impianto, un'attrezzatura, una rete o un sistema o una parte di un elemento, di un impianto, di un'attrezzatura, di una rete o di un sistema, necessari per la fornitura di un servizio essenziale", dove si deve intendere per «servizio essenziale»: un servizio fondamentale per il mantenimento di funzioni vitali della società, di attività economiche, della salute e della sicurezza pubbliche o dell'ambiente.
Fino all'edizione di ottobre 2025, il rapporto Clusit non ha mai preso in considerazione le specificità delle infrastrutture del patrimonio culturale e dei cyber attacchi che le vedono come obiettivi. Tra le "vittime per categoria", infatti, non compare il settore cultura. Nell'edizione di ottobre 2025, è stato invece introdotto un Focus dal titolo "Analisi degli incidenti Cyber nel settore culturale italiano tra il 2020 e il 2024", a cura di un Gruppo di lavoro costituitosi all'interno del Corso di Alta Formazione Cyberhumanities for Heritage Security dell'Università di Roma Tre. Il Direttore del Corso è il prof. Emanuele Bellini (emanuele.bellini@uniroma3.it) che è stato tra i primi, in Italia, ad occuparsi delle problematiche legate alla cybersicurezza nel contesto del patrimonio culturale, e che si ringrazia per l'essenziale contributo alla redazione di questo lavoro.

[32] Il Perimetro di Sicurezza Nazionale Cibernetica (PSNC) è stato istituito con il D.L. 21 settembre 2019, n. 105. Il successivo D.P.R. 30 luglio 2020, n. 131 ha esplicitamente elencato i settori considerati di interesse nazionale. La Direttiva UE 2022/2555 ha ulteriormente ampliato tale perimetro includendo nuovi settori critici.

[33] Direttiva UE 2022/2555, Art. 6 – Definizioni: «Sicurezza dei sistemi informatici e di rete»: la capacità dei sistemi informatici e di rete di resistere, con un determinato livello di confidenza, agli eventi che potrebbero compromettere la disponibilità, l'autenticità, l'integrità o la riservatezza dei dati conservati, trasmessi o elaborati o dei servizi offerti da tali sistemi informatici e di rete o accessibili attraverso di essi.

[34] La Digital Preservation Coalition definisce la "Digital Preservation" come "the series of managed activities necessary to ensure continued access to digital materials for as long as necessary" e ne individua i principali rischi: "Rapid technological change may leave content unusable or unintelligable as the software that interprets it becomes obsolete. Without committed resources, the storage and management of digital content will not be possible. Organizational change might leave digital content without a committed custodian. Digital preservation requires a series of actions over time to ensure digital content remains alive, discoverable, accessible and usable": https://www.dpconline.org/digipres/what-is-digipres.

[35] Direttiva UE 2022/2555, Art. 6 – Definizioni: «incidente»: un evento che compromette la disponibilità, l'autenticità, l'integrità o la riservatezza di dati conservati, trasmessi o elaborati o dei servizi offerti dai sistemi informatici e di rete o accessibili attraverso di essi.

[36] Regolamento UE 2019/881, Art. 2 – Definizioni: «cibersicurezza»: l'insieme delle attività necessarie per proteggere la rete e i sistemi informativi, gli utenti di tali sistemi e altre persone interessate dalle minacce informatiche e «minaccia informatica»: qualsiasi circostanza, evento o azione che potrebbe danneggiare, perturbare o avere un impatto negativo di altro tipo sulla rete e sui sistemi informativi, sugli utenti di tali sistemi e altre persone.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-685

Verso il riconoscimento del C3E italiano delle biblioteche

> C3E is composed by complementary entities that are not only interconnected, but inter-reliant and work together to achieve a common value proposition or mission. C3E should be considered critical since "is crucial for the maintenance of vital societal functions" (Council Directive 2022/2557/EC on Resilience of Critical Entities). In fact, it is imperative to treat the (digital) Cultural Heritage as a critical infrastructure because of its role for the society's identity and cohesion.
> (Bellini 2025)

Un "C3E – *Cultural Cyber Critical Ecosystem"* è un ecosistema composto da più entità interconnesse e interdipendenti[37] tra loro, che lavorano insieme per il mantenimento della identità e della coesione sociale di una comunità. Tale *Mission*, che è anche la prima delle componenti del C3E, si può declinare nell'ecosistema biblioteche a partire dalle missioni della biblioteca pubblica definite in (IFLA, UNESCO 2022) cui si può aggiungere, più sul versante della tutela, l'istituto del Deposito Legale, che ha il precipuo fine di "conservare la memoria della cultura e della vita sociale italiana".[38] La seconda componente del C3E è quella degli *agents,* da intendere sia come comunità e organizzazioni (enti e istituzioni culturali, operatori di settore, associazioni, imprese culturali e creative ecc. e tutti i loro utenti o clienti), che agenti software. Secondo l'Indagine ISTAT sulle biblioteche pubbliche e private 2022[39], le biblioteche italiane aperte al pubblico sono 8.131, di cui 6.453 pubbliche e 1.678 private. Gli utenti che hanno usufruito di almeno un servizio nel corso dello stesso anno sono quantificati in 5.7 milioni, ma sono oltre 35 milioni gli accessi fisici. I numeri assoluti sul personale impiegato nelle biblioteche non sono disponibili ma, considerate le percentuali di addetti riportate nell' Indagine ISTAT sulle biblioteche 2023[40], possiamo certamente parlare di qualche decina di migliaia di operatori. D'altra parte, sappiamo che anche l'attenzione verso gli agenti software è in forte crescita. Lo dimostrano iniziative come il rilascio

[37] Esistono numerosi studi sulla interdipendenza delle infrastrutture critiche e sul conseguente "Effetto Domino" che può essere provocato dal fallimento di una o più di queste strutture, anche e soprattutto a seguito di un cyberattacco. Interessante, a questo proposito, il progetto DOMINO attivato nell'ambito del programma della Commissione Europea "*Prevention, Preparedness and Consequence Management of Terrorism and other Security related risks*" 2007-2013, che ha sviluppato un modello di valutazione del rischio e di propagazione degli eventi di *failure* alle infrastrutture economiche e sociali che operano nella nostra società: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=SWD:2018:331:FIN. Tra gli output del progetto, è stato realizzato anche il software di valutazione del rischio, la *DOMINO Platform*, che prende in considerazione tra gli item che caratterizzano il livello di qualità della vita anche "*Arts, sports, leisure and social activities, cultural traditions*": https://home-affairs.ec.europa.eu/document/download/0f2758c0-3cb2-4d9e-adfd-a725450cee96_en?-filename=Panel%202_Luisa%20Franchina_DOMINO.pdf.

[38] Legge 15 aprile 2004, n. 106 - Art. 1: https://www.normattiva.it/eli/stato/LEGGE/2004/04/15/106/ORIGINAL.

[39] ISTAT - Statistiche culturali 2022 - Dati sulle Biblioteche - Tavola 4.12: https://www.istat.it/tavole-di-dati/statistiche-culturali-anno-2022/, citata in (MiC 2024). Se si considerano anche le biblioteche non stabilmente aperte al pubblico, il numero sale considerevolmente. L'Anagrafe delle Biblioteche Italiane, al 18/7/2025, ne censisce 13.685: https://anagrafe.iccu.sbn.it/it/.

[40] ISTAT - Statistiche culturali 2023 - Dati "Biblioteche" - Tavola 4.9: https://www.istat.it/tavole-di-dati/statistiche-culturali-anno-2023/.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-685

di APIs da parte dell'ICCU[41] o i servizi basati sui dati in corso di sviluppo da parte dell'ICDP – Digital Library[42], nell'ambito del Piano nazionale di digitalizzazione del patrimonio culturale (PND)[43]. Sembra poi superfluo ribadire come, in questo momento, una grandissima quantità di traffico verso qualsiasi tipologia di servizio digitale esposto in rete sia generata dai crawler, alla ricerca di dati per l'addestramento degli LLMs[44]. Gli agenti interagiscono con i *Digital Cultural Assets* (DCA) ovvero con il Patrimonio culturale digitale nativo e non nativo, cuore dell'ecosistema. (Bellini 2025) identifica tre macro-criteri per la quantificazione del valore dei DCA:

- *Production Cost* (PC): il costo di produzione dei beni quindi i costi di digitalizzazione o i costi di produzione del bene nativo digitale.
- *Functional Value* (FV): una sorta di valore indotto dai possibili usi e riusi dei beni.
- *Cultural Value* (CV): il valore culturale e storico attribuito dalla comunità di esperti.

Il PC è in genere più facilmente quantificabile, almeno per i beni prodotti in tempi più recenti, mentre il FV e il CV possono chiaramente mutare molto in base al contesto, al tempo e ad altri fattori contingenti[45]. Nel calcolo del valore dei DCA basato sul costo di produzione possono essere conteggiati, innanzitutto, gli investimenti fatti nel corso degli ultimi 40-50 anni nella catalogazione del patrimonio ovvero nella produzione di metadati descrittivi. Il solo Opac SBN indicizza 21.256.681 record per 125.675.014 localizzazioni[46]. Vi è poi il costo della digitalizzazione del patrimonio delle biblioteche: le digitalizzazioni attualmente accessibili dall'Opac SBN sono 951.889[47], alle quali, entro il 2026, si aggiungeranno le digitalizzazioni provenienti dai cantieri del PND. Per la digitalizzazione di 107 mila microfilm conservati presso il Centro Nazionale per lo Studio del Manoscritto (CNSM) saranno spesi complessivamente 9.2 milioni di euro[48]. Altri 27.9 milioni sono stati investiti per la digitalizzazione della documentazione catastale conservata negli Archivi di Stato e dei giornali quotidiani postunitari, appartenenti alle maggiori Biblioteche statali[49]. Inoltre, Regioni e Province autonome hanno ricevuto 26 milioni di euro per la digitalizzazione di materiali cartacei e fotografici, e 650.000 euro per la digitalizzazione di microfilm di documenti archivistici e bibliografici di cui sono custodi[50].

---

41 https://www.iccu.sbn.it/it/attivita-servizi/dati-aperti/.

42 Si vedano il catalogo degli e-services di I.PaC sulla Piattaforma Digitale Nazionale Dati (PDND) e la piattaforma DPaaS "un ambiente di sviluppo e servizi tecnologici innovativi, per creare prodotti innovativi basati sui dati in modo semplice e scalabile": https://digitallibrary.cultura.gov.it/approfondimenti/ecomic-dpaas/.

43 https://digitallibrary.cultura.gov.it/il-piano/.

44 Sul tema si veda (De Robbio 2025).

45 Si pensi, ad esempio, al dibattito sulla normativa sul riuso delle immagini dei beni culturali che vede contrapposti coloro che difendono gli incassi diretti derivanti dalla vendita delle licenze d'uso e coloro che sostengono che consentire il libero riuso genererebbe un indotto molto superiore; a questo proposito si veda (Modolo 2025).

46 Dato al 14/7/2025: https://opac.sbn.it/.

47 Dato al 14/7/2025: https://opac.sbn.it/.

48 https://digitallibrary.cultura.gov.it/avvisi/pnrr-digitalizzazione-del-patrimonio-culturale-online-la-procedura-di-gara-da-92-milioni-di-euro-per-la-categoria-microfilm-di-manoscritti-antichi/.

49 https://digitallibrary.cultura.gov.it/avvisi/pnrr-digitalizzazione-del-patrimonio-culturale-online-la-procedura-di-gara-da-279-milioni-di-euro-per-la-categoria-carta-documentazione-catastale-registri-e-mappe-e-giornali-quotidiani/.

50 https://pnrr.cultura.gov.it/gare-per-la-digitalizzazione-del-patrimonio-culturale-di-regioni-e-province-autonome-m1c31-1-strategie-e-piattaforme-digitali-per-il-patrimonio-culturale/.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-685

Il PC del Patrimonio nativo digitale per le biblioteche può forse essere inteso più come costo di acquisizione delle risorse native digitali, quali i costi per l'accesso permanente a repertori e banche dati, o quelli di acquisizione e gestione del deposito legale digitale.
Pur evitando di addentrarsi in analisi sul valore indotto dall'uso e riuso delle risorse e dei servizi delle biblioteche[51], che esulerebbe forse le finalità di questo contributo, il FV dell'ecosistema delle biblioteche può essere, però, certamente ricondotto almeno ai servizi di consultazione e prestito delle risorse, sia digitali che no, tra i quali si annoverano anche servizi di rete come l'ILL e il Document Delivery. Infine, in relazione al processo che porta la comunità di esperti ad attribuire un CV ai DCA, sembra doveroso sottolineare il ruolo che ha la creazione del catalogo che, non a caso, è anche una delle *functions* principali che l'ecosistema svolge nei confronti dei DCA, durante tutto il loro ciclo di vita: "Il processo di contabilizzazione di un *heritage item* come *asset* necessita innanzitutto di una fase di rilevazione (*recognition*), che consiste nel raccogliere ed elaborare i dati disponibili al fine di rappresentarli nei prospetti contabili" (Aprile, Grandis, e Mocavini 2018). Altre funzioni sono quelle di digitalizzazione e/o acquisizione, metadatazione, accesso e fruizione, valorizzazione, conservazione e, non ultimo, cybersicurezza. Lo svolgimento di queste funzioni richiede chiaramente specifiche competenze e risorse tecnologiche. L'ultima componente del C3E sono gli *Hybrid Infrastructure Systems*, essenziali non solo per la tenuta degli asset digitali ma anche di quelli tradizionali. Essi comprendono una enorme varietà di infrastrutture e servizi cyber, come cataloghi e inventari online[52], ILS e LSP[53], biblioteche e repository digitali[54], siti web, apps, sistemi di autenticazione, piattaforme per l'ILL, il Document Delivery e il prestito digitale, servizi basati su IA, sistemi di reportistica e di analisi dei dati, apparati di rete, sistemi di controllo degli accessi, sistemi di monitoraggio ambientale delle sale e dei magazzini, sistemi di allarme. Questa varietà è uno dei fattori che rende la sicurezza del C3E così difficile da proteggere, perché ciascuno di questi sistemi è potenzialmente anche un *single point of failure*[55]. Di seguito, ci si soffermerà più nel dettaglio sulle infrastrutture che sottendono alcune delle funzioni chiave del C3E delle biblioteche italiane ma, intanto, è bene ricordare che, nell'ambito del PND, sono in corso di realizzazione una serie di infrastrutture e servizi che serviranno anche il comparto biblioteche (Cerullo et al. 2024). In ultimo, un C3E dovrebbe garantire il mantenimento nel tempo di tre *properties* caratterizzanti e interdipendenti (Bellini 2025):

- Affidabilità (*trust*): la capacità di preservare nel tempo le principali caratteristiche dei DCA ovvero accessibilità, autenticità, integrità, leggibilità[56].

[51] A questo proposito, di sicuro interesse sarebbe una rassegna degli studi esistenti sul rapporto tra biblioteche e istruzione scolastica e formazione continua, e tra biblioteche e ricerca scientifica.
[52] Nel 2023, il 72,6% delle biblioteche italiane aveva un catalogo in formato digitale: ISTAT - Statistiche culturali 2023 - Dati "Biblioteche" - Tavola 4.16: https://www.istat.it/tavole-di-dati/statistiche-culturali-anno-2023/.
[53] Nel 2023, il 45,6% delle biblioteche aveva una piattaforma web per l'erogazione di servizi on-line, https://www.istat.it/tavole-di-dati/statistiche-culturali-anno-2023/.
[54] Nel 2023, il 15,8% delle biblioteche aveva una teca digitale, propria o condivisa a livello locale o nazionale, https://www.istat.it/tavole-di-dati/statistiche-culturali-anno-2023/.
[55] https://en.wikipedia.org/wiki/Single_point_of_failure.
[56] Cfr. Glossario dei termini e degli acronimi allegato alle Linee Guida AgID sulla formazione, gestione e conservazione dei documenti informatici: https://docs.italia.it/AgID/documenti-in-consultazione/lg-documenti-informatici-docs/it/bozza/allegati.html.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-685

- Resilienza (*resilience)*: la capacità di prevedere e gestire i rischi cyber potenziali[57].
- Sostenibilità (*sustainability)*: da intendersi come sostenibilità tecnica, ambientale, organizzativa, culturale ed economica.

SBN e Magazzini Digitali come componenti portanti del *Library Cyber Critical Ecosystem* nazionale

Nel contesto di quello che, a questo punto, potremmo definire il *Library Cyber Critical Ecosystem* italiano, le infrastrutture afferenti a SBN e al servizio Magazzini Digitali[58] non possono che essere considerate componenti portanti per l'assolvimento delle funzioni dell'ecosistema stesso, vale a dire:

- Censimento del patrimonio: 7.120 delle 8. 131 biblioteche aperte al pubblico sono parte della rete SBN[59]. Ciò significa che, al netto del noto problema dei dati "sommersi"[60], la maggioranza delle informazioni sul patrimonio bibliografico o di altra natura delle biblioteche italiane è custodita all'interno dell'Indice SBN.
- Affidabilità delle informazioni sul lungo periodo: l'autorevolezza delle istituzioni che producono il censimento del patrimonio culturale è garanzia, nel lungo periodo, dell'affidabilità delle informazioni contenute.
- Tutela del patrimonio fisico: il catalogo SBN va considerato nella sua duplice veste di strumento di accesso e strumento di tutela. Esso, infatti, previene perdite e sottrazioni illecite e ne consente, laddove possibile, il ripristino.
- Circolazione dei documenti a livello locale, nazionale e internazionale: il catalogo SBN, come punto di accesso iniziale ai servizi di circolazione locale, e il servizio di ILL SBN sono infrastrutture essenziali per il prestito e la consultazione dei documenti.
- Supporto tecnologico al processo di patrimonializzazione[61] dei documenti nativi digitali: Magazzini digitali, come servizio di conservazione e accesso di lungo periodo ai documenti nativi digitali, è attualmente l'unica infrastruttura bibliotecaria potenzialmente in grado di sostenere il processo di patrimonializzazione del nativo digitale che, viceversa, è soggetto a perdite più o meno accidentali dovute a cause come il mancato riconoscimento del valore del loro contenuto informativo[62], l'obsolescenza tecnologica o la modifica unilaterale delle condizioni di accesso definite da editori, produttori o distributori dei documenti nativi digitali.

[57] Il primo documento che definisce ufficialmente il concetto di ciberresilienza è il Regolamento UE 2024/2847 c.d. Cyber Resilience Act: http://data.europa.eu/eli/reg/2024/2847/oj.

[58] https://bncf.cultura.gov.it/servizi/magazzini-digitali/.

[59] Dato al 25/7/2025: https://www.iccu.sbn.it/it/SBN/poli-e-biblioteche/.

[60] In questa dicitura è possibile far rientrare i dati non ancora recuperati dai cataloghi legacy, compresi quelli cartacei, il non catalogato e i dati presenti solo nei cataloghi di Polo.

[61] "Per processo di patrimonializzazione si identifica l'insieme delle azioni, promosse su istanze sociali di tipo culturale, tecnico-scientifico e giuridico-amministrativo, attraverso le quali un qualsiasi oggetto, materiale, immateriale o digitale, viene considerato degno di sopravvivere al deperimento naturale per essere conservato nel tempo come testimonianza di civiltà" (ICDP 2023).

[62] A questo proposito, si pensi in particolare all'attività di creazione e gestione degli archivi del Web.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-685

I rischi specifici, reali o potenziali, per l'affidabilità, la resilienza e la sostenibilità di queste infrastrutture afferiscono a tre ambiti: quello legale, quello economico-gestionale e quello tecnologico.

Contesto legale

Nonostante SBN e Magazzini Digitali assolvano in tutto o in parte alle funzioni appena elencate, rispettivamente da 40 e 20 anni circa, un pieno riconoscimento del loro valore a livello istituzionale e legale non può dirsi compiuto. SBN è stata inserita tra le "basi dati di interesse nazionale" all'interno del Piano triennale per l'informatica nella Pubblica Amministrazione 2019-2021[63], ma nei successivi aggiornamenti il riferimento è stato espunto senza spiegazioni. Allo stesso tempo, la non esecutività dell'obbligo di deposito legale dei documenti nativi digitali[64] ha già causato e sta continuando a causare la perdita di moltissima documentazione di interesse nazionale. A ciò si aggiunge che nemmeno la più recente normativa comunitaria in materia di cybersicurezza, la Direttiva UE 2022/2555 (c.d. NIS2) e il D.lgs. 4 settembre 2024, n. 138 di recepimento della NIS2, sembra fornire indicazioni dirimenti per il settore cultura[65]. Questo, infatti, non viene nominato né tra i settori ad alta criticità[66], come forse è normale che sia, né tra gli altri settori critici[67], tra i quali viene invece compreso quello della Ricerca. Tuttavia, il legislatore italiano dimostra una certa sensibilità alla questione nominando nell'Allegato III, dedicato alle Amministrazioni centrali, regionali, locali e di altro tipo, cui certamente moltissimi istituti e luoghi della cultura afferiscono, anche gli "Enti produttori di servizi assistenziali, ricreativi e culturali"; e includendo nell'Allegato IV, caso unico nel panorama europeo, "Soggetti che svolgono attività di interesse culturale", qui da intendersi quindi come soggetti privati. Bisogna anche sottolineare che il Ministero della Cultura è Autorità di settore NIS[68] e, tra i suoi compiti di supporto all'Agenzia per la cybersicurezza nazionale (ACN)[69], vi è quello di convalidare l'elenco dei soggetti NIS per settore di competenza[70], proponendo eventuali ulteriori identificazioni governative. Purtroppo, il registro dei soggetti NIS non è pubblicamente disponibile, e non sappiamo se lo sarà mai, dal momento che, vigente la Direttiva NIS1[71], il registro era secretato.

Modello gestionale e di sostenibilità economica

Il modello gestionale e di sostenibilità economica di SBN, come ideato all'inizio degli anni '80

[63] https://docs.italia.it/italia/piano-triennale-ict/pianotriennale-ict-doc/it/2019-2021/05_dati-della-pubblica-amministrazione.html.
[64] Si veda infra (Bergamin e Messina 2025).
[65] www.gazzettaufficiale.it/eli/id/2024/10/01/24G00155/sg.
[66] I settori ad alta criticità sono elencati nell'Allegato I al D.lgs. 4 settembre, n. 138.
[67] D.lgs. 4 settembre, n. 138, Allegato II.
[68] D.lgs. 4 settembre, n. 138, art. 11.
[69] https://www.acn.gov.it/.
[70] Soggetti pubblici e privati appartenenti alle tipologie di cui agli allegati I, II, III e IV del D.lgs. 4 settembre 2024, n. 138. Tali soggetti, in base all' art. 7 del medesimo decreto, sono tenuti all'obbligo di registrarsi sulla piattaforma dedicata della ACN: https://www.acn.gov.it/portale/nis/registrazione.
[71] http://data.europa.eu/eli/dir/2016/1148/oj.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-685

del '900 e consolidato nei due decenni successivi, rappresenta un *unicum* positivo, anche in una prospettiva allargata al contesto internazionale. Gli organi di governo SBN sono, però, in attesa di rinnovo da ormai quasi dieci anni[72], e questo ha comportato un totale addossamento dell'onere decisionale e della responsabilità operativa sull'ICCU che, teoricamente, ne dovrebbe essere il "gestore"[73]. Il venir meno del modello gestionale originale, a sua volta basato su una architettura tecnologica nativamente decentrata e distribuita, ha avuto un forte impatto anche sulla sostenibilità economica che ne discendeva. E le difficoltà di sostenere una infrastruttura così complessa hanno via via fatto propendere per un suo accentramento tecnologico. L'attuale modello di gestione e sviluppo sia del sistema Indice che dei sistemi di Polo, infatti, non sembra facilitare l'aggiornamento e l'innovazione, anche a causa di costi assai elevati, non più dettati dal mercato ma da situazioni di *lock-in* di fatto[74]. Stessa sorte subisce anche il servizio "Magazzini Digitali". Nato come sperimentazione sulle base delle previsioni dell'art. 37 del D.P.R. 3 maggio 2006, n. 252, è stato inizialmente implementato grazie alla collaborazione tecnico-scientifica e il co-finanziamento non solo delle due Biblioteche nazionali centrali, ma anche della Biblioteca nazionale Marciana e della Fondazione Rinascimento Digitale[75]. La prolungata assenza del regolamento tecnico attuativo del deposito legale digitale ha, tuttavia, distolto attenzione e fondi dal servizio per lunghi periodi, facendo convergere sulla Biblioteca nazionale centrale di Firenze la quasi totalità delle responsabilità tecnico-gestionali oltre che degli oneri finanziari, e rendendolo per questo non pienamente efficiente.

Potenziali criticità tecnologiche

La migrazione, in corso di completamento nel 2025, di tutte le infrastrutture e i servizi SBN[76] al Polo Strategico Nazionale (PSN)[77], per il tramite del Ministero della Cultura, ha avuto e avrà certamente un impatto fortemente positivo sulla questione della sicurezza. È bene specificare, però, che all'interno del PSN, l'approccio alla sicurezza è basato sullo *Shared Security Responsibility Model*, uno strumento attraverso il quale Cloud Service Provider (il PSN) e Cloud Service Customer (la PA che usufruisce dei servizi cloud) definiscono e regolano in che modo la responsabilità e l'accountability per la sicurezza dei dati e delle risorse venga suddivisa nell'ambito di uno specifico servizio Cloud[78]. Ciò significa che per le biblioteche il problema della sicurezza non si esaurisce

[72] https://www.iccu.sbn.it/it/SBN/organi-di-governo-sbn/.

[73] È doveroso sottolineare che coloro che si sono succeduti, in questi anni, alla direzione dell'Istituto centrale per il Catalogo Unico hanno più volte sollecitato, nelle opportune sedi, il rinnovo degli organi di governo SBN e che si sono fatti promotori della prima Assemblea dei Poli tenutasi nel 2015 e di quella in programma a dicembre 2025; purtroppo questi appelli sono rimasti evidentemente inascoltati.

[74] Nonostante l'attenzione sempre più diffusa per l'adozione di soluzioni hardware standard e software open source, sia oggettivi ostacoli tecnologici, quali la migrazione di grandi quantità di dati, che i costi del *reverse engineering* di soluzioni non largamente adottate e/o sostenute da una vasta community di sviluppo (spesso basate su linguaggi e formati estremamente customizzati), rendono di fatto la possibile concorrenza delle aziende sul mercato molto limitata.

[75] https://wayback.archive-it.org/org-1284/20231204094728/https://www.depositolegale.it/accordo-mibac-editori/.

[76] Indice e Opac SBN, SBNCloud, e altre banche dati come Manus e Edit XVI. Le infrastrutture per la gestione e conservazione del patrimonio digitalizzato, essendo parte del PND, sono ospitate dal Consorzio Cineca.

[77] https://www.polostrategiconazionale.it/.

[78] https://www.polostrategiconazionale.it/chi-siamo/sicurezza/matrici-di-responsabilita-condivisa-della-sicurezza/.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-685

con il passaggio al PSN. Allo stesso modo, la scelta della BNCF, nel 2023, di entrare a far parte del Consorzio Cineca per la gestione dei servizi di Magazzini Digitali[79], dettata primariamente dall'esigenza di affidarsi ad un partner pubblico dotato di competenze specifiche e infrastrutture efficienti, si è rivelata vincente anche dal punto di vista della sicurezza[80]. Se, però, si amplia la prospettiva all'intero *Library Cyber Critical Ecosystem* italiano, lo scenario tecnologico appare maggiormente problematico. Innanzitutto, se è vero che teoricamente tutti i dati e i servizi bibliotecari della PA centrale sono transitati o è previsto che transitino sul PSN, non è detto che questo passaggio copra davvero il 100% dei sistemi. Si pensi, ad esempio, a cataloghi e siti web, di istituzioni grandi e piccole, spesso non considerati critici, che vengono mantenuti in locale o in hosting a provider terzi non sempre dotati delle certificazioni previste da ACN. Dotati delle certificazioni previste ACN dovrebbero, invece, essere tutti i provider selezionati dalle PA locali per la migrazione ad ambienti cloud incentivata nell'ambito del Piano Nazionale di Ripresa e Resilienza[81], e in parte già realizzata, anche prima del PNRR, soprattutto in alcune Regioni in cui sono storicamente presenti società e infrastrutture dedicate allo sviluppo e supporto ICT[82]. Così come la rispondenza ai requisiti di sicurezza dovrebbe essere imprescindibile nella scelta di fornitori di servizi ormai largamente diffusi nelle biblioteche, come il prestito digitale o l'accesso alle banche dati[83]. Trattandosi di un caso che interessa il servizio di Web archiving di Magazzini Digitali, bisogna anche citare l'eventualità in cui l'hosting di taluni servizi sia affidato a provider terzi non certificati ACN, a causa dall'effettiva mancanza di alternative sul mercato, che equivarrebbe all'impossibilità di erogare i servizi stessi[84]. Appare pertanto chiaro come anche sul piano delle soluzioni tecnologiche il panorama sia particolarmente variegato, e come da ciò possano dipendere livelli di cybersicurezza differenti. Certo, purché esistano dei backup e siano rispettate almeno le previsioni del GDPR, i potenziali

[79] Il Consorzio Cineca ospita già il servizio di conservazione e accesso alle tesi di dottorato (https://tesidottorato.depositolegale.it), ma è in corso di definizione anche l'affidamento per la reingegnerizzazione del servizio di rilascio del National Bibliography Number (NBN): https://bncf.cultura.gov.it/servizi/magazzini-digitali/#cap-5. Infine, è stato discusso il progetto di creazione di una infrastruttura per la conservazione e l'accesso di lungo periodo agli altri prodotti della ricerca.

[80] Le infrastrutture e i servizi cloud per la PA del Consorzio sono conformi a quanto previsto dal Regolamento ACN in materia: https://www.acn.gov.it/portale/cloud.

[81] https://padigitale2026.gov.it/misure.

[82] Si pensi a Lepida S.c.p.A. in Emilia-Romagna (https://www.lepida.net/) o a TIX in Toscana (https://www.cloud.toscana.it/tix/).

[83] Il principale riferimento in questo ambito è il Catalogo delle Infrastrutture digitali e dei Servizi cloud dell'ACN: https://www.acn.gov.it/portale/catalogo-delle-infrastrutture-digitali-e-dei-servizi-cloud.

[84] Il servizio di Web archiving di Magazzini Digitali viene erogato attraverso la piattaforma Archive-it di Internet Archive che, seppure non certificata ACN, garantisce la conservazione di copie ridondate dei dati in datacenter indipendenti e la loro disponibilità per il download in qualunque momento: https://archive-it.org/archive-it/. Bisogna, però, ricordare che la tenuta di un servizio di conservazione non dipende solo dalle soluzioni tecnologiche adottate, ma anche dalla robustezza istituzionale ed economica del soggetto conservatore nonché dal contesto legale e culturale in cui il soggetto opera. Brewster Khale, fondatore di Internet Archive, ha più volte dichiarato come una delle principali minacce alla loro attività derivi dalle battaglie legali intentategli dagli editori sulla liceità di alcuni servizi in relazione alle norme USA sul copyright. Per difendersi nei tribunali, infatti, la Fondazione ha perso centinaia di migliaia di dollari e rischia sanzioni milionarie, che minerebbero la capacità di Internet Archive di mantenere le proprie infrastrutture e i propri servizi così come fatto finora. A ciò, si aggiungono le attuali politiche del governo Trump che non solo hanno tagliato i fondi destinati alle istituzioni culturali e di ricerca, pubbliche e private, ma hanno anche creato un clima poco favorevole al lavoro di Internet Archive: https://americanlibrariesmagazine.org/2025/06/04/newsmaker-brewster-kahle/#:~:text=Since%20founding%20the%20Internet%20Archive,him%20feel%20%E2%80%9Cvery%20encouraged.%E2%80%9D.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-685

danni materiali potrebbero essere ritenuti contenuti, ma la criticità di un servizio si misura anche e soprattutto in base alla mission dell'istituzione: la prolungata indisponibilità dell'accesso a siti web, cataloghi o repository di biblioteca può andare non solo a immediato discapito dell'attività di studenti e ricercatori, ma soprattutto della reputazione dell'istituzione[85]. Per non parlare dei sistemi digitali per il controllo fisico degli ambienti che se non disponibili, anche per un breve periodo, possono costituire un problema per l'incolumità fisica delle persone e per la salvaguardia del patrimonio culturale. Se, ad esempio, ad essere attaccato fosse il sistema di controllo varchi di una biblioteca, l'istituto potrebbe non essere più nelle condizioni di sapere quante persone sono all'interno dell'edificio o quante di loro hanno lasciato la biblioteca con documenti rari o preziosi. Infine, in un ecosistema in cui tutte le componenti sono interdipendenti, non è possibile ignorare che, visto anche il contesto normativo descritto, le infrastrutture tecnologiche delle istituzioni culturali private possano essere ancora di più da attenzionare.

Conclusioni: dalla *Cyber security* alla *Cyber resilience*

Considerata la complessità del quadro fin qui delineato, che implica anche evidenti difficoltà nell'attribuzione univoca delle responsabilità in questo ambito, sarebbero innanzitutto auspicabili azioni di sensibilizzazione ed advocacy sul tema della cybersicurezza da parte delle istituzioni deputate alla tenuta delle infrastrutture bibliotecarie, come l'ICCU e le Biblioteche nazionali centrali, ma anche da parte delle associazioni professionali, come l'AIB o l'AIUCD. Azioni che dovrebbero essere seguite da una formazione specifica, da differenziare in base al livello di coinvolgimento nei processi che garantiscono la sicurezza di sistemi e servizi ma che tocchi, appunto, tutti i livelli. Sembra, infatti, di dover rilevare come proprio in un settore come quello delle biblioteche, che vanta una tradizione millenaria nella tutela del patrimonio culturale, si fatichi non poco ad affrontare un doveroso cambiamento culturale nella gestione del rischio cyber. Una formazione mirata degli addetti ai lavori, e degli utenti che interagiscono con il patrimonio, potrebbe sperabilmente almeno evitare le continue lamentele sulla richiesta di generare password "inutilmente complicate". Una volta acquisita consapevolezza dei rischi, però, per evitare danni economici e reputazionali talvolta molto ingenti, è necessario investire in sicurezza. L'elenco delle "misure tecniche, operative e organizzative adeguate [...] alla gestione dei rischi posti alla sicurezza dei sistemi informativi e di rete [...] nonché per prevenire o ridurre al minimo l'impatto degli incidenti per i destinatari dei [...] servizi", previste nel D.lgs. 4 settembre 2024, n. 138[86], mostra proprio l'importanza di approcciarsi alla gestione della cybersecurity come alla gestione di un processo articolato, che va dalla analisi delle vulnerabilità e dei rischi potenziali (c.d. *vulnerability assessment*) alla definizione delle azioni concrete afferenti ad una più ampia *cyber resilience strategy* in caso di attacco. Una rassegna delle misure di cybersicurezza specifica per biblioteche digitali si trova in (Tammaro e Bellini 2024). Ancor prima, sarebbe importante chiarire il quadro normativo illustrato e definire, quindi, per ciascuna funzione del C3E delle biblioteche quali sono i livelli di servizio che devono essere garantiti in termini di resilienza. Sulla base degli obiettivi, sarà possi-

[85] https://www.ilpost.it/2024/05/22/british-library-conseguenze-attacco-informatico/.
[86] D.lgs. 4 settembre 2024, n. 138, art. 24 commi 1 e 2.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-685

bile scegliere il modello gestionale e le soluzioni tecnologiche adeguati. Ciò che desta maggiore preoccupazione sono ovviamente gli ingenti investimenti di cui necessita la cybersicurezza, per queste attività costanti di monitoraggio e aggiornamento di tutti i sistemi critici. Tuttavia, come ci insegna il caso della British Library, il costo della prevenzione è comunque sempre inferiore al costo del fallimento (The British Library 2024).

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-685

Riferimenti bibliografici*

Aprile, Rocco, Fabio Giulio Grandis, e Fabrizio Mocavini. 2018. «Heritage come asset nella contabilità economico-patrimoniale delle Amministrazioni pubbliche.» *Rivista della Corte dei conti* 3-4: 429–40. https://www.corteconti.it/Download?id=7cbc64ff-575f-46fd-93a6-39dd52b6fb2f.

Bellini, Emanuele. 2025. «Cyber Humanities for Heritage Security». Commun. ACM 68, 12 (December 2025), 112–117. https://doi.org/10.1145/3735659.

Cerullo, Luigi, Margherita Bartoli, Lina Antonietta Coppola, Costantino Landino, Antonio Davide Madonna, Antonella Negri, Giovanni Pescarmona, Chiara Fauda Pichet, Margherita Porena, e Valentina Rossetti. 2024. «Verso la creazione di un Ecosistema digitale nazionale per la Cultura.» *DigItalia* 19 (2): 11–48. https://doi.org/10.36181/digitalia-00101.

Clusit - Associazione Italiana per la Sicurezza Informatica. 2012. *Rapporto Clusit sulla sicurezza ICT in Italia*. https://clusit.it/wp-content/uploads/download/Rapporto_Clusit%202012.pdf.

Clusit - Associazione Italiana per la Sicurezza Informatica. 2025. *Rapporto Clusit sulla Cybersecurity in Italia e nel mondo*. https://clusit.it/rapporto-clusit/.

De Robbio, Antonella. 2025. «I bot dell'AI all'assalto delle biblioteche pubbliche.» *Il BO Live*. 26 luglio 2025. https://ilbolive.unipd.it/it/news/societa/bot-dellai-allassalto-biblioteche-pubbliche.

ICDP (Istituto Centrale per la Digitalizzazione del Patrimonio Culturale) - Digital Library. 2023. *Piano nazionale di digitalizzazione del patrimonio culturale (PND)*. https://docs.italia.it/italia/icdp/icdp-pnd-docs/it/v1.1-febbraio-2023/index.html.

IFLA (International Federation of Library Associations and Institutions) e UNESCO (United Nations Educational, Scientific and Cultural Organization). 2022. «Manifesto IFLA-UNESCO delle biblioteche pubbliche 2022.» *AIB studi* 62 (2): 431–34. https://doi.org/10.2426/aibstudi-10097.

Messarra, Luca, Chris Freeland, e Juliya Ziskina. 2024. «Vanishing Culture: A Report on Our Fragile Cultural Record.» *Internet Archive*. https://blog.archive.org/2024/10/30/vanishing-culture-a-report-on-our-fragile-cultural-record/.

MiC - Ministero della Cultura. 2024. *Minicifre della cultura. Una raccolta di dati statistici sulla cultura*. https://www.fondazionescuolapatrimonio.it/wp-content/uploads/2024/12/Minicifre-della-cultura_edizione-2024.pdf.

Modolo, Mirco. 2025. «Libero riuso delle immagini del patrimonio culturale pubblico: profili di compatibilità con il codice dei beni culturali e del paesaggio.» In *Strumenti del Dipartimento di Giurisprudenza di Siena,* a cura di Mario Perini, 3: 29–56. Florence: Firenze University Press, USiena Press. https://doi.org/10.36253/979-12-215-0702-7.07.

Neglia, Grazia, Mariarosaria Angrisano, Ippolita Mecca, e Francesco Fabbrocino. 2024. «Cultural Heritage at Risk in World Conflicts: Digital Tools' Contribution to Its Preservation.» *Heritage* 7 (11): 6343–65. https://doi.org/10.3390/heritage7110297.

* Ultima data di consultazione dei siti web: 26 luglio 2025.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-685

Ovenden, Richard. 2021. *Bruciare libri: la cultura sotto attacco: una storia millenaria*. Tradotto da Luisa Doplicher e Daniele A. Gewurz. Milano: Solferino.

Tammaro, Anna Maria, e Emanuele Bellini. 2024. «Il ruolo della cybersecurity per le biblioteche digitali: sfide attuali e strategie di protezione.» *DigItalia* 19 (2): 100–14. https://doi.org/10.36181/digitalia-00104.

The British Library. 2024. «Learning lessons from the cyber-attack. British Library cyber incident review.» *The British Library*. https://www.bl.uk/home/british-library-cyber-incident-review-8-march-2024.pdf/.

Uddin, Rafe, e Daniel Thomas. 2024. «British Library to burn through reserves to recover from cyber attack.» *Financial Times*, maggio.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-697

JLIS.it

Digital legal deposit: cooperation, preservation, and new access opportunities

Giovanni Bergamin[(a)], Maurizio Messina[(b)]

a) Formerly National Central Library of Florence, https://orcid.org/0000-0002-2912-5662
b) Formerly Marciana National Library, https://orcid.org/0009-0004-7011-4043

Contact: Giovanni Bergamin, giovanni.bergamin@gmail.com; Maurizio Messina, mauriziomess@gmail.com
Received: 31 July 2025; **Accepted:** 22 October 2025; **First Published:** 15 January 2026

ABSTRACT
This paper examines the profound transformation of library collections in the digital era and the pivotal role of legal deposit. In Italy, the incomplete legal framework for digital resource deposit has resulted in the loss of two decades of cultural production. Legal deposit remains the primary mechanism for building a national cultural heritage collection, a concept widely embraced.
International cooperation models are examined, such as the UK's collaborative system, which mitigated the impact of a severe cyberattack on the British Library in 2023. The study underscores the need for robust cybersecurity and considers decentralized storage solutions like private IPFS networks.
New AI-driven tools for cataloguing and metadata creation, such as Annif, and innovative access methods are analyzed. Italian law now permits Text and Data Mining (TDM) on legal deposit collections for scientific research. A Retrieval-Augmented Generation (RAG) model is proposed for *computational access*, ensuring transparency by linking results to source documents.
The paper advocates for a new cooperative governance model, potentially supported by a publicly owned in-house company, to deliver advanced and sustained technological services to legal deposit institutions, ensuring effective management of digital legal deposit systems and aligning with libraries' evolving roles in cultural and social ecosystems.

KEYWORDS
Digital Legal Deposit Collection; Cybersecurity; Automated library metadata creation; Computational access; Text and Data Mining (TDM); Retrieval-Augmented Generation (RAG).

Deposito legale digitale: cooperazione, conservazione e nuove modalità di fruizione

ABSTRACT
Il presente contributo analizza la profonda trasformazione delle collezioni bibliotecarie nell'era digitale e il ruolo fondamentale del deposito legale. In Italia, l'incompletezza del quadro normativo per il deposito delle risorse digitali ha comportato la perdita di due decenni di produzione culturale. Il deposito legale è universalmente riconosciuto come il principale meccanismo per la costituzione della collezione del patrimonio culturale nazionale.
Vengono esaminati modelli di cooperazione internazionale, come il sistema collaborativo del Regno Unito, che ha mitigato l'impatto di un grave attacco informatico subito dalla British Library nel 2023. Lo studio sottolinea l'esigenza di robuste infrastrutture di cybersecurity e prende in considerazione soluzioni di archiviazione decentralizzata, quali le reti IPFS private.
Sono altresì analizzati nuovi strumenti basati sull'intelligenza artificiale per la catalogazione e la creazione di metadati, come Annif, nonché metodi di accesso innovativi. La normativa italiana consente ora il Text and Data Mining (TDM) sulle collezioni in deposito legale per la ricerca scientifica. Viene proposto un modello di Retrieval-Augmented Generation (RAG) per l'accesso computazionale, che garantisce trasparenza collegando i risultati ai documenti fonte.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-697

Il contributo sostiene infine l'adozione di un nuovo modello di governance cooperativa, potenzialmente sorretto da una società in-house a capitale pubblico, finalizzato a erogare servizi tecnologici avanzati e sostenibili per le istituzioni depositarie, assicurando una gestione efficace dei sistemi di deposito legale digitale e allineandosi all'evoluzione del ruolo delle biblioteche negli ecosistemi culturali e sociali.

PAROLE CHIAVE
Collezione del deposito legale digitale nazionale; Cybersecurity; Creazione automatizzata dei metadati bibliotecari; Accesso computazionale; Text and Data Mining (TDM); Retrieval-Augmented Generation (RAG).

Il cambiamento della natura delle collezioni bibliotecarie nell'era digitale è profondo e tale da costituirne il maggiore tratto identitario. Non si tratta certo di un fenomeno nuovo: ogni cambiamento nelle modalità di produzione e diffusione della conoscenza si è accompagnato, in una mutua relazione di causa/effetto, al mutamento dell'universo (lato sensu) bibliografico (Meschini 2021), e le collezioni delle biblioteche hanno necessariamente recepito sia i mutamenti del mercato editoriale che le diverse modalità d'uso dell'informazione registrata indotte dall'evoluzione tecnologica.

Nel mondo digitale le biblioteche, oltre a doversi districare fra la solida tradizione del *possesso* di documenti e la frontiera continuamente mutevole dell'*accesso* alle risorse informative, con tutti i connessi problemi delle tecniche di controllo bibliografico di collezioni divenute ibride e multiformi (Guerrini 2021), si sono scoperte non più sole nella loro consueta attività di mediazione informativa ma immerse in un ambiente fortemente concorrenziale, con svariati altri attori impegnati, con successo, nel medesimo *core business* (Kempf 2013). Le strategie di gestione della concorrenza, delle alleanze e degli scontri con gli attori commerciali, con i monopolisti dell'informazione scientifica, gli aggregatori di risorse, i servizi dei motori di ricerca risentono delle situazioni locali e vedono in posizione migliore le biblioteche che operano all'interno di reti cooperative ben strutturate e capaci di maggiore forza contrattuale. Le biblioteche restano dunque a pieno titolo, come del resto sono state storicamente, nella filiera che va dalla produzione alla disseminazione della conoscenza, e all'interno della filiera possono instaurare relazioni di servizio reciprocamente proficue con gli altri attori che ne fanno parte o, al contrario, venire marginalizzate. Si tratta quindi di valorizzare le strutture identitarie delle biblioteche, e la collezione (o raccolta) è certamente fra queste.

Nella comunità professionale sono emerse sensibilità e posizioni diverse al riguardo, derivanti dal ruolo che le biblioteche rivestono all'interno di vari ecosistemi, quello della filiera della conoscenza, appunto, con le politiche del libro, della lettura e dell'informazione, quello della ricerca, della produzione e della divulgazione scientifica, quello del *welfare* sociale e culturale, quello della formazione lungo tutto l'arco della vita e delle varie forme di *literacy*, quello di agenti per la costruzione di comunità informate e critiche e per la partecipazione democratica dei cittadini ai processi sociali e politici. Si tratta di ecosistemi dai confini labili, spesso intersecantisi fra loro, che fanno della biblioteca un unicum fra le altre organizzazioni di servizio. A dettare il prevalere di uno o dell'altro sono le comunità territoriali, o scientifiche, o di qualsivoglia altro genere, cui le biblioteche o le reti di biblioteche fanno riferimento, o con cui interagiscono, e da cui sono in definitiva legittimate. In tale contesto è parso di assistere ad uno svilimento delle politiche del libro e dell'informazione a favore delle altre funzioni, ma questo sembra a volte essere più una pretesa che un dato di realtà. Il policentrismo culturale italiano ha fatto sì che molte delle biblio-

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-697

teche considerate di base possiedano anche collezioni storiche e alimentino con il loro ulteriore sviluppo il sedimentarsi della memoria sociale e culturale delle comunità di riferimento, mentre, contemporaneamente, molte biblioteche storiche siano attive anche nei servizi di lettura pubblica, nella circolazione dei documenti e nell'accesso pubblico alle collezioni digitali, facendo riferimento a comunità più ampie di quelle territoriali. Un motivo in più per considerare obsolete tante distinzioni delle biblioteche per tipologie, che paiono più tipologie istituzionali che di servizio. La redazione di alcune eccellenti carte delle collezioni da parte delle biblioteche pubbliche di base,[1] che prevedono sistemi scientificamente accorti per il loro sviluppo e si presentano anche sotto forma di patto con le comunità di riferimento, stanno a confermare tali assunti.
La *Raccomandazione del Comitato dei Ministri agli Stati membri sulla legislazione e la politica delle biblioteche in Europa* (Comitato dei Ministri agli Stati membri sulla legislazione e la politica delle biblioteche in Europa 2023), adottata dal Consiglio d'Europa nel 2023, e la sua *Appendice*, delineano il quadro di sviluppo delle politiche bibliotecarie europee nel breve e medio periodo, riconoscendo, in particolare, il ruolo delle biblioteche come "istituzioni accessibili al pubblico di natura culturale, educativa e sociale" e come servizi di natura essenziale, anche in caso di minacce globali, ai fini dell'accesso libero e gratuito all'informazione. Profondamente radicate nelle politiche del libro e dell'informazione (Comitato dei Ministri agli Stati membri sulla legislazione e la politica delle biblioteche in Europa 2023, titolo 4, par. 10) (i termini collezione e collezioni compaiono nel testo di 16 pagine 19 volte),[2] esercitano un ruolo fondamentale per lo sviluppo di società democratiche e sono presidio della libertà di espressione e di pensiero. Comunque si considerino le collezioni bibliotecarie, da raccolte di oggetti musealizzati a infrastrutture virtuali per la ricerca a dimensione internazionale e cooperativa (VRE),[3] la loro centralità rimane dunque evidente.

Il Deposito legale

La medesima *Raccomandazione*, sopra richiamata (Comitato dei Ministri agli Stati membri sulla legislazione e la politica delle biblioteche in Europa 2023, par. 14), dedica un intero paragrafo al Deposito legale , "mezzo principale per creare le collezioni del patrimonio culturale nazionale", e ciò ci porta all'argomento centrale di questo contributo, focalizzato soprattutto sul deposito legale delle risorse digitali. E in questo caso l'idea di una collezione nazionale, cartacea e digitale, strumento strategico per la costruzione e la salvaguardia dell'eredità culturale delle nazioni, sembra universalmente accettata. Come si diceva, le singole biblioteche o reti di biblioteche possono privilegiare nelle loro politiche di servizio uno o l'altro degli ecosistemi sopra citati, ma la disponibilità di un'infrastruttura nazionale intesa come Collezione di ultima istanza resta imprescindibile per il perseguimento di qualunque loro obiettivo. Anche le biblioteche pubbliche che più si caratterizzano per i servizi legati al *welfare* hanno bisogno di un *background* di contenuti (risorse e strumenti di controllo bibliografico e di gestione) affidabili e sicuri.
Il Deposito legale in Italia è prescritto dalla Legge 106/2004 e dal Regolamento di attuazione DPR

[1] Si veda per tutte la *Carta delle collezioni delle biblioteche comunali pistoiesi*, Biblioteca San Giorgio, Biblioteca Forteguerriana https://www.sangiorgio.comune.pistoia.it/carta-delle-collezioni-3/.
[2] Diversa interpretazione in (Vitiello 2024).
[3] Si veda (Kempf 2013; Connaway et al., 2010).

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-697

252/2006, ed è previsto anche per le risorse digitali (documenti diffusi tramite rete informatica, nella lettera della legge). Il quadro normativo è però, si direbbe colpevolmente, incompleto: il deposito di tali risorse infatti attende ancora di essere definito (DPR 252, art. 37) con un ulteriore regolamento.

L'art. 1, comma 2 della Legge sopra citata si pone la finalità di costituire "l'archivio nazionale e regionale della produzione editoriale". La legge riprende un uso autorevole del termine 'archivio' collegato alle funzioni di una biblioteca nazionale così come proposto da Diego Maltese che in vari interventi parla di "archivio nazionale del libro", in grado di conservare per ogni documento "una globalità di riferimenti culturali" rendendone possibile una "'lettura' a tutti i livelli"(Maltese 1963, 87; 1977, 286; 1977, 290; 1979, 450).[4]

Si tratta di un uso analogico del termine 'archivio' che mette in primo piano quel "vincolo naturale, necessario e determinato" (Penzo Doria 2022, 164) che caratterizza la particolarità della raccolta di una biblioteca nazionale: in altre parole, la "produzione letteraria nazionale" ovvero tutto quello che in forma registrata (non solo – in senso stretto – scritta) documenta la "memoria della cultura e della vita sociale italiana" (art. 1, c. 1 Legge 106/2004). Maltese riprende e fa propria l'analogia proposta da Desiderio Chilovi (1903, 11): "una Biblioteca Nazionale Centrale [...] deve rappresentare, come in un archivio, tutto ciò che si pubblica". Una "sorprendente anticipazione" di questa analogia la possiamo trovare nel 1882 sempre a Firenze nella Biblioteca nazionale nella testimonianza di Domenico Comparetti durante la visita ispettiva della Commissione parlamentare "d'inchiesta sopra le biblioteche, le gallerie e i musei del Regno". Comparetti, sentito in qualità di utente della Biblioteca, dichiara: "Come c'è un archivio delle carte, vi deve essere pure un archivio della stampa. Questo è il carattere che dovrebbe darsi alla Biblioteca Nazionale" (Traniello 2000, 235).

È parimenti interessante notare come già nei dibattiti di una delle Commissioni Cibrario del 1869,[5] istituita dal Regno d'Italia il 20 luglio di quell'anno per occuparsi del riordino delle biblioteche del Regno, era presente l'idea di una raccolta di tutte le pubblicazioni a stampa presso la Nazionale di Firenze, istituita nel 1861, con finalità culturali e di messa a disposizione del pubblico. Questo si ipotizzava, però, in virtù della legge sulla proprietà letteraria, era cioè cosa diversa dall'idea di un controllo di natura giurisdizionale sulle pubblicazioni attuata tramite la legge sulla stampa, principio che in seguito prevarrà. Già Paolo Traniello (1995, 222)[6] aveva notato l'"originalità" e la "maggiore modernità" di questo approccio, che avrebbe consentito da subito il coinvolgimento degli editori "nell'azione pubblica di controllo". [7]

Si è ritenuto di ricordare questa vicenda perché in realtà l'idea di un deposito delle pubblicazioni come condizione per il riconoscimento della proprietà intellettuale si è successivamente sviluppata in parallelo all'evoluzione della normativa sul deposito legale come la conosciamo e sperimentiamo attualmente. L'art. 103 della Legge 633/1941 sulla Protezione del diritto d'autore ha infatti istituito

[4] Questo contributo usa prevalentemente il termine 'documento' (anziché il termine 'risorsa') sia per sottolineare la particolarità della collezione (o raccolta) del deposito legale, sia perché il termine è usato dalla L. 106/2004. Per il termine 'lettura' si veda più avanti il paragrafo *Nuove modalità di fruizione*.

[5] https://it.wikipedia.org/wiki/Commissione_Cibrario.

[6] Ringraziamo Claudio Leombroni per la segnalazione di questo articolo.

[7] Un sistema di questo tipo è vigente negli Stati Uniti.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-697

il Registro pubblico generale delle opere protette, di cui il successivo art. 105 disciplina il deposito presso il Ministero della cultura. L'attualità di questi temi è confermata da un recente provvedimento ministeriale che ha istituito il *Repertorio delle opere dei creatori digitali*, ai sensi dell'art. 27, comma 2 della Legge 206/2023, recante "Disposizioni organiche per la valorizzazione, la promozione e la tutela del made in Italy". Il MiC, con D.M. 3.10.2024, ha dunque introdotto anche per i "creatori digitali" il deposito delle proprie opere, da attuarsi sulla base di specifiche linee guida.[8]
Ci si trova dunque nella situazione paradossale per cui la normativa sul deposito legale, a quasi vent'anni dall'emanazione del Regolamento 252/2006, e dopo anni di sperimentazione volontaria su piattaforme gestite dalla BNCF, e con procedure ottimizzate, non prevede ancora il deposito delle risorse digitali,[9] mentre un analogo deposito è previsto dalla normativa sulla tutela del made in Italy su una piattaforma di fatto inesistente, e con una procedura ad alto tasso di adempimenti burocratici,[10] che non sembra porsi obiettivi di conformità allo standard OAIS.[11]
Il ritardo nel perfezionamento della normativa sul deposito legale ha comportato fin qui il mancato deposito, ovvero la perdita, di vent'anni di produzione culturale digitale.

La cooperazione

Vale la pena di richiamare alcuni punti, che trovano fondamento nella *Dichiarazione IFLA sul Deposito legale*[12] del 2011 e, per quanto riguarda le risorse digitali, nella *Carta sulla conservazione del patrimonio digitale*, adottata dalla Conferenza generale dell'UNESCO il 17 ottobre 2003[13]

- A parte alcune eccezioni, come l'Olanda, il deposito legale è prescritto dalle leggi nazionali. Le biblioteche depositarie operano quindi sulla base di un mandato cogente ed esclusivo, non vi è in questo caso concorrenza con altri agenti.
- Il deposito legale implica necessariamente e conferma il principio della cooperazione fra tutti gli agenti della filiera della produzione, gestione e disseminazione della conoscenza coinvolti. La normativa italiana prevede esplicitamente la cooperazione fra i soggetti obbligati,[14] le biblioteche nazionali centrali, titolari prime del deposito, e le biblioteche titolari del deposito dei documenti della produzione culturale regionale, individuate con i Decreti Ministeriali 28.12.2007 e 10.12.2009.

Con riferimento al secondo punto, la cooperazione implica che tutti gli agenti della filiera devono trovare, pur all'interno degli obblighi di legge o delle prassi consolidate del deposito legale, il

[8] https://biblioteche.cultura.gov.it/it/documenti/2024/Guida_depositoRPGenerale-opere_protette_19_02_25.pdf.
[9] Si vedano al riguardo le considerazioni di (Storti 2024, 25, 28-30).
[10] Compreso il versamento di una marca da bollo di 16 € per ogni deposito.
[11] *Open Archival Information System.* ISO 14721:2025 https://www.iso.org/standard/87471.html.
[12] https://www.ifla.org/publications/ifla-statement-on-legal-deposit-2011/.
[13] https://www.unesco.it/wp-content/uploads/pdf/UploadCKEditor/carta_UNESCO_it.pdf.
[14] Legge 106/ 2004, art. 3:
1. I soggetti obbligati al deposito legale sono:
a) l'editore o comunque il responsabile della pubblicazione, sia persona fisica che giuridica;
b) il tipografo, ove manchi l'editore;
c) il produttore o il distributore di documenti non librari o di prodotti editoriali similari;
d) il Ministero per i beni e le attività culturali, nonché il produttore di opere filmiche.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-697

proprio vantaggio, con il minor pregiudizio economico possibile per i soggetti obbligati, il pieno adempimento delle finalità generali delle istituzioni depositarie, la massima tutela possibile dell'interesse dei cittadini ad un accesso pieno e libero alla conoscenza ed alle risorse informative e culturali. In questo senso la *Dichiarazione IFLA* sopra richiamata sottolinea come le istituzioni depositarie siano tenute ad una gestione responsabile e trasparente delle risorse depositate, i soggetti obbligati non debbano essere tenuti a procedure di deposito irragionevoli o troppo complicate, e che le risorse depositate siano rese accessibili al pubblico senza irragionevole pregiudizio degli interessi dei detentori dei diritti sulle risorse medesime.

Il 20 novembre 2024 la sezione Biblioteche Nazionali dell'IFLA e la Biblioteca Nazionale di Singapore hanno organizzato un webinar sul deposito legale[15] con centinaia di partecipanti da tutto il mondo, anche in vista del prossimo aggiornamento delle linee guida IFLA sul deposito legale nell'era digitale. Non potendo in questa sede dilungarsi troppo sulle esperienze presentate al seminario, si può tentare di ricavarne alcune indicazioni utili per capire come nei diversi contesti nazionali si sia cercato fin qui di dare attuazione ai principi generali sopra elencati, vissuti spesso come contraddittori nelle prassi normative e organizzative delle singole realtà nazionali.

Una delle esperienze più interessanti è quella dell'Estonia, la cui legislazione prevede dal 2017[16] il deposito legale anche dei file da cui si ricavano le pubblicazioni a stampa, nonché di quelle che sono definite *web publications*,[17] con una definizione peraltro piuttosto generica. Kairi Felt, Responsabile del Dipartimento per lo sviluppo delle collezioni della Biblioteca nazionale estone, ha illustrato il modello organizzativo del deposito legale estone in termini di agenzia di servizi: in un portale apposito gli editori elettronici possono effettuare l'upload di e-book e dei file delle pubblicazioni pronte per la stampa, richiedere i numeri standard ISBN, ISSN, ISMN, scaricare i codici a barre EAN in formato vettoriale, gestire diritti ed eventuali restrizioni all'accesso pubblico, che di norma avviene da postazioni specifiche site in 5 biblioteche di deposito (con esclusione di copia e download), recuperare in ogni momento i file precedentemente caricati. Gli editori vengono così liberati dagli oneri di gestione di un magazzino digitale e dalla complessità e dai costi della conservazione digitale di lungo periodo.

L'intervento di Sabine Springer e Christoph Wohlstein, della Deutsche Nationalbibliothek (DNB) ha evidenziato come la normativa sul deposito legale abbia difficoltà a riflettere i cambiamenti e la globalizzazione dei mercati editoriali, in cui le piattaforme internazionali stanno rimpiazzando i mercati nazionali mettendo in crisi il principio della sede dell'editore come criterio per individuare il soggetto obbligato al deposito, mentre le piattaforme di *streaming* mettono sul mercato media compositi e interattivi, che si aggiornano in continuazione, e che combinano testi, musica, *podcast* e altri contenuti che è difficile far ricadere nelle fattispecie previste dalla normativa come oggetto del deposito legale. Una parte consistente dell'eredità culturale nazionale sfugge agli attuali strumenti di controllo, conservazione, garanzia di accessibilità nel lungo periodo, rischiando di mettere in crisi l'istituto stesso del deposito.

[15] https://www.ifla.org/news/ifla-national-libraries-section-organises-webinar-on-legal-deposit-a-strategic-tool-for-building-tomorrows-heritage/.

[16] *Legal Deposit Act* https://www.riigiteataja.ee/en/eli/514092016001/consolide.

[17] *Legal Deposit Act*, capitolo 1, § 3, 4.1: "web publication means a collection of identifiable information that has been made publicly available on the Internet".

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-697

Anche dall'esperienza inglese, illustrata da Linda Arnold-Stratford, Responsabile della struttura di collegamento fra le 7 biblioteche inglesi di deposito legale, emerge la centralità di una relazione positiva con l'industria editoriale, basata su piattaforme tecnologiche condivise che riducono il carico di adempimenti per gli editori in fase di deposito, e prevede il loro coinvolgimento negli organi di governo e gestione del deposito legale, insieme ai dirigenti delle biblioteche e a rappresentanti governativi (come in realtà si era iniziato a fare in Italia già a metà degli anni 2000). Il modello fortemente cooperativo del sistema di deposito legale inglese[18] ha consentito di limitare i danni del grave attacco informatico che ha messo fuori uso i sistemi della British Library il 28 ottobre del 2023. Il principale nodo di conservazione, presso la Biblioteca Nazionale di Scozia, è stato isolato nel momento in cui ci si è resi conto della gravità dell'attacco, e successivamente replicato sull'infrastruttura commerciale di *storage* di Amazon Web Services. Ma tutti i sistemi di *harvesting,* archiviazione e accesso alle risorse digitali, cioè tutto l'ambiente di gestione del deposito legale, sono finiti fuori uso e lo erano ancora dopo quasi un anno, perché basati su software non più supportati che sarebbe stato rischioso ripristinare.

Nel complesso, il webinar ha evidenziato alcuni punti nodali nel dibattito contemporaneo sul deposito legale, ma si tratta di temi da affrontare comunque con pragmatismo, e tenendo conto che molte opinioni convergono sull'idea della necessaria selettività del deposito legale.[19]

Tornando alla situazione in Italia, è opportuno mettere in luce qualche altro aspetto, con riferimento sia al deposito delle risorse digitali e al *web archiving*, che al rapporto fra archivi nazionali e regionali della produzione editoriale.[20] Innanzitutto, se per le risorse digitali più tradizionali, come e-book e periodici elettronici, è ancora possibile basarsi sulla sede legale del soggetto obbligato per individuare le porzioni della produzione editoriale di competenza delle biblioteche depositarie regionali, la situazione è più confusa per i siti web e per le risorse digitali composite, ibride o multiformi (come evidenziato sopra dall'intervento di Springer e Wohlstein).[21]

In secondo luogo la normativa vigente lascia qualche dubbio in merito alla effettiva responsabilità delle regioni nella gestione di un archivio di deposito regionale per i documenti diffusi tramite rete informatica.

In ogni caso la natura dei documenti depositati e considerazioni di efficienza richiederebbero soluzioni basate sulla cooperazione interistituzionale (Storti 2023, 48). La gestione del deposito legale dovrebbe essere organizzata in modo unitario, valorizzando il contributo sia del livello nazionale sia di quello regionale. A livello regionale la vicinanza delle istituzioni con i territori è importante nella fase di acquisizione e nella fase di fruizione. In fase di acquisizione occorre considerare che, soprattutto per determinate categorie di documenti (si pensi ad esempio al *web archiving* o al *social*

[18] Il sistema distribuito è composto da 7 biblioteche: British Library St. Pancras, National Library of Wales, British Library Boston Spa, National Library of Scotland, Bodleian Library Oxford, University Library Cambridge, Trinity College Dublin; le prime quattro sono nodi del Digital Library Store (DLS), il Magazzino digitale, fra queste il nodo scozzese replica i magazzini di tutti gli altri nodi.

[19] Così già (Mandillo 2000; Vitiello 2007).

[20] Sulle relazioni fra archivi nazionali e regionali (e su molti altri aspetti della normativa) restano imprescindibili gli approfondimenti di (Puglisi 2007; 2020).

[21] È bene comunque specificare che non tutte le risorse digitali vanno depositate nelle biblioteche, oggetto del presente contributo. I documenti sonori e video, la grafica d'arte, le fotografie e i video d'artista sono destinati ad altri istituti, individuati dagli art. 14 e 20 del DPR 252/2006.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-697

archiving) il livello regionale può contribuire in maniera significativa nell'applicazione dei criteri di selezione.[22] In fase di fruizione occorre ricordare che l'art. 38 del DPR 252/2006 differenzia le condizioni di fruibilità sulla base dello status di origine dei documenti oggetto di deposito legale digitale:

- il comma 1 prevede che le istituzioni depositare possano dare accesso in rete ai documenti appartenenti alla raccolta solo se quei documenti sono "in origine accessibili liberamente in rete";
- il comma 2 prevede un accesso limitato "a utenti registrati che accedono da postazioni situate all'interno degli istituti depositari" per documenti in "origine accessibili a determinate condizioni, quali licenze o altri contratti attributivi del diritto all'accesso e all'utilizzazione del documento".

In questo contesto, le biblioteche depositarie regionali rivestono un ruolo cruciale. Poiché l'accesso al deposito legale digitale richiede anche, nonostante la natura digitale dei documenti, punti di accesso fisicamente determinati, è importante che questi punti siano vicini il più possibile agli utenti.

La condivisione di una piattaforma tecnologica distribuita e basata su cloud è pienamente compatibile con una gestione differenziata degli archivi digitali, distinguendo le responsabilità a livello nazionale e regionale. Questo approccio consente di definire i confini degli archivi regionali come sottoinsiemi logici dell'archivio nazionale, rendendoli più flessibili e adattabili alla natura dei documenti digitali.

La conservazione

La necessità di evitare che funzioni strategiche come il deposito legale dipendano da sistemi *legacy* è un requisito ben noto nella letteratura sulla conservazione digitale, ma quanto si è appreso dall'esperienza (peraltro non unica) della British Library[23] va ben oltre il rispetto dei requisiti o la messa in atto di nuove soluzioni tecnico-organizzative, come la separazione fisica delle collezioni digitali di deposito legale, e del loro ambiente di gestione, dal resto delle collezioni digitali della Biblioteca. Quell'esperienza ha aperto un capitolo nuovo nella storia oramai non breve della conservazione digitale: fermi restando i suoi tradizionali obiettivi, ovvero garantire attraverso sistemi certificabili il mantenimento nel lungo periodo di una sequenza di bit, la sua leggibilità da parte di una macchina, l'accessibilità permanente ai contenuti espressi da quella sequenza di bit (anche attraverso le necessarie procedure di migrazione o emulazione), il mantenimento dell'autenticità, cioè dell'identità e dell'integrità, degli oggetti digitali che quella sequenza esprimono, e la loro disponibilità permanente anche in quanto dati per qualunque uso consentito, la loro attuazione richiede un ripensamento all'interno di una prospettiva più ampia di *cybersecurity*. Anche lo stesso sistema distribuito di archivi digitali affidabili perché certificati, ai quali sia stato formalmente

[22] L'art. 37 comma 3 del DPR 252 2006 prevede i criteri di selezione nella fase di sperimentazione del deposito legale digitale. La normativa da emanare per l'avvio del deposito legale dovrebbe prendere in conto anche questo tema.

[23] Anche organizzazioni non governative, che si ritenevano meno appetibili per il crimine informatico, come Internet Archive, hanno subito attacchi nell'ottobre 2024, con interruzione dei servizi (DDoS, Denial of services) e violazione delle credenziali di 31 milioni di utenti.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-697

attribuito il diritto-dovere di conservare e mantenere l'accessibilità degli oggetti digitali nel lungo periodo, che in Italia non è ancora stato possibile realizzare, dovrà prendere in conto i rischi crescenti legati al crimine informatico.[24]

Paradossalmente, la pervasività della trasformazione digitale, i sistemi cooperativi e fortemente distribuiti, come quelli qui auspicati per il deposito legale, le loro crescenti interconnessioni e interdipendenze aumentano le "superfici di attacco"[25] da parte del crimine informatico. E nelle stesse condizioni si trovano anche i grandi data center, la cui complessità strutturale e gestionale e la molteplicità dei possibili accessi aumentano parimenti la possibilità di falle nei sistemi di sicurezza.

In tale contesto, sul sistema del deposito legale, che ha il compito di "conservare la memoria della cultura e della vita sociale italiana"[26] ricadono crescenti responsabilità, che pare velleitario lasciare sulle sole spalle delle istituzioni depositarie, fra l'altro, come previsto, a costo zero. L'alterazione dei contenuti culturali, rispondente ad interessi immediati di natura economica o anche geo-politica, costituisce infatti un vulnus tale da intaccare credibilità e reputazione delle istituzioni culturali e scientifiche, e delle biblioteche digitali, in maniera grave e con conseguenze imprevedibili, si pensi al sistema delle citazioni ed alla valutazione della ricerca scientifica, ma anche alle distorsioni degli eventi di cronaca e dell'attualità politica. Di qui anche la centralità del *web archiving* e della conservazione della produzione culturale digitale nella sua integrità e autenticità.

Anche le politiche per l'accesso aperto possono diffondere agevolmente documenti, e in generale contenuti e risorse, che siano stati malevolmente alterati, pregiudicando la corretta conservazione e trasmissione della memoria delle comunità. Non sembra ancora esserci sufficiente coscienza che l'ecosistema dell'Eredità culturale digitale costituisce un'"infrastruttura culturale critica ibrida",[27] cioè composta da istituzioni (GLAM/MAB) e infrastrutture tecnologiche, risorse digitali, processi, comunità sociali (professionisti, agenti software e utenti) che con quelle risorse interagiscono, a formare appunto un ecosistema sociale e tecnologico complesso. Tale infrastruttura va annoverata fra quelle critiche, e come tale andrebbe riconosciuta anche a livello normativo, al pari di quelle dei sistemi di pagamento, dei sistemi di trasporto pubblico, dei sistemi energetici. Con in più il fatto che una sua alterazione può intaccare, con la memoria, l'identità stessa delle comunità.

L'esperienza della British Library prima richiamata evidenzia la necessità di politiche di *cybersecurity* che valutino attentamente le architetture disponibili, sempre tenendo conto che la conservazione e la sicurezza sono processi, e che nessuna tecnologia di per sé è in grado di garantirle. Sebbene il cloud centralizzato in grandi data center sia spesso considerato l'unica soluzione praticabile, esistono modelli decentralizzati che rappresentano alternative valide o complementari. Un possibile scenario per il deposito legale digitale potrebbe essere l'adozione di una rete privata IPFS,[28] in grado di garantire un'infrastruttura distribuita per la conservazione a lungo termine.

[24] Sullo scenario della *cybersecurity* per le istituzioni del patrimonio culturale si veda (Tammaro e Bellini 2024).

[25] https://en.wikipedia.org/wiki/Attack_surface.

[26] Legge 106/2004, art. 1, comma 1.

[27] Così Emanuele Bellini, nel webinar *Conservazione e sicurezza dei beni culturali digitali* per la Fondazione Scuola del Patrimonio, 28 gennaio 2025.

[28] Per un abbozzo di descrizione di IPFS si veda: https://it.wikipedia.org/wiki/InterPlanetary_File_System, e il sito ufficiale https://ipfs.tech/. Per un caso d'uso si veda: https://observer.com/2017/05/turkey-wikipedia-ipfs/.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-697

L'IPFS è intrinsecamente progettato per offrire un sistema di archiviazione distribuito e indirizzabile tramite un identificatore basato sui contenuti che gestisce: in questo modo, file e metadati associati possono essere replicati, verificati e recuperati indipendentemente da qualsiasi singolo punto di guasto. Ad esempio una modifica malevola di un file di un nodo non ha nessun impatto sulle repliche di quel file presenti in altri nodi (Bergamin e Gobbo 2025, [6]). Occorre inoltre considerare l'avvio del Polo Strategico Nazionale (PSN),[29] che mira a mettere in sicurezza dati e applicazioni della Pubblica Amministrazione italiana attraverso un cloud basato su quattro grandi data center (due nel Lazio e due in Lombardia). In un quadro ancora in definizione, il PSN potrebbe svolgere un ruolo chiave come dark archive[30] e per l'hosting di servizi avanzati come LLM e RAG, descritti di seguito.

Il deposito legale, strumento costitutivo dell'eredità sociale e culturale e dell'identità stessa delle comunità, risponde quindi ad esigenze profonde della società, che richiedono non solo la garanzia della continuità di servizi da ritenersi essenziali, ma l'adozione di tecnologie proattive in grado di individuare, con le opportune azioni di monitoraggio, le possibili vulnerabilità, con tecniche già sperimentate in altri domini similmente critici. Nonostante la locuzione "cambio di paradigma" sia spesso abusata, parrebbe lecito usarla in questo caso: i progetti, gli strumenti, le tecniche, i modelli organizzativi della conservazione digitale devono sviluppare nativamente capacità di resilienza tali da essere in grado di mantenere il più possibile la propria identità e integrità di fronte a eventi distruttivi esterni.

Lo scenario del deposito legale interseca, e interroga, quindi altri scenari, più ampi e articolati.

Nuove modalità di fruizione

Il deposito legale genera un corpus completo e sistematico della produzione culturale di un determinato Paese, che si accumula e si stratifica nel tempo. Il risultato è una raccolta unica e rappresentativa, che documenta in modo organizzato e continuo l'evoluzione della cultura, della conoscenza e della creatività. In questa unicità si trova l'opportunità di una fruizione che può essere su almeno due livelli (qui Maltese usa la metafora della 'lettura'): un livello orientato ai contenuti della singola risorsa bibliografica, e un livello orientato a insiemi di documenti legati fra loro da un "vincolo naturale, necessario e determinato" ovvero portatori di una "globalità di riferimenti culturali". Tra le ricerche possibili al secondo livello, è inclusa la possibilità di effettuare sul corpus creato attraverso il deposito legale ricerche di tipo longitudinale[31] (ad esempio, studiando l'evoluzione di un fenomeno nel tempo attraverso i documenti del corpus) o di tipo trasversale[32] (ad esempio, analizzando lo stesso fenomeno in un determinato momento su diversi documenti del corpus).

Con il deposito legale delle pubblicazioni digitali e con la possibilità di accesso al full text emerge la necessità di interrogarsi sul ruolo di mediazione che le biblioteche possono e devono offrire. Una domanda cruciale è: esistono vincoli giuridici nella legislazione italiana per l'uso del full text delle pubblicazioni digitali oggetto di deposito legale? L'art. 3 della Direttiva UE

[29] https://www.polostrategiconazionale.it/.
[30] https://dictionary.archivists.org/entry/dark-archives.html.
[31] https://en.wikipedia.org/wiki/Longitudinal_study.
[32] https://en.wikipedia.org/wiki/Cross-sectional_study.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-697

2019/790 (DSM)[33] prevede un'eccezione obbligatoria al diritto d'autore per la "estrazione, per scopi di ricerca scientifica, di testo e di dati" (così viene tradotto il *text and data mining* (TDM)[34] "effettuate da organismi di ricerca e istituti di tutela del patrimonio culturale [...] da opere o altri materiali cui essi hanno legalmente accesso". In Italia il recepimento della Direttiva DSM ha modificato la legge sul diritto d'autore (L. 633/1941) con l'introduzione dell'art. 70-ter che nei primi 3 commi regola la possibilità di effettuare il TDM per scopi di ricerca scientifica anche per le raccolte delle biblioteche. Il comma 2 definisce il TDM in modo ampio, includendo implicitamente tecniche di intelligenza artificiale (IA), incluso l'uso o l'addestramento di modelli linguistici di grandi dimensioni (LLM)[35] o lo sviluppo di sistemi di Retrieval-Augmented Generation (RAG).[36] Tuttavia – viene prescritto al comma 1 – è fondamentale rispettare le limitazioni del medesimo comma: i risultati del TDM possono essere comunicati al pubblico solo se trasformati in opere originali che non riproducano direttamente contenuti protetti da diritto d'autore. In generale, l'uso delle tecnologie legate all'IA deve essere visto come un nuovo tipo di tecnologia culturale e sociale (Farrell et al. 2025) che permette agli esseri umani di riusare in modo più efficiente la conoscenza accumulata da altri esseri umani, ma che non deve dimenticare l'etica della ricerca (IFLA FAIFE 2020).

Come generalmente riconosciuto l'IA in biblioteca può avere almeno "due possibili campi di applicazione": "assistere l'utente nella ricerca e il bibliotecario nella catalogazione e nell'indicizzazione per soggetto" (Cheti 2025, 347). In questo momento nel campo dell'IA stiamo assistendo a profondi cambiamenti: l'emergere dell'IA generativa sta prendendo sempre più spazio nelle nostre modalità di accesso al sapere e alla conoscenza ed è profondamente differente dall'IA tradizionale (Storey et al. 2025, [9]). In estrema sintesi: l'IA tradizionale (predittiva o discriminativa) è progettata per analizzare dati esistenti e fornire previsioni, classificazioni o decisioni basate su schemi appresi. Non crea nuovi contenuti, ma interpreta e categorizza quelli che già ha; l'IA generativa è progettata per creare (generare) contenuti nuovi e originali che non esistevano prima, basandosi sui pattern e sulle relazioni che ha appreso in fase di addestramento da grandi quantità di dati esistenti. Occorre aggiungere che l'IA generativa può essere usata anche in compiti fino ad oggi affidati all'IA tradizionale come ad esempio l'utilizzo di modelli generativi per la classificazione di documenti (Chae e Davidson 2025).

Se diamo un rapido sguardo alle iniziative in corso integrate nei servizi di fornitura di contenuti e nei sistemi di gestione delle biblioteche offerti da fornitori come Clarivate (ProQuest, Ex Libris), Elsevier ed EBSCO troviamo applicazioni che fanno uso dell'IA generativa e che offrono strumenti principalmente in tre aree: assistenza nella creazione delle query di ricerca; creazione di riassunti di singoli documenti o di insiemi di documenti risultato di una ricerca; creazione e arricchimento dei metadati (Taylor et al. 2025). Senza entrare in dettagli che non sono funzionali al tema di questa ricerca, la letteratura professionale consultata (Breeding 2025; Taylor et al. 2025)

[33] Recepita e pubblicata in GU il 27.11.2021, DL 8 novembre 2021, n. 177. *Attuazione della direttiva (UE) 2019/790 del Parlamento europeo e del Consiglio, del 17 aprile 2019, sul diritto d'autore e sui diritti connessi nel mercato unico digitale e che modifica le direttive 96/9/CE e 2001/29/CE.*

[34] https://en.wikipedia.org/wiki/Text_mining.

[35] https://it.wikipedia.org/wiki/Modello_linguistico_di_grandi_dimensioni. Una raccolta di risorse sui LLM: https://github.com/Hannibal046/Awesome-LLM.

[36] https://it.wikipedia.org/wiki/Retrieval_augmented_generation.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-697

presenta un quadro con molte iniziative, con applicazioni generalmente offerte in versione beta e dove talvolta l'acronimo IA è usato dai produttori con funzioni di marketing più che per descrivere una determinata tecnologia (Taylor et al. 2025).

Sul versante delle biblioteche nazionali possiamo trovare un'iniziativa che introduce un cambiamento reale nei modi di gestione delle raccolte del deposito legale. È collegata alla diffusione di Annif (Suominen, Inkinen, e Lehtinen 2022) che si autodefinisce strumento open source di indicizzazione automatizzata[37] per soggetto, basato su più algoritmi (modelli) e pensato per biblioteche, archivi e musei.[38] Si tratta di una piattaforma, principalmente basata su metodi dell'IA tradizionale, che offre anche l'uso di *ensemble* ovvero di una tecnica che combina le previsioni di più modelli (algoritmi) di *machine learning* per ottenere una previsione complessiva più robusta e accurata rispetto a quella di un singolo modello. Questa piattaforma, nata nel 2017 per la Biblioteca nazionale finlandese, si è diffusa progressivamente a livello internazionale.[39] In estrema sintesi l'obiettivo di Annif è quello di mettere a confronto le parole chiave estratte dal testo di un determinato documento con un vocabolario controllato per individuare quelle più efficienti ai fini dell'indicizzazione per soggetto di quel documento. Recentemente tra i modelli che possono far parte di un *ensemble* possiamo trovare a titolo sperimentale Xtransformer, ovvero un modello che si avvale di algoritmi utilizzati dall'IA generativa. Anche questo modello nella fase di addestramento riceve come input il testo dei documenti e le parole chiave assegnate manualmente e impara le relazioni che gli consentiranno di suggerire le parole chiave appropriate in fase di produzione. A differenza degli altri modelli, però, Xtransformer non mette a confronto direttamente gli elementi del testo (del documento, delle parole chiave, del vocabolario controllato) ma lavora con le loro rappresentazioni numeriche vettoriali (*embedding*[40]) basate su modelli di *transformer*[41] che, come è noto, sono alla base dei moderni LLM.[42]

Tra gli utilizzatori di Annif troviamo la Deutsche Nationalbibliothek (DNB) che dal 2022 utilizza questa piattaforma per l'indicizzazione per soggetto (GND)[43] e per la classificazione (DDC)[44] delle risorse bibliografiche digitali pervenute per deposito legale.[45] Non si tratta di un uso sperimentale, ma di un uso in esercizio che coinvolge un numero davvero ragguardevole di pubblicazioni.[46] Tenendo conto che il deposito legale digitale è descritto nella Bibliografia nazionale tedesca (Serie O dedicata agli ebook e articoli di riviste accademiche) si può constatare che nel catalogo generale

[37] https://www.treccani.it/enciclopedia/participio_(Enciclopedia-dell'Italiano)/ (per l'uso in funzione aggettivale del participio passato).

[38] "Annif is a multi-algorithm automated subject indexing tool for libraries, archives and museums": https://github.com/NatLibFi/Annif.

[39] https://annif.org/.

[40] https://it.wikipedia.org/wiki/Word_embedding.

[41] https://en.wikipedia.org/wiki/Transformer_(deep_learning_architecture).

[42] I dettagli del nuovo modello di ML Annif - Xtransformer - sono reperibili in (Suominen, Inkinen, e Lehtinen 2025).

[43] https://www.dnb.de/EN/Professionell/Standardisierung/GND/gnd_node.html.

[44] Per una ricerca sull'uso di Annif per l'assegnazione automatizzata della Classificazione Decimale Dewey si veda (Golub et al. 2024).

[45] https://www.dnb.de/EN/Professionell/ProjekteKooperationen/Projekte/KI/ki_node.html.

[46] Dal 2022 la Biblioteca Nazionale Centrale di Firenze (BNCF) sperimenta Annif: https://github.com/Bncf/Annif-Bncf/wiki/Il-progetto-Annif%E2%80%90BNCF. Per l'uso degli strumenti IA in BNCF si veda (Storti 2024). Per informazioni sull'avvio della sperimentazione nel 2011: (Bergamin e Lucarelli 2012).

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-697

della biblioteca il numero di record bibliografici a oggi prodotti per questa Serie è intorno ai 4 milioni e mezzo.[47]

Il compito del deposito legale digitale è stato assegnato alla DNB nel 2006[48] e applicato a regime dal 2010 quando la DNB ha deciso che per questa tipologia di pubblicazioni non si sarebbe più applicata la catalogazione 'uno a uno' o 'libro alla mano', ma una catalogazione automatizzata di tipo incrementale[49]. Alla catalogazione descrittiva fornita dall'editore (con un controllo di autorità automatizzato che fa uso del GND per i nomi di persona) viene aggiunta l'indicizzazione di tipo semantico (soggetto e classe) grazie a un sistema automatizzato (dal 2012 la piattaforma proprietaria Averbis e – come precedentemente riportato – dal 2022 Annif). Il record MARC21 che veicola queste informazioni bibliografiche contiene l'indicazione dell'origine automatizzata dei campi relativi alla soggettazione e alla classificazione. Dal 2012 MARC21 fornisce il campo 883 denominato *Metadata provenance* per ospitare informazioni sulle modalità di creazione dei metadati: è così possibile dichiarare se sono completamente o parzialmente generati dalla macchina, quale software di indicizzazione è stato usato, oltre che riportare il valore di *confidenza*[50] o di attendibilità che lo stesso algoritmo 'stima' di avere raggiunto. Questo nuovo flusso di lavoro viene completato dal controllo di qualità a campione effettuato da catalogatori della DNB.[51]

Oltre all'utilizzo a regime di Annif la DNB sta sperimentando anche le potenzialità dell'IA generativa per fare sostanzialmente lo stesso lavoro di Annif: assegnare parole chiave a documenti utilizzando un determinato vocabolario controllato. Il DNB-AI-Project[52] ha presentato al Workshop SemEval 2025,[53] organizzato dalla Technische Informationsbibliothek (TIB),[54] un prototipo per l'indicizzazione per soggetto automatizzata utilizzando un *ensemble* di differenti LLM (Kluge e Kähler 2025). Il Workshop prevedeva una gara di indicizzazione con la fornitura ai partecipanti di due insiemi di documenti: il primo insieme di addestramento completato con le parole chiave ricavate da un vocabolario controllato (GND) e assegnate manualmente dai catalogatori della TIB; il secondo con un un insieme di documenti da indicizzare con lo stesso vocabolario (GND). I risultati sono incoraggianti: "Our system is fourth in the quantitative ranking [...], but achieves

[47] Alla data del 18 luglio 2025: 4.379.606 per la Serie O. La query che riporta questo risultato: https://portal.dnb.de/opac/simpleSearch?query=cod%3D%22ro%22+&cqlMode=true.

[48] https://www.dnb.de/EN/Ueber-uns/Geschichte/geschichte_node.html: "On 29 June, the 'Law Regarding the German National Library' comes into force. Its key amendment is the expansion of the collection mandate to include online publications. The law also renames the library with its sites in Leipzig, Frankfurt am Main and Berlin as the Deutsche Nationalbibliothek, the German National Library".

[49] "The possibilities for evaluating digital content and web-based cooperation, as well as the development of technical procedures for interlinking content and establishing thematic relationships have paved the way for rethinking the library indexing methods. Cataloguing no longer has to be viewed as a one-off operation, but instead can be considered as a cyclical process in which cataloguing data are repeatedly changed and updated" (DNB 2017).

[50] https://it.wikipedia.org/wiki/Intervallo_di_confidenza.

[51] Informazioni più complete sul trattamento in DNB delle pubblicazioni digitali in (Gömpel, Junger, e Niggemann 2010); su Averbis e su Annif in (Mödden 2022). In questo ultimo lavoro viene riportata anche la percentuale media dei record bibliografici sottoposti a controllo di qualità dal 2017 al 2021 è stata del 18% (Mödden 2022, 262).

[52] https://www.dnb.de/EN/Professionell/ProjekteKooperationen/Projekte/KI/ki_node.html.

[53] *The 19th International Workshop on Semantic Evaluation* (https://semeval.github.io/).

[54] https://www.tib.eu/en/tib/profile/foundation "The Technische Informationsbibliothek – German National Library of Science and Technology (TIB) became an independent public-law foundation of the Federal State of Lower Saxony in January 2016".

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-697

the best result in the qualitative ranking conducted by subject indexing experts" (Kluge e Kähler 2025); l'utilizzo degli LLM riduce notevolmente i costi dell'addestramento (basta inviare al LLM pochi esempi di documenti con indicizzazioni assegnate manualmente, con la tecnica del *few-shot prompting*[55]). Un aspetto problematico è però dato dagli alti costi di elaborazione: "Particularly larger models use up an enormous amount of GPU-resources that may be infeasible in productive settings" (Kluge e Kähler 2025).
Se torniamo al TDM e alle nuove possibilità di accesso al full text occorre dire che – almeno nella letteratura scientifica – non troviamo per quanto riguarda il deposito legale un panorama di iniziative analogo a quello dell'indicizzazione automatizzata. Una pubblicazione di riferimento nel campo della *digital preservation* e che affronta questo tema è sicuramente *Computational Access: A beginner's guide for digital preservation practitioners* (DPC 2022):

> The term computational access relates to the ability to enable users to access collections within a digital preservation repository [...] in order to analyze, interrogate, or extract new meaning from that material (through, e.g., data or text mining, machine learning).

Non vi sono però resoconti di iniziative in corso sul *computational access* nel contesto di una raccolta di deposito legale. Oggi gli strumenti che potrebbero far parte del *computational access* possono essere individuati negli LLM e nei RAG. Un LLM addestrato sul corpus del deposito legale sarebbe senza dubbio uno strumento con rilevanti potenzialità. Tuttavia, è fondamentale considerare gli alti costi di sviluppo e il fatto che, per sua natura, un LLM non è progettato per collegare i risultati di una ricerca direttamente ai documenti specifici da cui ha appreso (Roncaglia 2023, 97). Un percorso più efficiente e anche più realistico dal punto di vista dei costi potrebbe essere il modello di *computational access* schematizzato in Figura 1. Con la consapevolezza che particolarmente in questo campo le tecnologie si stanno muovendo molto rapidamente, il modello presentato propone una soluzione già sperimentata (con molte varianti) in altri settori (Fan et al. 2024), ovvero l'uso di un sistema RAG.[56] L'obiettivo qui non è quello di esaminare tutti i dettagli tecnologici, ma quello di mettere in evidenza gli elementi strategici che possono aiutare nella valutazione di questo strumento. Quello che in ogni caso è importante (e irrinunciabile) sono gli obiettivi del modello qui proposto:

- garantire una assoluta trasparenza e verificabilità tra il risultato di una ricerca e i documenti che fanno parte della raccolta del deposito legale digitale;
- facilitare la fruizione della raccolta del deposito legale soprattutto per quanto riguarda il 'vincolo' - sopra richiamato - che tiene assieme i documenti che ne fanno parte.

[55] https://en.wikipedia.org/wiki/Large_language_model#Extensibility.
[56] Per l'origine del RAG: (Lewis et al. 2020); due lavori che fanno il punto sul RAG: (Gao et al. 2024; Zhao et al. 2024); RAG nel contesto dei nuovi strumenti di fruizione: (Di Marcantonio 2025); una raccolta di risorse sul RAG: https://github.com/Danielskry/Awesome-RAG.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-697

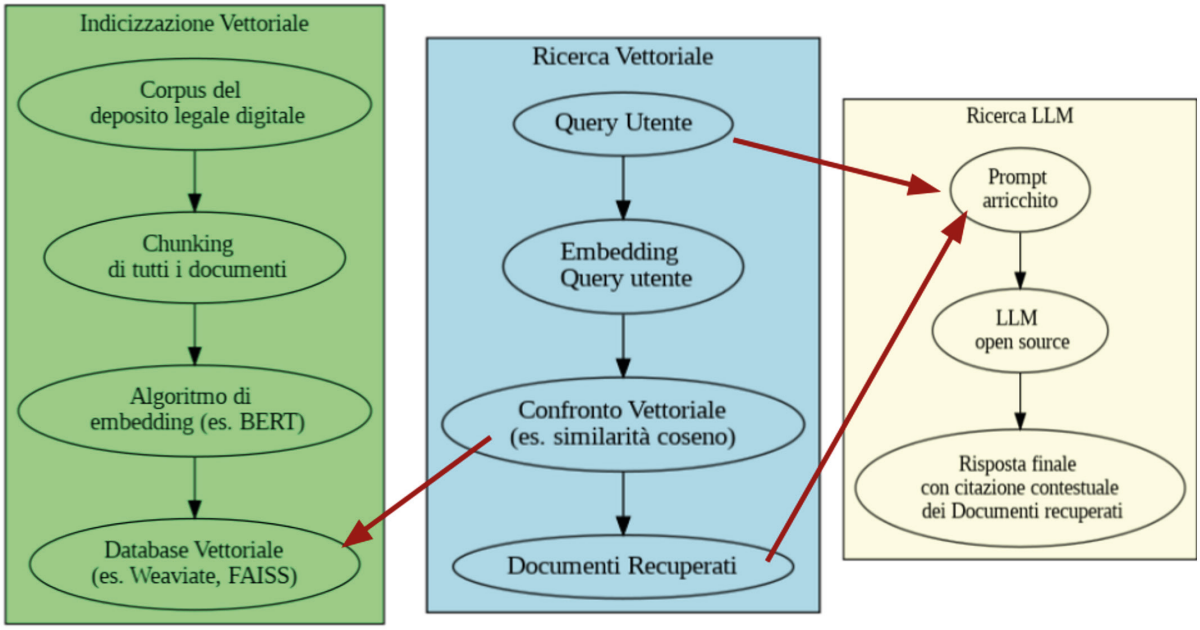


Fig. 1 Modello di computational access con un sistema RAG per il deposito legale digitale

Il contenitore con sfondo verde di Figura 1 rappresenta schematicamente il sistema di indicizzazione tipico di un'architettura RAG:

- tutti i documenti del corpus del deposito legale vengono suddivisi in unità di testo, denominate *chunk* (es. paragrafi, frasi o blocchi di lunghezza fissa). La modalità di segmentazione è definita in base al tipo di applicazione e all'algoritmo di *embedding* utilizzato (es. BERT[57]);
- un algoritmo di embedding (es. BERT) analizza i termini di ciascun *chunk* nel loro contesto e genera una rappresentazione numerica vettoriale (*embedding*) di lunghezza fissa che codifica il 'significato' complessivo del *chunk*;
- gli *embedding* sono memorizzati in un database vettoriale associati a informazioni che facilitano la ricerca dei documenti, come il *chunk* stesso e l'identificativo del documento di provenienza (es URI).

Il contenitore di Figura 1 con sfondo blu illustra i passi di una ricerca vettoriale in un sistema RAG.

- la query dell'utente viene convertita in un *embedding* utilizzando lo stesso algoritmo (es. BERT) impiegato in fase di indicizzazione;
- viene effettuato un confronto di 'vicinanza semantica' tra la query convertita e i contenuti del database vettoriale; l'algoritmo di confronto (es. similarità del coseno[58]) dipende dal tipo di applicazione;
- i risultati restituiranno i documenti del corpus 'semanticamente' più vicini.

[57] https://it.wikipedia.org/wiki/BERT.
[58] https://en.wikipedia.org/wiki/Cosine_similarity.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-697

Infine il contenitore con sfondo giallo di Figura 1 rappresenta il ruolo di un Large Language Model (LLM) nella ricerca:

- la query dell'utente e i *chunk* rilevanti recuperati, insieme ai loro metadati (tra questi l'URI del documento di provenienza), costituiscono il "prompt arricchito"[59] inviato a un LLM;
- un LLM open source (es. LLaMA,[60] Mistral[61]) è sufficiente per elaborare il prompt sfruttando le sue capacità di processare il linguaggio naturale in vari ambiti quali "la comprensione, la traduzione, la generazione e la previsione di nuovi contenuti";[62]
- la risposta finale della ricerca dovrebbe comprendere la sintesi dei risultati ottenuti e la citazione contestuale dei passaggi rilevanti dei documenti recuperati.

Il disegno di dettaglio di una implementazione RAG per il deposito dovrà tenere conto delle tecnologie disponibili al momento della realizzazione. In ogni caso la ricerca in un database vettoriale che indicizza il corpus del deposito legale digitale si propone di superare i limiti dell'indicizzazione full-text di tipo tradizionale (es. Elasticsearch, Apache Solr)[63] che si basa sostanzialmente sulla corrispondenza 'letterale' di parole chiave. Questo strumento per il *computational access* sembra essere una strada promettente per nuove modalità di fruizione e in grado di rispondere alle molteplici esigenze di 'lettura' della raccolta del deposito legale digitale.

Come conclusione del tutto provvisoria a questa sezione dedicata alle nuove modalità di fruizione del deposito legale, possiamo dire che è arrivato il momento per progettare – non solo per il deposito legale – "un percorso di collaborazione per il futuro che metta insieme il meglio dello sforzo umano con il meglio dello sforzo della macchina. Gli esseri umani uniti alle macchine rendono infinitamente meglio degli esseri umani da soli o delle macchine da sole" (Broussard 2019, cap 1.1). Sul versante della catalogazione (o creazione di metadati) – come abbiamo visto nel caso della DNB – il controllo di autorità con vocabolari controllati[64] e il controllo di qualità del risultato richiedono il lavoro professionale del bibliotecario. Così anche i nuovi strumenti di fruizione potranno risultare utili ed efficienti solo se saremo in grado di misurare la loro efficienza. Anche qui abbiamo bisogno del lavoro professionale del bibliotecario. Senza questo "non avremmo né domande intelligenti, né risposte pertinenti; neppure sapremmo interpretare una risposta del sistema come pertinente o meno. In altre parole, non avremmo nessun controllo sulle sue prestazioni" (Cheti 2023, 348).

Un nuovo modello gestionale?

A fronte di tutte le precedenti considerazioni, pare legittimo domandarsi quale modello gestionale e organizzativo si possa ritenere più adatto per una gestione efficace del sistema cooperativo del deposito legale.

[59] In inglese 'augmented prompt': lo stesso 'augmented' dell'acronimo RAG.
[60] https://it.wikipedia.org/wiki/Llama_(modello_linguistico).
[61] https://it.wikipedia.org/wiki/Mistral_AI.
[62] https://www.treccani.it/vocabolario/neo-modello-linguistico-di-grandi-dimensioni_(Neologismi)/.
[63] Sia Elasticsearch che Apache Solr hanno comunque recentemente integrato anche estensioni che permettono la ricerca vettoriale (https://www.elastic.co/elasticsearch; https://solr.apache.org/).
[64] Così come definiti da ISO 25964 (comprendono Tesauri e Liste di autorità).

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-697

Nel 2023 il Gruppo di lavoro sulle biblioteche digitali dell'AIB ha pubblicato un documento dal titolo *Piano d'azione per un'infrastruttura nazionale della conoscenza* (AIB 2023). Con tale locuzione si intende

> un'infrastruttura di base, cui le stesse biblioteche possano fare riferimento per lo sviluppo dei propri servizi, che funga da piattaforma per l'accesso universale alla conoscenza registrata sia sotto forma di documenti che di dati (AIB 2023, 19).

La sua natura è pubblica e nazionale, e le sue funzionalità devono garantire la raccolta, "la registrazione, l'organizzazione, la conservazione permanente e l'accesso universale"(AIB 2023, 15) alle risorse informative e culturali, espresse appunto anche come dati.
Il Piano individua tre pilastri dell'infrastruttura, per ciascuno dei quali vengono considerati i limiti attuali e le possibili linee di sviluppo: – la rete del Servizio Bibliotecario Nazionale, – il deposito legale e la sperimentazione realizzata con il progetto Magazzini Digitali, – le istituzioni, i principi e le prassi che costituiscono il dominio della Scienza Aperta. Nel riconoscere la centralità del sistema del deposito legale, inteso come s'è detto come strumento d'elezione per la raccolta e la conservazione dell'eredità culturale, il Piano rileva sia la contraddizione tra tale centralità e la progressiva marginalizzazione delle biblioteche nelle politiche pubbliche, che l'inadeguatezza delle strutture organizzative, gestionali e dei modelli di finanziamento delle biblioteche stesse. La situazione è particolarmente critica nelle biblioteche nazionali, a partire dalle due centrali, che del deposito legale condividono la massima responsabilità, le cui dotazioni organiche e finanziarie non sono nemmeno lontanamente paragonabili a quelle delle istituzioni sorelle dei maggiori paesi europei ed extraeuropei. Colpisce ad esempio l'assenza nei nostri istituti centrali, fra i quali si annovera anche l'ICCU, di veri e propri uffici ricerca e sviluppo, che sono spesso motori d'innovazione e sperimentazione, e laboratori di pensiero scientifico e di modelli organizzativi. Non che a Firenze e Roma non si faccia ricerca e sperimentazione,[65] ma si tratta di attività necessariamente residuali, portate avanti grazie alla volontà ed alle straordinarie competenze di singoli funzionari. Storicamente, l'assimilazione delle biblioteche e dei loro servizi al dominio dei beni culturali[66], se da un lato ha permesso di ancorarle ad almeno un settore significativo delle politiche pubbliche, le ha pure costrette a subirne ristrettezze ed oscillazioni, ed ha rallentato se non impedito la formazione di una coscienza civile e politica della natura critica ed essenziale dei servizi bibliotecari, in particolare di quelli nazionali, come si è visto sopra a proposito di conservazione e *cybersecurity*.
Come si diceva in apertura di questo contributo, inoltre, le biblioteche sono partecipi di una pluralità di ecosistemi, non solo di quello dei beni culturali.
È dunque necessario immaginare modelli di organizzazione, di governo e di sostenibilità finanziaria adeguati agli scopi che l'infrastruttura della conoscenza, come qui delineata, si pone. A questo proposito il citato Piano d'azione (AIB 2023, 17, 132) sostiene che modelli organizzativi che possano operare secondo le norme del diritto privato, come le aziende detenute direttamente dall'Amministrazione, siano più adatti alla gestione di servizi ad alto contenuto di tecnologie.

[65] Si veda ad es. l'eccellente contributo di (Storti 2024).
[66] Su questi temi l'AIB si è espressa più volte, a partire dalla seconda tesi del Congresso di Viareggio del 1987 https://www.aib.it/aib/commiss/cnbp/tesi.htm, fino a (AIB 2023, 74).

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-697

Un modello praticabile perché già sperimentato con successo è quello dell'azienda *in house*, detenuta normalmente da una Pubblica Amministrazione, es. un ministero, e che eroga servizi alla medesima Pubblica Amministrazione.[67]
Le società *in house* sono definite dal DL 175/2016, *Testo unico in materia di società a partecipazione pubblica*, art. 2, comma o.
Due sono gli aspetti previsti dal DL che è opportuno qui mettere in luce:

- l'art. 16, comma 3, prevede che oltre l'ottanta per cento del fatturato della società debba derivare dallo svolgimento dei compiti ad essa affidati dall'ente pubblico o dagli enti pubblici soci. La produzione ulteriore di fatturato "è consentita solo a condizione che la stessa permetta di conseguire economie di scala o altri recuperi di efficienza sul complesso dell'attività principale della società".
- l'art. 19, comma 2 prevede che "Le società a controllo pubblico stabiliscono, con propri provvedimenti, criteri e modalità per il reclutamento del personale nel rispetto dei principi, anche di derivazione europea, di trasparenza, pubblicità e imparzialità e dei principi di cui all'articolo 35, comma 3, del DL 165/2001".

Pur nei limiti sopra definiti, dunque, la società *in house* può svolgere attività di mercato, producendo valore ulteriore per l'ente pubblico che la detiene, e può assumere con relativa autonomia il personale dotato delle competenze tecniche necessarie allo svolgimento dei propri compiti istituzionali. Una società *in house* partecipata, per conto del MiC, dalle due biblioteche nazionali centrali (istituzioni depositarie centrali), e per conto delle Regioni dalle biblioteche depositarie regionali, potrebbe quindi offrire servizi tecnologici – che richiedono continuità operativa e un alto livello di specializzazione – al sistema cooperativo del deposito legale con criteri di efficacia ed efficienza.[68]
Non si può peraltro tacere il fatto che l'iter per la costituzione di una società a partecipazione pubblica, oltre a dover essere motivato analiticamente dall'amministrazione proponente con riferimento alla sua necessità per il perseguimento delle proprie finalità istituzionali, anche sul piano della convenienza economica e della sostenibilità finanziaria,[69] risulta particolarmente complesso. Richiede infatti un decreto del Presidente del Consiglio dei ministri, su proposta del Ministro dell'economia e delle finanze di concerto con i ministri competenti per materia, previa deliberazione del Consiglio dei ministri,[70] e prevede i controlli dell'Autorità garante della concorrenza e del mercato e della Corte dei conti.
È evidente quindi la necessità di una forte e chiara volontà politica, che andrebbe "costruita" con il lavoro delle istituzioni depositarie, a livello nazionale e regionale, e con l'appoggio e specifiche attività di *advocacy* da parte dell'associazione e dell'intera comunità professionale.
Si può infine notare che la più recente riforma del Ministero della Cultura, attuata con il DPCM

[67] Valga per tutte l'esempio di Sogei, Società generale d'informatica SpA: è la società di information technology detenuta al 100% dal Ministero dell'economia e delle finanze, che fornisce servizi all'amministrazione economico-finanziaria e ad altre amministrazioni dello Stato. Un elenco esemplificativo di società pubbliche o a capitale misto pubblico-privato è riportato in (AIB 2023, 130-32).
[68] Questo contributo è focalizzato sul deposito legale, ma l'oggetto di una società pubblica come quella qui ipotizzata potrebbe positivamente estendersi all'offerta di servizi tecnologici per la gestione complessiva dei servizi bibliotecari nazionali.
[69] DL 175/ 2016, art. 5.
[70] DL 175/ 2016, art. 7.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-697

57/2024, ha stabilito con l'art. 3, comma 7 d) l'afferenza della Direzione generale Biblioteche e istituti culturali, da cui dipende il deposito legale, al Dipartimento per le attività culturali del MiC, interrompendo una tradizione normativa ed organizzativa che ha sempre visto le biblioteche afferire, con gli archivi e i musei, agli organismi deputati alla tutela. Tale scelta è stata oggetto di critica, ma potrebbe in realtà costituire una buona occasione per portare le istituzioni deputate all'erogazione dei servizi bibliografici e bibliotecari nazionali nell'alveo proprio di quelle che erogano servizi culturali essenziali e critici, come s'è detto sopra.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-697

Riferimenti bibliografici

AIB (Associazione italiana biblioteche, Gruppo di lavoro sulle biblioteche digitali). 2023. *Piano d'azione per l'infrastruttura nazionale delle conoscenza.* Roma: Associazione italiana biblioteche.

Bergamin, Giovanni, e Lorenzo Gobbo. 2025. "IPFS for the long term digital preservation". (In corso di pubblicazione).

Bergamin, Giovanni, e Anna Lucarelli. 2012. "The Nuovo Soggettario as a service for the linked data world." *JLIS.it* 4 (1): 213. https://doi.org/10.4403/jlis.it-5474.

Breeding, Marshall, "2025 library systems report," *American libraries*, 1 maggio 2025, https://americanlibrariesmagazine.org/2025/05/01/2025-library-systems-report/.

Broussard, Meredith. 2019. *La non intelligenza artificiale: come i computer non capiscono il mondo.* Milano: Angeli.

Chae, Youngjin, e Thomas Davidson. 2025. "Large language models for text classification: from zero-shot learning to instruction-tuning." *Sociological methods & research*, https://doi.org/10.1177/00491241251325243.

Cheti, Alberto. 2025. "L'accesso per soggetto: dal catalogo a schede al catalogo online, con uno sguardo all'Intelligenza Artificiale." *AIB Studi*, 64 (3): 321–52. https://doi.org/10.2426/aibstudi-14127.

Chilovi, Desiderio. 1903. *L'archivio della letteratura italiana e la Biblioteca nazionale centrale di Firenze.* Firenze: Bemporad (testo reperibile anche in https://www.aib.it/aib/stor/testi/chilovi3.htm).

Comitato dei Ministri agli Stati membri sulla legislazione e la politica delle biblioteche in Europa. 2023. *Raccomandazione del Comitato dei Ministri agli Stati membri sulla legislazione e la politica delle biblioteche in Europa*. https://eblida.org/wp-content/uploads/2024/04/Traduzione_EBLIDA_Raccomandazioni_Consiglio_d_Europa.pdf.

Connaway, Lynn Silipigni, e Timothy J. Dickey. 2010. *Towards a profile of the researcher of today: What can we learn from JISC projects? Common themes Identified in an analysis of JISC Virtual Environment and Digital Repository Projects*. https://repository.jisc.ac.uk/418/2/VirtualScholar_themesFromProjects_revised.pdf.

Di Marcantonio, Giorgia. 2025. "Intelligenza artificiale, Large Language Models (LLMs) e Retrieval-Augmented Generation (RAG). Nuovi strumenti per l'accesso alle risorse archivistiche e bibliografiche." *Bibliothecae.it* 13 (1): 161-73.

DNB (Deutsche Nationalbibliothek). 2017. *General outline and introduction of the future subject cataloguing procedure for publications at the German National Library.* https://www.dnb.de/SharedDocs/Downloads/EN/Professionell/Erschliessen/konzeptWeiterentwicklungInhaltserschliessungMai2017.html.

DPC (Digital Preservation Coalition). 2022. *Computational Access: A beginner's guide for digital preservation practitioners.* http://doi.org/10.7207/compaccess22-01.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-697

Fan, Wenqi , Yujuan Ding, Liangbo Ning, Shijie Wang, Hengyun Li, Dawei Yin, Tat-Seng Chua, e Qing Li. 2024. "A Survey on RAG meeting LLMs: towards retrieval-augmented Large Language Models." In *KDD '24: Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining,* 6491-501. https://doi.org/10.1145/3637528.3671470.

Farrell, Henry, Alison Gopnik, Cosma Shalizi, e James Evans. 2025. "Large AI models are cultural and social technologies." *Science* 387(6739), 1153-6. https://doi.org/10.1126/science.adt9819.

Gao, Yunfan, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, e Haofen Wang. "Retrieval-Augmented Generation for Large Language Models: A Survey." Preprint, sottomesso il 27 marzo 2024. https://doi.org/10.48550/arXiv.2312.10997.

Golub, Koraljka, Osma Suominen, Ahmed Taiye Mohammed, Harriet Aagaard, e Olof Osterman. 2024. , "Automated Dewey Decimal Classification of Swedish library metadata using Annif software." *Journal of Documentation* 80 (5): 1057-79. https://doi.org/10.1108/JD-01-2022-0026.

Gömpel, Renate, Ulrike Junger, e Elisabeth Niggemann. 2010, "Veränderungen im Erschließungskonzept der Deutschen Nationalbibliothek." *Dialog mit Bibliotheken* 22 (1): 20–2 https://d-nb.info/1009804510/34.

Guerrini, Mauro. 2021. "Il controllo bibliografico nell'era digitale: identificatori, metadati, punti d'accesso e rispetto del contesto culturale e linguistico." *Biblioteche oggi trends* 7 (1). https://www.doi.org/10.3302/2421-3810-202101-092-1.

Humphreys, Kenneth William. 1964. "The role of the national Library: a preliminary statement." *Libri* 14: 356-68.

IFLA FAIFE (Committee on Freedom of Access to Information and Freedom of Expression). 2020. *IFLA statement on libraries and artificial intelligence.* https://repository.ifla.org/handle/20.500.14598/1646.

Lewis, Patrick, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, e Douwe Kiela, 2020. "Retrieval-Augmented Generation for knowledge-intensive NLP tasks." In *Advances in Neural Information Processing Systems 33 (NeurIPS 2020).* https://proceedings.neurips.cc/paper_files/paper/2020/file/6b493230205f780e1bc26945df7481e5-Paper.pdf.

Kempf, Klaus. 2013. *Der Sammlungsgedange im Digitalen Zeitalter - L'idea della collezione nell'età digitale.* Fiesole: Casalini libri.

Kluge, Lisa, e Maximilian Kähler. "DNB-AI-Project at SemEval-2025 Task 5: An LLM-Ensemble approach for automated subject indexing". Preprint, sottomesso il 30 aprile 2025. https://doi.org/10.48550/arXiv.2504.21589.

Maltese, Diego. 1963. "Bibliografia, commercio librario e archivio nazionale del libro." *Bollettino d'informazioni AIB* 3 (3): 84-9.

Maltese, Diego. 1977. "Natura e formazione dell' archivio nazionale del libro." *Bollettino d'informazioni AIB* 17(4): 286-94.

Maltese, Diego. 1979. "Natura e funzioni della Biblioteca nazionale centrale." *Il ponte* 35 (4): 446-60.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-697

Mandillo, Anna Maria. 2000. "La nuova legge sul deposito legale: una riforma non solo per le biblioteche." *AIB notizie* 14 (3): 4-7. https://www.aib.it/aib/editoria/n14/02-03mandillo.htm.

Meschini, Federico. 2021. "Chi controlla il controllo bibliografico universale? Il sogno dei bibliotecari nell'ecosistema digitale." *AIB studi: rivista di biblioteconomia e scienze dell'informazione* 61 (3): 623-27. https://doi.org/10.2426/aibstudi-13325.

Mödden, Elisabeth. 2022. "Artificial Intelligence, Machine Learning and Bibliographic Control. DDC Short Numbers - Towards Machine-Based Classifying." *JLIS.it* 13 (1): 256-64. https://doi.org/10.4403/jlis.it-12775.

Penzo Doria, Gianni. 2022. "Una nuova definizione di archivio." *JLIS.it* 13 (2): 156-73. https://doi.org/10.36253/jlis.it-465.

Puglisi, Paola. 2007. "Deposito legale: la bicicletta nuova". *Bollettino AIB* 47(1-2): 11-41. https://bollettino.aib.it/article/view/5203.

Puglisi, Paola. 2020. "Deposito legale quattordici anni dopo: come, quando, quanto e perché." *AIB studi* 60 (3): 591-614. https://aibstudi.aib.it/article/view/12477/11788.

Roncaglia, Gino. 2023. *L'architetto e l'oracolo.* Bari: Laterza.

Storey, Veda C., Wei Thoo Yue, J. Leon Zhao, e Roman Lukyanenko. 2025. "Generative artificial intelligence: evolving technology, growing societal impact, and opportunities for information systems research." *Information Systems Frontiers.* https://doi.org/10.1007/s10796-025-10581-7.

Storti, Chiara. 2023. "'Resource not found': cultural institutions, interinstitutional cooperation and collaborative projects for web heritage preservation." *JLIS.it* 14 (2): 39–52. https://doi.org/10.36253/jlis.it-533.

Storti, Chiara. 2024. "I requisiti per l'addestramento degli strumenti di Intelligenza Artificiale e il deposito legale delle risorse digitali: una riflessione sul contesto normativo italiano e il ruolo delle biblioteche." *Digitalia* 19 (1): 23-35. https://doi.org/10.36181/digitalia-00092.

Suominen, Osma, Juho Inkinen, e Mona Lehtinen. 2022. "Annif and Finto AI: Developing and implementing automated subject indexing." *JLIS.it* 13 (1): 265–82. https://doi.org/10.4403/jlis.it-12740.

Suominen, Osma, Juho Inkinen, e Mona Lehtinen. "Annif at SemEval-2025 Task 5: traditional XMTC augmented by LLMs." Preprint, sottomesso il 28 aprile 2025. https://doi.org/10.48550/arXiv.2504.19675.

Tammaro, Annamaria, e Emanuele Bellini. 2024. "Il ruolo della cibersecurity per le biblioteche digitali: sfide attuali e strategie di protezione." *Digitalia* 19 (2): 100-14. https://digitalia.cultura.gov.it/article/view/3071/2151.

Taylor, Jaime, Kelly Dagan, Margaret Youngberg, Therese Kaufman, e Johanna Radding. 2025. "A Survey of AI tools in library tech: accelerating into and unlocking streamlined enhanced convenient empowering game-changers." *Journal of electronic resources librarianship*, May, 1–14. https://doi.org/10.1080/1941126X.2025.2497738.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-697

Traniello, Paolo. 1995. "La legislazione italiana sul deposito obbligatorio: l'eredità ottocentesca." *Bollettino AIB* 35 (2): 221-31.

Traniello, Paolo. 2000. "A proposito di archivio del libro: riflessioni su una sorprendente anticipazione di Domenico Comparetti." *Bollettino AIB* 40 (2): 233-40.

Vitiello, Giuseppe. 2007. "Come si consolida un'anomalia bibliotecaria: a proposito della nuova legge sul deposito legale in Italia". *Biblioteche oggi* 25 (1): 9-21. https://www.bibliotecheoggi.it/media/download/get/28840afe-b5d9-45ff-97a6-09e942cc7ac8/original.

Vitiello, Giuseppe. 2024. "Dalla Raccomandazione del Consiglio d'Europa (2023) al Rapporto OMC sulle biblioteche del Consiglio dell'Unione europea (2026)." *Biblioteche oggi* 42 (5): 3-14. https://doi.org/10.3302/0392-8586-202405-003-1.

Zhao, Penghao, Hailin Zhang, Qinhan Yu, Zhengren Wang, Yunteng Geng, Fangcheng Fu, Ling Yang, Wentao Zhang, Jie Jiang, e Bin Cui. "Retrieval-augmented generation for ai-generated content: a survey." Preprint, sottomesso il 29 febbraio 2024. https://doi.org/10.48550/arXiv.2402.19473.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-682

JLIS.it

Collaboration, Cooperation, and Community: ITALE and SBN*

Raffaella Sprugnoli[(a)], Silvia Ceccarelli[(b)]

a) University of Pisa, https://orcid.org/0009-0003-4212-3068
b) University of Insubria, https://orcid.org/0009-0003-1827-486X

Contact: Raffaella Sprugnoli, raffaella.sprugnoli@unipi.it; Silvia Ceccarelli, silvia.ceccarelli@uninsubria.it
Received: 28 July 2025; **Accepted:** 20 October 2025; **First Published:** 15 January 2026

ABSTRACT
The article retraces the history and evolution of the collaboration between ITALE, the Italian community of Ex Libris users, and the National Library Service (SBN), focusing on ITALE's role in promoting cooperation among library institutions. It outlines the main stages of the technical and institutional dialogue with ICCU, aimed at ensuring interoperability between library management systems (initially Aleph, later Alma) and the SBN infrastructure. Special attention is given to the challenges posed by the management of the UNIMARC format and its relationship with the SBN-MARC protocol, as well as to the role of ITALE working groups in providing technical support and negotiating with Ex Libris. In the final section, the article explores future perspectives, particularly the introduction of Linked Open Data and the evolution of the UNIMARC format, identifying the synergy between ITALE and ICCU as a crucial element in guiding the Italian library community toward new descriptive paradigms and greater international visibility of bibliographic data.

KEYWORDS
ITALE; National Library Service (SBN); Aleph/Alma; Library cooperation; Bibliographic standards; Technological evolution.

Collaborazione, cooperazione e comunità: ITALE e SBN

ABSTRACT
L'articolo ripercorre la storia e l'evoluzione della collaborazione tra ITALE, la comunità italiana degli utenti Ex Libris, e il Servizio Bibliotecario Nazionale (SBN), concentrandosi sul ruolo svolto da ITALE nella promozione della cooperazione tra istituzioni bibliotecarie. Vengono illustrate le principali tappe del dialogo tecnico e istituzionale con l'ICCU, volto a garantire l'interoperabilità tra i sistemi gestionali (prima Aleph, poi Alma) e l'infrastruttura SBN. Particolare attenzione è rivolta alle sfide poste dalla gestione del formato UNIMARC e dal rapporto con il protocollo SBN-MARC, nonché al ruolo dei gruppi di lavoro ITALE nel supporto tecnico e nella negoziazione con Ex Libris. Nella parte finale, l'articolo esplora le prospettive future, in particolare l'introduzione dei Linked Open Data e l'evoluzione del formato UNIMARC, individuando nella sinergia tra ITALE e ICCU un elemento cruciale per accompagnare la comunità bibliotecaria italiana verso nuovi paradigmi descrittivi e una maggiore visibilità internazionale dei dati bibliografici.

PAROLE CHIAVE
ITALE; Servizio Bibliotecario Nazionale (SBN); Aleph/Alma; Cooperazione bibliotecaria; Standard bibliografici; Evoluzione tecnologica.

* Le autrici desiderano ringraziare Claudia Burattelli, Laura Soro e Silvia Trevenzoli per i preziosi suggerimenti.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-682

ITALE, la Comunità italiana utenti ex libris

L'Associazione ITALE è nata ed ha operato dal 1993 al 2007 come Associazione Italiana Utenti ALEPH.
Alla fine del 2007 si è allargata e trasformata, sul modello del Gruppo Internazionale degli Utenti Ex Libris (IGeLU), in **Associazione italiana degli utenti dei prodotti Ex Libris** accogliendo al suo interno gli utenti di tutti i prodotti Ex Libris (Aleph, Alma, Primo, Summon, Leganto, ecc.).
Nel 2019 lo Statuto è stato modificato in modo da poter accogliere nell'associazione anche le istituzioni che adottano i prodotti Ex Libris in lingua italiana, pur se non risiedenti in Italia: ITALE funge infatti da riferimento per le questioni relative alla traduzione in lingua italiana dei prodotti Ex Libris.
ITALE è, da sempre, una associazione di enti e non di persone. Sono attualmente iscritte ad ITALE 49 istituzioni che hanno adottato i sistemi Ex Libris per l'automazione dei servizi bibliotecari (3 istituzioni Svizzere).
Le aree di intervento sono quelle relative all'automazione bibliotecaria, all'applicazione degli standard internazionali, alla valorizzazione delle collezioni e alla gestione dei servizi, settori nei quali l'Associazione organizza seminari e incontri periodici, coordinandosi sia con IGeLU, sia con gli altri gruppi nazionali di utenti.
L'Associazione svolge funzioni di coordinamento tra gli associati per gli aspetti concernenti le richieste di sviluppo del software, la standardizzazione nell'uso delle relative funzioni, rappresenta le istanze e le esigenze degli utenti italiani all'interno di IGeLU e nei confronti di Ex Libris.
Presentandosi come luogo privilegiato per la condivisione di esigenze e lo scambio di esperienze, l'Associazione si occupa inoltre, della diffusione, fra i propri membri, di tecnologie, norme di comportamento per la gestione dei servizi, sistemi innovativi nei settori della comunicazione e della trasmissione di testi e di dati.
ITALE opera tramite lo staff delle istituzioni associate, bibliotecari e tecnici in massima parte, riuniti in Gruppi di Lavoro su specifici ambiti e prodotti con il coordinamento del Presidente e del Comitato. Sono attualmente attivi 7 gruppi di lavoro: Alma, Alma – Risorse Elettroniche, Alma Digital, Discovery Tool, Resource Sharing (ILL/DD), Leganto, Standard di catalogazione e dialogo con SBN. L'associazione si riunisce in assemblea 2 volte l'anno. All'assemblea dei soci si accompagnano di norma il workshop dei gruppi di lavoro e un seminario tecnico.
Tutte le informazioni su ITALE e le sue attività sono pubblicate sul sito web https://itale.igelu.org/.
ITALE nasce, vive e si può sviluppare solo grazie all'impegno ed alla partecipazione dei bibliotecari e dei tecnici dedicati alle biblioteche. L'esperienza di ITALE mostra come attraverso la condivisione di esperienze e conoscenze, il confronto di idee, il coordinamento di esigenze comuni, sia possibile superare ostacoli e sfide dinanzi ai quali le singole forze di ognuno sarebbero insufficienti.
L'associazione, grazie all'impegno dei suoi membri e all'abitudine al lavoro di gruppo e alla condivisione di esperienze e competenze è diventata una vera Comunità di utenti e come tale svolge al meglio il suo ruolo di rappresentante delle esigenze degli utenti italiani nei confronti degli altri gruppi di utenti e di Ex Libris.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-682

La cooperazione in ITALE

Tra le istituzioni ITALE sono state sperimentate diverse forme di cooperazione:

- il Catalogo collettivo Bicocca - Insubria che fino al 2019 riuniva in un unico catalogo basato su Aleph le collezioni delle università di Milano Bicocca e dell'Insubria
- il Sistema Bibliotecario Genovese, che, fino al 2018 riuniva in un unico catalogo basato su Aleph le collezioni del comune e quelle dell'università
- il consorzio SBART (Sistemi Bibliotecari di Ateneo Regione Toscana) che costituisce un felice esempio di cooperazione nella cooperazione: al suo interno è stata realizzata l'implementazione consortile del Discovery Primo per il quale è stata sperimentata una struttura particolare. Esso è suddiviso in quattro ambienti corrispondenti ai cataloghi delle istituzioni partecipanti (Università di Firenze, Università di Pisa, Servizio Bibliotecario Senese – comprensivo di Università di Siena, Università per Stranieri di Siena e rete REDOS – e Scuola Superiore Sant'Anna di Pisa) più un ambiente cumulativo, la cosiddetta vista SBART[1] che costituisce un vero e proprio metacatalogo. Le quattro università hanno inoltre attivato la condivisione delle anagrafiche degli utenti i quali possono quindi usufruire dei servizi in maniera fluida e semplificata[2]
- l'adesione delle istituzioni ITALE al protocollo ILL SBN per il servizio di prestito interbibliotecario. L'integrazione con ILL-SBN è una prassi ormai consolidata in molte biblioteche ITALE che ne apprezzano la capacità di semplificare i flussi di lavoro e migliorare l'efficienza del servizio
- lo scambio "Alma to Alma" permette di gestire l'intero flusso di lavoro in un unico ambiente e apre la possibilità di prendere accordi e di gestire le richieste con partner internazionali, grazie alla *Resource Sharing Directory*
- il nuovo Protocollo di intesa per la fornitura dei servizi di prestito interbibliotecario e fornitura di documenti in reciproco scambio gratuito tra le istituzioni aderenti alla associazione ITALE avviato nel 2025
- il progetto, non ancora giunto a realizzazione, per l'integrazione del modulo Resource sharing di Alma con il circuito NILDE per il document delivery.

L'associazione favorisce, stimola, supporta e coordina, sin dalla sua fondazione, la cooperazione tra le istituzioni e propone progetti di innovazione, anche tecnologica per la gestione dei servizi bibliotecari.

[1] https://sbart-network.primo.exlibrisgroup.com/discovery/search?vid=39SBART_NETWORK:SBART_UNION&lang=it.

[2] Oltre al Discovery gli atenei toscani condividono anche il catalogo mediante l'area Network del gestionale Alma (cfr. *infra*) nella quale è stato attivato il colloquio con Indice SBN del polo SBT (Sistemi Bibliotecari Toscani). Da questo particolare assetto è derivata la necessità di armonizzare le pratiche catalografiche dei tre atenei caratterizzate da difformità sia a livello teorico che di prassi lavorative. A questo scopo è stato creato, nel 2019, un gruppo di lavoro interateneo che si occupa della redazione di linee guida comuni e cura la manutenzione del catalogo. Nel 2024, poi, il polo SBT ha attivato una convenzione con la Biblioteca Nazionale Centrale di Firenze finalizzata *in primis* all'arricchimento del Thesaurus del Nuovo soggettario ma che prevede anche altri ambiti di collaborazione tra cui, ad esempio, la valutazione di "procedure per l'interoperabilità con altre risorse online di comune interesse" e lo sviluppo di "tematiche di comune interesse anche nell'ambito di progetti che coinvolgano altri partner (ad esempio, Wikidata, applicazioni legate a linked open data, indicizzazione automatica, ecc.)".

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-682

Nonostante l'assenza di un catalogo condiviso, nella Comunità ITALE la cooperazione tra le biblioteche e i bibliotecari si è estesa fino a coprire tutti gli aspetti della vita delle biblioteche, dalla valorizzazione delle risorse, alla gestione dei servizi di back office e front office.

La collaborazione tra ITALE e ICCU/SBN

Il rapporto tra l'Associazione ITALE e l'Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche (d'ora in avanti ICCU) rappresenta un caso emblematico di cooperazione tra comunità professionali e istituzioni pubbliche nel contesto dell'evoluzione dei sistemi bibliotecari italiani.
Fin dalla sua fondazione, ITALE ha identificato come strategico il dialogo con ICCU, sottolineando la necessità che gli utenti Aleph – il software gestionale in uso tra la fine degli anni '90 del secolo scorso e gli anni '10 di quello corrente – potessero interagire con il sistema del Servizio Bibliotecario Nazionale (d'ora in avanti SBN).
Con questo scopo, a partire dagli anni 2000, ITALE ha promosso un dialogo costante con ICCU, volto a favorire l'interoperabilità tra Aleph e l'infrastruttura nazionale di SBN, con particolare attenzione ai protocolli di scambio dati, alla catalogazione condivisa, per giungere poi, in una fase più avanzata, ai servizi di prestito interbibliotecario.
Volendo tracciare schematicamente una storia[3] dei rapporti tra l'Associazione e ICCU, questa potrebbe essere articolata in quattro fasi.
Una **fase iniziale** (2000–2004), quella del **riconoscimento della necessità di dialogo**. L'attenzione si concentra sulla possibilità di cooperazione tecnica, in particolare sullo scarico dei record bibliografici in formato UNIMARC, e sull'opportunità offerta dalla nascita del portale SBN per facilitare lo scambio di dati. In particolare, dal 2004, i nuovi sviluppi dell'Indice SBN e del protocollo SBN-MARC rendono tecnicamente possibile l'interconnessione tra Aleph e SBN.
La **seconda fase** (2004–2007) è quella della **sperimentazione e del consolidamento tecnico**: ITALE avvia un piano di migrazione nazionale per gli utenti Aleph, accompagnato da una fase di formazione e test. In parallelo, viene avviato il progetto di collegamento tra Aleph500 e Indice2 SBN, cui si affianca la costituzione da parte di ICCU di un Gruppo di mantenimento del protocollo SBN-MARC, incaricato di definire metodologie di test, certificare le release e supportare gli sviluppatori del software.
La **terza fase** (2007-2015) è quella del **consolidamento e dell'estensione della cooperazione al prestito interbibliotecario**. Nel 2007, in concomitanza con lo sviluppo da parte di Ex Libris del nuovo modulo ILL2 nella versione 18 di Aleph500, ITALE avvia al proprio interno il percorso che condurrà all'adesione al servizio ILL-SBN. Anche questa fase è caratterizzata da un confronto costante tra tutti gli attori coinvolti. Vengono avviati i test del colloquio fra ILL-SBN e ILL2 e viene istituito un tavolo di lavoro comune con lo scopo di individuare le eventuali criticità studiando al contempo anche la possibilità di nuovi sviluppi da sottoporre al team di Ex Libris. Il percorso per arrivare alla piena interoperabilità è lungo e articolato e conosce anche fasi di stallo. Nel 2013, il progetto di col-

[3] La storia dei rapporti tra ITALE e ICCU è stata ricostruita sulla base dei verbali dell'Associazione disponibili sul sito di ITALE nella sezione Assemblee (accessibile da parte dei soci itale previa autenticazione).

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-682

loquio via ISO-ILL raggiunge una fase avanzata, con test completati tra gateway SBN e client Aleph e l'estensione della sperimentazione a nuove istituzioni aderenti ad ITALE. La piena operatività è raggiunta nel 2015 con la predisposizione del protocollo di intesa ITALE-ICCU e l'adesione dell'Università di Firenze cui seguiranno negli anni successivi anche altre istituzioni.

In questa fase il colloquio tra Aleph500 e SBN ha raggiunto un buon livello di efficienza e consente la cooperazione con Indice ai vari livelli di adesione previsti dal Protocollo SBN-MARC. È noto che le modalità partecipazione al catalogo SBN da parte dei diversi Poli ad esso aderenti sono determinate nella configurazione profilo di catalogazione di ciascuno di essi. L' adesione al catalogo si articola quindi in livelli che vanno da quello minimo con il quale i Poli si limitano a fornire la sola segnalazione di possesso di documenti già presenti nel catalogo, a quello massimo che comporta la possibilità di incrementarlo aggiungendovi documenti in esso ancora non presenti.[4]

Il livello massimo di cooperazione è quello che prevede i cosiddetti "allineamenti" che implicano la uniformazione dei dati del catalogo locale ai corrispondenti dati dell'Indice, accogliendo le correzioni apportate da altri.

Aleph ha ottenuto la certificazione sia per il livello 3[5] che per il livello 4 di adesione a Indice e tre università, Padova, Firenze e Genova, lo utilizzano per la catalogazione in colloquio con SBN. Tra queste solo l'Università di Padova ha sperimentato l'adesione a livello 4, che, pur funzionante, presenta alcune criticità soprattutto legate alla difficoltà nella ricezione degli allineamenti.

Un'altra importante criticità rilevata nella gestione del colloquio con Indice è legata alla non perfetta corrispondenza tra il Protocollo SBN-MARC e il formato UNIMARC.

Il protocollo SBN-MARC è stato sviluppato per permettere il colloquio tra sistemi locali (come Aleph) e l'Indice SBN. Tuttavia, pur ispirandosi al formato UNIMARC, esso ne rappresenta una versione semplificata e adattata alle esigenze del sistema SBN. Da ciò scaturiscono significativi discostamenti dal formato sintetizzabili come segue:

- non perfetta corrispondenza tra campi e sottocampi nei due formati
- interpretazioni diverse di alcuni elementi catalografici
- limitazioni nella granularità delle informazioni trasmesse.

L'uso non letterale dello standard UNIMARC in SBN ha implicato per gli utilizzatori di Aleph un adattamento della pratica catalografica che è giunto in alcuni casi sino allo scostamento dallo standard con l'adozione di scelte ad esso non conformi. L'aspetto della coerenza dei dati tra sistemi diversi e la necessità di strumenti di conversione o mappatura tra i due formati ha un impatto importante nella vita dell'associazione traducendosi nella costante analisi delle criticità e delle soluzioni che nel corso del tempo vengono elaborate mediante la continua negoziazione tra gli utenti di Aleph e gli sviluppatori di Ex Libris. Questa relazione, dinamica e complessa, è anche alla base delle discussioni avviate all'interno dell'associazione su argomenti quali la sostenibilità a lungo termine dell'interoperabilità e le riflessioni sulla fattibilità teorica e tecnica dell'abbandono di UNIMARC in favore di MARC 21.[6]

[4] https://norme.iccu.sbn.it/index.php?title=Norme_comuni/La_catalogazione_partecipata/I_livelli_di_adesione.

[5] Il livello 3 di adesione prevede: cattura e localizzazione per possesso, creazione e correzione dei record non condivisi. Il livello 4 prevede: cattura e localizzazione per possesso e gestione, creazione, correzione ed allineamento. https://www.iccu.sbn.it/it/SBN/poli-e-biblioteche/tipologia-poli/index.html.

[6] La discussione sull'opportunità di passare da UNIMARC a MARC 21 comincia nel 2008. L'ipotesi del cambiamento

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-682

La quarta fase, in cui ci troviamo tuttora, inizia a metà degli anni '10 e segna l'apertura di uno scenario completamente nuovo. Essa si pone come un momento di svolta sia nelle modalità di cooperazione all'interno di ITALE, che nel rapporto con ICCU-SBN. È in questo periodo infatti che le istituzioni italiane si preparano ad adottare **Alma**, il nuovo gestionale sviluppato da Ex Libris destinato a sostituire Aleph.[7]

Alma, uno dei primi **URM** ad affacciarsi sul mercato,[8] rappresenta infatti un cambio radicale di paradigma con il quale si abbandona l'approccio "a compartimenti stagni" tipico degli ILS che prevedeva l'utilizzo di una pluralità di sistemi a seconda della tipologia di materiale che si desiderava gestire (materiale a stampa, risorse elettroniche, digitalizzazioni ecc.). Alma, infatti, integra in un'unica piattaforma il supporto a tutte le operazioni di biblioteca (selezione, acquisizione, gestione dei metadati, digitalizzazione e servizi) per qualunque tipo di materiale combinando e riunendo le funzioni di un ILS con quelle di un ERM. Alma inoltre, include in modo nativo (a differenza di quanto accadeva nella maggior parte degli ILS tradizionali) Analytics, uno strumento di analisi e reportistica particolarmente duttile e potente.

Alma,[9] infine, anche in questo differenziandosi dagli ILS, non è un software da installare e gestire su server locale, ma è un SaaS (Software as a Service), un servizio basato su un'infrastruttura cloud gestita interamente da Ex Libris con tutti i vantaggi e gli svantaggi che questo comporta: nessun impegno di tipo sistemistico per le istituzioni che adottano il sistema da una parte, minore possibilità di intervento e quindi ridotta capacità di personalizzazione per venire incontro a specifiche esigenze dall'altra.

Come è facile capire, il passaggio ad Alma si è configurato non come una semplice "evoluzione", ma come una vera e propria "rivoluzione"[10] nella misura in cui ha imposto, in virtù della sua complessa architettura, un ripensamento totale delle modalità di gestione delle risorse e dei servizi.

Proprio la complessa e sfidante architettura di Alma si è rivelata anche un potente motore per la cooperazione. La gestione centralizzata, basata per il data management su metadati e flussi di lavoro unificati, la stessa riduzione delle possibilità di personalizzazione cui si è accennato sopra, hanno di fatto favorito l'integrazione di flussi di lavoro, la condivisione delle risorse e lo sviluppo di prassi comuni.

Alma, in definitiva, ha fornito non solo una piattaforma tecnologica avanzata, ma anche uno spazio condiviso di lavoro, confronto e innovazione, dimostrando come un sistema gestionale possa diventare leva strategica per la cooperazione istituzionale e la costruzione di un ecosistema collaborativo.

viene respinta in considerazione delle criticità che esso avrebbe generato nei rapporti con SBN e BNI, che adottano UNIMARC. Il tema è riproposto nel 2016, in occasione della migrazione ad Alma, ma anche in questo caso prevale l'esigenza di mantenere il colloquio con SBN.

[7] L'annuncio della nascita di Alma a gennaio 2011 (https://exlibrisgroup.com/press-release/ex-libris-announces-the-cloud-based-alma-library-management-service/), segna il cambiamento strategico dell'azienda rispetto ad Aleph che da quel momento, pur senza essere ufficialmente dismesso, non conosce più sviluppi sostanziali.

[8] Interessante disamina delle caratteristiche più innovative di Alma nello stesso anno in cui ne viene annunciato il lancio nell'analisi in (Breeding 2011).

[9] Per maggiori dettagli sulle caratteristiche di Alma si veda la descrizione presente nelle pagine della documentazione di Ex Libris: https://knowledge.exlibrisgroup.com/Alma/Product_Documentation/010Alma_Online_Help_(English)/010Getting_Started/010Alma_Introduction/010Alma_Overview.

[10] Per il confronto tra sistemi gestionali "evolutivi" e sistemi gestionali "rivoluzionari" si rimanda a (Breeding 2019).

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-682

Il GDL Standard catalografici e dialogo con SBN

L'esigenza di adottare prassi lavorative nuove, più adatte a raccogliere le sfide presentate dall'evoluzione tecnologica del sistema gestionale, e, più in generale, da un contesto in rapida evoluzione, è subito avvertita nell'ambito di ITALE. Nel 2015, infatti, ITALE promuove la creazione di gruppi di lavoro tematici, con la finalità di favorire la crescita professionale e lo scambio di esperienze. Il primo gruppo[11] di cui viene proposta la formazione è proprio quello dedicato agli standard bibliografici e al colloquio SBN-MARC-Alma/Aleph la cui nascita viene formalizzata nel 2016. Il gruppo raccoglie fin dall'inizio numerose adesioni, a testimonianza dell'interesse, evidentemente molto sentito, nei confronti del tema dei protocolli di catalogazione e dell'integrazione con l'Indice SBN.

Non casualmente questa fase di rinnovato fervore operativo funge anche da stimolo a valorizzare il rapporto di collaborazione con ICCU: all'inizio del 2016 un incontro tra la presidente ITALE Fernanda Canepa e la neoeletta direttrice dell'Istituto, Simonetta Buttò, getta le basi per il rafforzamento della collaborazione e la pianificazione di attività congiunte, tra cui il prestito ISO-ILL e il miglioramento dell'integrazione tra le piattaforme Ex Libris -Aleph500 e Alma- e SBN.

Nel 2017 viene perfezionato il protocollo d'intesa già esistente fra ITALE e l'ICCU per lo sviluppo dei percorsi progettuali consolidati nel corso degli anni, come l'integrazione con il protocollo SBN-MARC e la partecipazione alla cooperazione ILL/SBN. Significativamente in questo contesto di interlocuzioni tra ITALE e l'ICCU si delinea anche l'apertura verso nuove possibilità di cooperazione favorite proprio dall'introduzione di Alma presso i clienti italiani di Ex Libris.

Si colloca in questa fase anche la richiesta da parte di ICCU di adeguare i Poli alle versioni più recenti del protocollo SBN-MARC (2.02 e 2.03) cui Ex Libris risponde manifestando la propria disponibilità ad effettuare tale adeguamento solo per la piattaforma Alma,[12] ciò che rende di fatto obbligatorio il passaggio al nuovo gestionale per le istituzioni aderenti a SBN. Il primo polo ad inaugurare l'ingresso in SBN di Alma nel 2017, è stato quello dei Sistemi bibliotecari toscani SBT, adottando il livello 3 di adesione, il primo per il quale Alma aveva ottenuto nello stesso anno la certificazione.[13] Sono poi seguiti, nei due anni successivi, altri quattro poli: dapprima il polo SGE (Sistema Bibliotecario Università di Genova), già presente in SBN con Aleph, poi il polo UBG (Università di Bergamo), anch'esso ex Aleph ma di nuova costituzione in SBN al pari di SBT, entrambi aderenti a livello 3. Hanno quindi fatto seguito il polo PUV (Polo Universitario Veneto), ex Aleph, e il polo USM (Università Statale di Milano), proveniente da Sebina: questi ultimi hanno optato per l'adesione a livello 4, per la quale il software ottiene la certificazione nel 2019.

Alma, lo si è visto, è caratterizzata da un livello di centralizzazione tanto spinto da lasciare poco margine a personalizzazioni locali e da una aderenza più stringente ai formati standard. Questa

[11] Accanto al gruppo di lavoro su standard bibliografici e al colloquio SBN, la 56 Assemblea di ITALE del Novembre 2015, propone anche la creazione di un gruppo dedicato ai Discovery tool (Primo, Summon). Il numero dei gruppi di lavoro è destinato ad aumentare gradualmente nel corso degli anni fino ad arrivare all'assetto attuale nel quale sono attivi sette gruppi di lavoro (Alma, Alma Risorse Elettroniche. Alma Digital, Leganto, Discovery tool, Resource Sharing, Standard bibliografici e al colloquio SBN).

[12] La certificazione di Aleph è rimasta bloccata alla versione 1.9 del protocollo.

[13] Si veda a tale proposito il comunicato stampa di Ex Libris: https://librarytechnology.org/pr/22746/alma-di-ex-libris-consegue-la-certificazione-per-la-catalogazione-condivisa-sul-sistema-bibliotecario-nazionale.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-682

particolare architettura fa emergere con maggiore forza le discrepanze (già manifestatesi con Aleph come detto nel paragrafo precedente) tra UNIMARC e il protocollo SBN-MARC. Le versioni 2.02 e 2.3 del protocollo – quest'ultima rilasciata da ICCU nel 2018 – si presentano come più performanti delle precedenti nonché esplicitamente basate sul formato UNIMARC del quale recepiscono la semantica.[14] Ciò nonostante discrepanze continuano a sussistere creando difficoltà nell'integrazione con SBN. A questo si aggiungono i problemi legati al trattamento di alcune peculiarità catalografiche proprie di SBN, ma del tutto prive di significato in UNIMARC, quali ad esempio l'uso, nei nomi degli autori personali, dell'underscore o del cancelletto,[15] o anche l'uso degli asterischi nei nomi di ente. Anche l'adesione a livello 4 presenta criticità importanti legate soprattutto, come già era avvenuto per Aleph, alla ricezione degli allineamenti.

Del resto, anche lato Alma si rileva una gestione non sempre pienamente soddisfacente del formato UNIMARC. Nonostante quanto dichiarato nella documentazione di Ex Libris,[16] secondo cui Alma garantisce la piena compatibilità con gli standard internazionali, UNIMARC risulta un po' negletto rispetto a MARC 21, più curato da parte di Ex Libris, sia per quanto riguarda la completezza della sua implementazione in Alma, sia per quanto concerne lo sviluppo di nuove funzionalità progettate sempre in prima battuta unicamente per l'ambiente MARC 21 e solo in un secondo momento rese disponibili anche per gli altri formati.

In questo contesto di forte e continua tensione tra la non perfetta integrazione tra il protocollo SBN-MARC e Alma e l'esigenza di migliorare l'implementazione e la gestione di UNIMARC all'interno di Alma, si colloca l'attività del gruppo Standard catalografici e dialogo con SBN, attività che, a partire dalla seconda metà del 2017, cresce in volume facendosi sempre più ricca articolata. Essa si concretizza nella costante analisi delle problematiche via via emergenti e nel continuo confronto tecnico con Ex Libris tramite richieste di miglioramento funzionale inoltrate al supporto dell'azienda da parte dei vari membri del gruppo.

I risultati non sono mancati, seppur talvolta con tempi di realizzazione dilatati nel tempo. Il colloquio con Indice a livello 4 a partire dal 2022 registra continui e costanti miglioramenti, così anche le richieste di sviluppo relative a UNIMARC trovano, in buona parte, risposte che il gruppo e la comunità di ITALE ritengono soddisfacenti.

Agli ambiti di attività sopra descritti, se ne aggiunge poi un terzo relativo al miglioramento dell'integrazione tra i dati UNIMARC e Primo VE,[17] il nuovo Discovery tool che nel 2018 Ex Libris presenta alle istituzioni italiane come un innovativo modello di implementazione, destinato a sostituire le precedenti versioni di Primo (Total Care o Direct). L'esigenza imprescindibile per le biblioteche, collocate oggi in un ecosistema informativo sempre più vasto e complesso, di poter contare su un Discovery tool efficace, dà ragione dell'importanza del lavoro in questo ambito particolarmente bisognoso di attenzione e di interventi mirati. La mappatura di UNIMARC in Primo VE, infatti, si è rivelata lacunosa e carente tanto da rischiare di inficiare l'esperienza di ricerca e di

[14] https://www.iccu.sbn.it/it/SBN/catalogazione-e-manutenzione-del-catalogo-sbn/le-attivita-di-catalogazione-e-il-protocollo-sbnmarc/index.html.

[15] https://norme.iccu.sbn.it/index.php?title=Norme_comuni/Authority_file/Nomi_di_persona/Registrazione_di_authority/Trascrizione.

[16] https://knowledge.exlibrisgroup.com/Alma/Product_Materials/050Alma_FAQs/Metadata_Management/01_General.

[17] https://knowledge.exlibrisgroup.com/Primo/Product_Documentation/020Primo_VE/Primo_VE_(English)/010Getting_Started_with_Primo_VE/005Primo_VE_Overview.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-682

recupero dell'informazione da parte degli utenti finali.[18] Per migliorare questa situazione ITALE si è mossa in primo luogo tramite il gruppo di lavoro dedicato al Discovery tool[19] che ha poi in più occasioni interpellato il gruppo Standard catalografici e colloquio con SBN in virtù della sua specifica competenza in merito al formato UNIMARC. Il lavoro sul Discovery ha quindi agito come ulteriore attivatore di collaborazioni, sia in ambito nazionale, in seno all'associazione, che in ambito internazionale. Le questioni concernenti la mappatura e la visualizzazione dei campi UNIMARC in Primo VE sono state infatti anche al centro di collaborazioni con realtà europee che adottano il formato UNIMARC, in particolare con l'omologa associazione di utenti Ex Libris ACEF,[20] e con ABES[21] con cui sono stati aperti dialoghi e scambi stimolanti. Da queste cooperazioni sono scaturite le proposte che, grazie alla costante interlocuzione con Ex Libris, hanno condotto a progressivi miglioramenti nella integrazione del formato UNIMARC in Primo VE che oggi, pur ancora migliorabile, può essere considerata soddisfacente.

È importante sottolineare come, anche nell'ambito delle collaborazioni di respiro internazionale non sia mai stato perso di vista il rispetto della specificità nazionale italiana, e *in primis*, della specificità legata alla cooperazione con SBN, l'attenzione alla quale è sempre stata percepita come irrinunciabile. Si può dire, in questo senso, che le esigenze della comunità aderente a SBN hanno in qualche modo "dettato l'agenda" determinando scelte i cui esiti hanno coinvolto tutta la comunità italiana di utenti di Ex Libris, incluse quindi anche le istituzioni che hanno scelto di non collaborare con esso.[22] Questa impostazione non è stata percepita come un'imposizione, né tantomeno come un arbitrio, ma come la naturale conseguenza del ruolo svolto da ICCU nella sua qualità di agenzia catalografica nazionale. La sua funzione di coordinamento, indirizzo e garanzia della qualità catalografica rappresenta infatti un punto di riferimento irrinunciabile per l'intera comunità bibliotecaria italiana a prescindere dalla decisione di aderire alla catalogazione partecipata alimentando il catalogo collettivo.

[18] Un'analisi dettagliata delle problematiche legate allo standard UNIMARC in Primo VE in (Pastorini et al. 2020). Nel documento, approvato dalla 68. Assemblea dei Soci ITALE, sono descritti i problemi riscontrati. Quelli considerati bloccanti, tali cioè da impedire il buon esito della ricerca, sono: il cattivo funzionamento della navigazione tra record dovuta alla errata mappatura dei campi di legame 4XX, la mancanza della funzionalità di autocompletamento del titolo in fase di ricerca (attiva per il campo 245 $$a di MARC 21 e non attiva per il campo 200 $$a di UNIMARC), la mancanza degli indici per serie in ricerca avanzata (errata mappatura dei campi 225 $$a e 410 $$a). Sono descritti anche numerosi altri problemi ritenuti significativi ma non bloccanti.

[19] https://itale.igelu.org/gdl-discovery-tool/.

[20] https://a-c-e-f.org/.

[21] https://www.sudoc.abes.fr/cbs/.

[22] Un esempio interessante è quello che riguarda la gestione in Alma dei codici di relazione presenti nella Appendice B di UNIMARC/B al cui perfezionamento il gruppo Standard catalografici e dialogo con SBN ha lavorato, in sinergia con ExLibris, a partire dal 2023. L'elenco di codici caricato in Alma e quindi in uso presso tutte le istituzioni italiane clienti di Ex Libris, include, oltre a quelli previsti dallo standard, anche alcuni non standard sviluppati da ICCU e necessari a coloro che aderiscono alla catalogazione partecipata in SBN: Le norme di catalogazione in SBN, infatti, prescrivono come obbligatorio l'uso dei codici di relazione per il materiale musicale e per i testi per musica. L'uso è raccomandato per tutti gli altri tipi di materiale (https://norme.iccu.sbn.it/index.php?title=Norme_comuni/Collegamenti/Titolo-Nome/Codici_di_relazione). L'OPAC SBN, in coerenza con queste prescrizioni, nella ricerca avanzata offre la possibilità di filtrare i risultati in base al ruolo.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-682

Prospettive future di collaborazione

La nuova frontiera della collaborazione tra ITALE e il mondo SBN potrebbe collocarsi nel contesto ricco e stimolante dei Linked Open Data (d'ora in avanti LOD). I LOD sono chiaramente per Ex Libris un ambito strategico su cui concentrare il proprio impegno di azienda leader per i servizi bibliotecari, numerosi sono infatti i progetti di sviluppo previsti per il potenziamento delle funzionalità loro dedicate.[23]

A partire dal 2024[24] i principi LOD sono stati progressivamente incorporati in Alma, al fine di consentire l'esposizione e l'arricchimento dei dati bibliografici utilizzando gli standard del web semantico. Alma infatti supporta l'uso dei record BIBFRAME in tutti i suoi flussi di lavoro. Parallelamente permette l'arricchimento dei record MARC 21 tramite URI, rendendo così la catalogazione più interoperabile. Accanto agli sviluppi in Alma, ci sono anche quelli sul Discovery Primo, integrato con funzionalità LOD, come le schede informative e le pagine personali che aggregano dati autorevoli da fonti esterne.

Gli sviluppi finora implementati, come accennato in precedenza, hanno riguardato tuttavia esclusivamente il formato MARC 21 tra quelli "tradizionali". Per questo motivo, nella primavera del 2025, il team italiano di Ex Libris ha sollevato la questione dell'impiego dei LOD anche per il formato UNIMARC, proponendo uno studio volto a valutarne la fattibilità. L'ICCU è stato subito individuato come interlocutore imprescindibile, in quanto agenzia catalografica nazionale responsabile della gestione degli authority file del catalogo SBN. L'ICCU inoltre ha già maturato una solida esperienza nell'uso dei LOD di cui ha cominciato ad occuparsi fin dal 2013 per giungere negli ultimi anni "in ottemperanza alle politiche sui dati aperti che governano l'agire della pubblica amministrazione, [a rendere] disponibili i LOD e le API per l'interrogazione e il riuso dei dati dei servizi bibliografici nazionali." (Atturo e Ravelli 2024, 51).

La collaborazione tra ICCU e ITALE sul tema dei LOD, al momento in cui scriviamo ancora in fase embrionale e non formalizzata,[25] potrebbe rappresentare una significativa opportunità di dialogo e sviluppo congiunto per esplorare soluzioni tecniche condivise, che valorizzino l'infrastruttura catalografica nazionale e, al tempo stesso, aprano all'interoperabilità i dati bibliografici italiani in formato UNIMARC. In questo contesto, l'esperienza e il ruolo dell'ICCU risultano fondamentali per garantire coerenza metodologica, mentre ITALE può farsi promotore di un confronto tecnico e strategico tra gli utenti italiani di Alma ed Ex Libris, coinvolgendo anche istituzioni che, pur non aderendo al catalogo SBN, fanno parte della comunità nazionale e vedono nell'ICCU un punto di riferimento autorevole come agenzia catalografica nazionale e referente per lo standard UNIMARC. La collaborazione sui LOD sarebbe un'occasione per rinnovare il dialogo tra comunità professionali, rafforzando una visione comune sull'evoluzione della catalogazione e sull'apertura dei dati nel contesto globale.

[23] https://knowledge.exlibrisgroup.com/Alma/Product_Materials/010Roadmap/Linked_Open_Data.

[24] https://exlibrisgroup.com/announcement/linked-open-data-implementations-in-ex-libris-primo-and-alma/.

[25] Nel mese di luglio 2025 sono in corso interlocuzioni tra ICCU e ITALE. ICCU ha manifestato la propria disponibilità ad aprire un dialogo per delineare meglio la eventuale collaborazione e i suoi contenuti teorici e tecnici.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-682

Conclusioni

La collaborazione tra SBN e ITALE ha già dimostrato quanto sia prezioso il dialogo tra enti pubblici e comunità di utenti nella gestione dell'infrastruttura catalografica nazionale. Tuttavia, guardando al futuro, è evidente che l'efficacia di questa cooperazione dovrà misurarsi con sfide ben più ampie: la transizione verso modelli descrittivi compatibili con il web semantico e l'adozione di standard in grado di garantire interoperabilità e visibilità globale ai dati bibliografici.

UNIMARC ha rappresentato per decenni uno strumento solido e condiviso, ma la sua struttura presenta limiti di cui l'IFLA stessa ha mostrato di essere consapevole promuovendo l'iniziativa internazionale Future of UNIMARC[26] che mira ad una profonda trasformazione del formato con l'adozione del modello concettuale entità-relazione (ER) e l'armonizzazione con IFLA LRM[27].

Tutta la comunità degli utilizzatori di UNIMARC è chiamata ad avviare la riflessione su questo progetto per indagare le possibili traiettorie di evoluzione del formato valutando contestualmente la prospettiva di un suo futuro superamento in favore di soluzioni più flessibili e interoperabili.

Si tratta di un obiettivo collocato in un futuro certamente non prossimo ma che necessita, per potersi realizzare, di essere preparato.[28]

In questa ottica, la sinergia tra SBN e ITALE potrà costituire il terreno ideale per sperimentare soluzioni condivise, promuovere la formazione e accompagnare la comunità bibliotecaria italiana nella transizione verso nuovi paradigmi e nuove prassi di lavoro.[29]

La storia di ITALE è la storia di una comunità che ha saputo trasformare la cooperazione tecnica in collaborazione strategica e la condivisione di problemi in costruzione di soluzioni. ITALE si configura non solo come un aggregato di utenti di una medesima piattaforma tecnologica, ma come un vero e proprio laboratorio di riflessione e sperimentazione, capace di contribuire in modo significativo all'evoluzione dei modelli di gestione dei servizi e di trattamento delle informazioni guidato dall'obiettivo condiviso di migliorare l'esperienza degli utenti e di semplificare il lavoro dei bibliotecari.

ITALE rinnova quindi la propria disponibilità a collaborare attivamente alla costruzione e al rafforzamento della comunità SBN, nella consapevolezza che la qualità dei dati bibliografici e l'efficacia dei servizi informativi rappresentano elementi strategici per la conservazione e la valorizzazione del patrimonio culturale nazionale.

Solo così la rete bibliotecaria italiana potrà riconquistare un ruolo centrale nel dibattito internazionale, contribuendo attivamente alla costruzione di un web della conoscenza aperto e condiviso.

[26] https://www.ifla.org/g/unimarc-rg/projects/.

[27] Si veda la Focus Area 2 dell'Action Plan 2023-2025 del Permanent UNIMARC Committee https://repository.ifla.org/handle/20.500.14598/2918.

[28] Un'analisi sintetica ma esaustiva della situazione attuale caratterizzata dall'uso ampiamente consolidato di formati sviluppatisi grazie ad una storia pluridecennale e condivisi da una comunità molto vasta, unita all'indicazione di ciò che sarebbe necessario per favorire l'evoluzione verso una fase nuova in (Guerrini e Testoni 2022, 9-14).

[29] Si richiama qui il concetto di "*metanoia*" presente in (Guerrini e Testoni 2022) nonché, più diffusamente, in (Guerrini 2022, 31)

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-682

Riferimenti bibliografici*

Atturo, Valentina, e Elena Ravelli. 2024. "L'evoluzione dell'authority work in SBN. Dalle origini ad Alphabetica e prospettive future." *JLIS.it* 15 (1): 45-58. https://doi.org/10.36253/jlis.it-572.

Guerrini, Mauro. 2022. *Dalla catalogazione alla metadatazione. Tracce di un percorso*. Roma: Associazione italiana biblioteche.

Guerrini Mauro, Laura e Testoni. 2022. "Conversazione con Mauro Guerrini, autore del libro:" Dalla catalogazione alla metadatazione: tracce di un percorso" Roma: AIB, 2021." *Vedianche* 32 (1): 7-14. https://riviste.aib.it/vedianche/article/view/13753.

Breeding, Marshall. 2011. "Ex Libris Marks Progress in Developing URM." *Smart Libraries Newsletter* 31 (1): 3-6. https://librarytechnology.org/document/16117.

Breeding, Marshall. 2019. "Smarter Libraries through Technology: Evolution or Revolution?." *Smart Libraries Newsletter* 39 (5): 1–3. https://librarytechnology.org/document/24465/.

PUC (Permanent UNIMARC Committee). 2023. "Action Plan 2023-2025: UNIMARC Review Group". https://repository.ifla.org/handle/20.500.14598/2918.

Anna Maria Pastorini, Giuseppe Marino, Elena Fasola, Libera Marinelli, Alberto Rovelli, e Silvia Trevenzoli. 2020. "Primo VE e UNIMARC." *Zenodo*. https://doi.org/10.5281/zenodo.3928163.

* Data di ultima consultazione dei siti web: 31 luglio 2025.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-679

JLIS.it

Wikidata and SBN: An Assessment of Two Years of Work (2023–2025)*

Carlo Bianchini[(a)], Stefano Bargioni[(b)], Camillo Carlo Pellizzari di San Girolamo[(c)]

a) University of Pavia, https://orcid.org/0000-0002-6635-6371
b) Pontifical University Santa Croce, https://orcid.org/0000-0001-7679-2874
c) Scuola Normale Superiore, https://orcid.org/0000-0003-2699-1693

Contact: Carlo Bianchini, carlo.bianchini@unipv.it; Stefano Bargioni, bargioni@pusc.it; Camillo Carlo Pellizzari di San Girolamo, camillo.pellizzaridisangirolamo@sns.it
Received: 17 July 2025; **Accepted:** 24 October 2025; **First Published:** 15 January 2026

ABSTRACT
The article presents the theoretical foundations and the reconstruction of the intensive bibliographic control, correction, and reconciliation activities that have progressively strengthened the link between SBN authority records and Wikidata entities, particularly over the past two years. It provides a detailed account and analysis of the most recent interventions carried out through the collaboration between ICCU and the Wikidata community. Building on this experience, the paper proposes a set of key strategies to further enhance the representation of Italian cultural entities within the Semantic Web and the network of Linked Data. Two main directions are suggested to pursue this goal: developing reliable pathways of access to and from SBN in order to foster interoperability among information systems, and promoting awareness of best practices in metadata creation within ordinary cataloguing workflows.

KEYWORDS
Wikidata; SBN; Authority control; Authority data; Reconciliation.

Wikidata e SBN: un bilancio di due anni di lavoro (2023-2025)

ABSTRACT
L'articolo presenta i presupposti teorici e la ricostruzione dell'intensa attività di controllo bibliografico, correzione e riconciliazione, che hanno consentito che il legame tra le Voci di autorità di SBN e gli elementi di Wikidata si consolidasse progressivamente, in particolare nel corso degli ultimi due anni. L'articolo illustra e analizza in dettaglio gli interventi più recenti, effettuati grazie alla collaborazione tra l'ICCU e la comunità di Wikidata. Sulla base dell'esperienza maturata, vengono proposte alcune strategie fondamentali per rafforzare ulteriormente la presenza delle entità culturali italiane nel web semantico e nella rete dei dati collegati ("Linked Data"). Si suggerisce di perseguire questo obiettivo lungo due direttrici principali: offrire percorsi di accesso affidabili da e verso SBN, favorendo l'interoperabilità tra sistemi informativi, e promuovere la conoscenza delle migliori pratiche di metadatazione all'interno delle pratiche di catalogazione ordinaria.

PAROLE CHIAVE
Wikidata; SBN; Controllo di Autorità; Voci di autorità; Riconciliazione.

* Il testo è stato scritto in totale collaborazione e condivisione: tuttavia vanno ascritti a Stefano Bargioni il paragrafo 4, a Carlo Bianchini i paragrafi 1 e 2 e a Camillo Pellizzari i paragrafi 3 e 5.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-679

1. Introduzione

La profonda trasformazione della catalogazione – tra le attività tecniche più centrali del flusso delle informazioni bibliografiche all'interno delle biblioteche – verso la *metadatazione*, ovvero la 'catalogazione in era digitale' (Guerrini 2020; 2022; Bargioni et al. 2022) si basa, tra l'altro, sullo spostamento del focus dalle risorse bibliografiche, gli oggetti portatori di informazione (es. libri, periodici, articoli, opere di consultazione ecc.) risultanti da un processo di produzione editoriale, alle relazioni bibliografiche che le caratterizzano e che testimoniano quel processo, come quella di autorialità, di curatela, di traduzione ecc. Questo spostamento, che potrebbe sembrare solo di facciata, si fonda su una visione teorica espressa in un modello concettuale evoluto: IFLA LRM (Library Reference Model) (IFLA 2016; 2017; Bianchini 2017; Guerrini e Sardo 2018). In esso, come nei suoi precursori (IFLA Study Group on the Functional Requirements for Bibliographic Records 2009; IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR) 2009; IFLA Working Group on Functional Requirements for Subject Authority Records (FRSAR) 2010), si individuano le entità di interesse e se ne esplicano tanto gli attributi (o proprietà) quanto le relazioni. La riduzione del numero di attributi e l'aumento delle relazioni tra entità sono un aspetto centrale di questo modello e anche degli standard descrittivi internazionali che vi si rifanno, come RDA (JSC 2010; Bianchini e Guerrini 2014). Poiché per creare una relazione in modo efficace ed efficiente è indispensabile definirne con esattezza assoluta gli estremi, ovvero il soggetto e l'oggetto, diventano centrali la questione dell'identificazione delle entità coinvolte e il tema – vasto e complesso – dell''identità' (ovvero di stabilire principi per i quali un'entità non è più la stessa ma diventa 'altra'). Su questo secondo problema, la riflessione è ancora tutta agli inizi, come osserva Tiziana Possemato: "i concetti di 'entità' e di 'identità' in ambito catalografico […] non sono mai stati analizzati in relazione all'attività di identificazione e modellazione degli oggetti del dominio GLAM" (Possemato 2024, 29). Tuttavia, la metadatazione e l'approccio *entity modelling* richiedono un cambiamento di prospettiva nella registrazione dei dati caratterizzato dall'identificazione delle entità stesse, attraverso gli identificatori, strumenti tipici del web semantico (Manzoni 2022). Proprio rispetto all'importanza e al ruolo degli identificatori del web semantico, Wikidata (d'ora in poi WD) svolge una funzione fondamentale anche nell'ambito dei cataloghi.

2. Wikidata e il controllo di autorità

Identificare le entità è un compito fondamentale per creare le condizioni per raggiungere l'obiettivo del Controllo Bibliografico Universale (CBU), il progetto più ambizioso mai ideato in ambito biblioteconomico (Anderson 1974; 2000; Solimine 1995). Il CBU è stato la ragione dello sviluppo di tutti gli standard internazionali nell'ambito della catalogazione, dall'ISBD agli ICP, da FRBR a IFLA LRM. Tra i primi strumenti creati dalla comunità internazionale per condividere gli sforzi delle biblioteche mondiali nell'ambito del lavoro di autorità (Guerrini e Sardo 2003), c'è il Virtual International Authority File,[1] o VIAF (Tillett, Reyad, e Cristán 2006). Si tratta di un progetto basato sullo sforzo congiunto di molte biblioteche e agenzie bibliografi-

[1] https://viaf.org/.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-679

che nazionali – come la Library of Congress, la Deutsche Nationalbibliothek, la Bibliothèque nationale de France e, per l'Italia, l'ICCU con i dati di SBN – e realizzato da OCLC. Il VIAF è costituito da insiemi di forme di nomi – detti cluster – che raggruppano e collegano tra loro i punti di accesso autorizzati creati dalle agenzie nazionali in base alle regole in vigore nei diversi paesi. La creazione dei cluster si basa su un software che, tramite algoritmi, individua l'eventuale corrispondenza tra una voce prodotta da un'agenzia e un cluster già presente in VIAF, collegandola o creando per la voce un nuovo cluster. Questo processo complesso può generare molti errori (tipicamente, l'attribuzione di una forma del nome di una persona a un cluster che non la rappresenta o, al contrario, la sua mancata attribuzione al cluster corretto e la conseguente creazione di un cluster distinto). Il lavoro di controllo della qualità dei dati ottenuti con questo processo di clusterizzazione è particolarmente lungo e costoso: la tecnica più efficace ed economica si basa sulla comparazione tra i dati registrati nei cluster VIAF e altri dati di autorità che offrono altrettante garanzie di qualità. Il processo di clusterizzazione infatti prevede sempre un margine di errore, ma l'errore non si ripete sempre identico nei diversi dataset ed è questo che rende efficace il processo di confronto.

Dove trovare un altro insieme di dati di autorità che contiene nomi di persona di provenienza internazionale e intercontinentale? Poiché il VIAF è stato il primo progetto di controllo di autorità di dimensioni internazionali – che ha preceduto nel tempo anche l'International Standard Name Identifier[2] – la domanda non è puramente retorica. È indispensabile tentare un approccio che coinvolga il web semantico e in particolare un suo 'hub' fondamentale: Wikidata (WD). Anche se il VIAF e WD sono piuttosto diversi per obiettivi, ambito, approccio organizzativo e teorico, raccolta e gestione dei dati, entrambi svolgono un importante ruolo di reciproco controllo della qualità dei dati (Bianchini, Bargioni, e Pellizzari di San Girolamo 2021).

WD ha un background molto diverso da quello di altre istituzioni che si sono occupate del CBU dagli anni 70 del XX secolo. WD non si dedica infatti solo ai dati bibliografici o di autorità, malgrado il suo valore da questo punto di vista sia indiscutibile (Bianchini e Sardo 2022). Proprio per il suo approccio diverso, l'esistenza, l'uso e il riuso dei dati di WD hanno un impatto pratico e teorico molto forte sul modello che la comunità bibliotecaria ha del CBU. WD esemplifica inoltre un approccio differente all'identificazione e alla descrizione delle entità dell'universo bibliografico; infatti, in WD, la differenza teorica tra il punto di accesso autorizzato (perché preferito) e i punti di accesso varianti – che era già stata molto ridimensionata nel cluster del VIAF – sembra diventare sempre meno importante dal punto di vista pratico. Infine, con WD è stata stimolata anche la riflessione sulla descrizione bibliografica, perché l'approccio granulare nella registrazione dei dati (per triple o dichiarazioni) è piuttosto diverso e offre potenzialità che la descrizione tradizionale non consente.

Lo stretto legame tra WD e il mondo dei dati bibliografici si è manifestato in molti modi, ed esiste un'ampia letteratura in merito,[3] ma la sua filosofia aperta e l'adesione piena ai principi FAIR[4]

[2] https://isni.org/page/what-is-isni/: l'ISNI è uno standard ISO (27729:2025), utilizzato da numerose biblioteche, editori, database, e organizzazioni per la gestione dei diritti d'autore in tutto il mondo.

[3] Si rinvia alla bibliografia di WD su WD (https://w.wiki/EfQf) e alla bibliografia su Zotero del Gruppo DARIAH su Wikibase, WD e Digital Humanities (https://www.zotero.org/groups/5639268/dariah_dhwiki/library).

[4] https://www.go-fair.org/fair-principles/.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-679

favorisce molto la sperimentazione di soluzioni integrate con altri prodotti aperti. È il caso dell'estensione AuthorityBox per l'OPAC della Biblioteca della PUSC (Bargioni 2020), che si basa sul software open source Koha, ma anche dei sempre più frequenti casi di riuso dei dati, a partire da *Alphabetica* fino al progetto Parsifal di URBE (Pellizzari di San Girolamo 2024a).
Infine, la caratteristica di WD di identificare non solo le persone, ma qualsiasi oggetto o concetto del mondo reale (esseri viventi, luoghi, proprietà come colore, dimensioni e peso, concetti come amicizia, inflazione ed elettroforesi) la rende uno strumento utile per un confronto anche di strumenti di indicizzazione specializzati e generalisti, come una classificazione o il Nuovo Soggettario (Cencetti, Pellizzari di San Girolamo, e Viti 2025).

3. Wikidata e SBN. Una collaborazione consolidata

Come molti altri cataloghi nazionali, anche l'OPAC SBN si è avvalso delle potenzialità offerte dal confronto con WD per l'analisi della qualità dei suoi dati e l'individuazione di possibili strategie di miglioramento.
Un primo studio si era focalizzato sull'analisi della qualità e della quantità dei dati di autorità di SBN considerando due diversi set di dati e comparandoli a quelli disponibili su WD (Bianchini, Bargioni, e Pellizzari di San Girolamo 2022) al fine di stabilire se la qualità delle voci di autorità (d'ora in poi VID) dei vari livelli previsti da SBN – in particolare i livelli da 90 compreso in su – corrispondeva effettivamente ai requisiti di completezza e di qualità indicati da SBN stesso per quei livelli. Il caso di studio si basava su un principio fondamentale per la qualità dei dati, ovvero il confronto tra dataset diversi che descrivono, in tutto o in parte, lo stesso universo. Il confronto – o riconciliazione – tra dataset può avvenire tra SBN e WD, come nel caso di studio, ma anche con e tra qualsiasi altro dataset. Il risultato è garantito dal processo di riconciliazione, non solo dalla qualità dei dataset di partenza.
Dall'analisi della qualità dei VID e del potenziale offerto da WD come strumento di controllo per il miglioramento delle voci si è virtuosamente sviluppato un progetto di reciproca interconnessione tra WD e SBN.[5] Anche se già dal 2013 WD aveva cominciato a creare legami verso i VID tramite la proprietà P396 che li identifica in WD (Pellizzari di San Girolamo 2024b, 34), è stato nel 2023 che si è avuta una significativa svolta nel processo di riconciliazione, grazie ad alcune innovazioni cruciali da parte di ICCU nella gestione dei VID: nel luglio 2023 i VID di livello 51 e 71 sono stati resi disponibili nell'OPAC, in aggiunta a quelli di livello 90 o superiore già disponibili, con reindirizzamenti per i VID fusi e implementazione di LOD e API (Atturo e Ravelli 2024).
Di conseguenza negli ultimi due anni (metà 2023-metà 2025), grazie al notevole incremento sia del numero di VID disponibili nell'OPAC e quindi riconciliabili con WD, sia dell'interesse da parte di ICCU e della comunità di WD per il processo di riconciliazione, le collaborazioni e i progetti riguardo ai VID sono molto aumentati, con tangibili miglioramenti in entrambi i dataset.

[5] https://w.wiki/EfQs è la pagina WD che raccoglie tutti i materiali relativi al progetto, costantemente aggiornata.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-679

3.1 L'impegno di ICCU con Wikidata per SBN

Le innovazioni da parte di ICCU nella gestione dei VID non hanno riguardato solo l'apertura dei dati:

- nel dicembre 2024 è stata rinnovata la composizione del *Gruppo di lavoro per la gestione e la manutenzione dell'Authority file di SBN. Nomi;*[6] scopo di questo gruppo di lavoro è la revisione qualitativa dei VID e la realizzazione di progetti di bonifica mirati, nonché la discussione di *best practices* relative a tali bonifiche;
- nello stesso mese è stato anche nominato un Comitato di Coordinamento per la Gestione e manutenzione dell'Authority File di SBN;[7]
- nel marzo 2025 è stata pubblicata la normativa per il trattamento dei VID relativi agli enti.[8]

Il progetto più ampio portato avanti in questi due anni da ICCU in collaborazione con la comunità di WD, nell'ambito della consolidata collaborazione fra Wikimedia Italia e ICCU (istituzionalizzata tramite una serie di convenzioni firmate nel 2015,[9] 2018[10] e 2022,[11] quest'ultima in corso di validità), ha riguardato l'arricchimento di gruppi di VID con dati ricavati da WD: a partire dall'insieme dei VID registrati in WD, si considera un sottoinsieme di VID carenti di informazioni e in tali VID si inseriscono dati tratti da WD per rendere le persone rappresentate da tali VID univocamente identificabili. Alla fine della prima fase del progetto circa 16200 VID sono stati arricchiti nel luglio 2024.[12] Elena Ravelli, coordinatrice dell'Area di attività per l'elaborazione e diffusione degli standard e delle norme catalografiche, e per la didattica dell'ICCU, ha presentato il progetto durante i Wikidata Days Bologna 2024, un evento dedicato all'interazione tra WD e le biblioteche accademiche (Ravelli 2024). La seconda fase del progetto, che arricchirà un nuovo gruppo di VID, è in avanzata fase di svolgimento.

3.2 La comunità Wikidata per SBN

Anche sul versante della comunità WD, i progressi si sono dispiegati in varie forme.
Un aspetto importante ha riguardato il miglioramento degli strumenti a disposizione degli utenti per lavorare alla riconciliazione, sia manuale sia semiautomatica, dei VID con WD: lo sviluppo di strumenti come Wikilinker per SBN (settembre 2023), WikiPlayground (febbraio 2025) e VID-Q reconciliator (marzo-aprile 2025), di cui si tratta dettagliatamente nel par. 4, ha consentito di progredire molto nella riconciliazione.
Si sono inoltre svolti numerosi eventi, in presenza e online, per spiegare le modalità di riconciliazione tra SBN e WD e coinvolgere più volontari, soprattutto bibliotecari: si segnalano in particolare due workshop relativi a SBN e WD presso il Convegno delle Stelline (21 marzo 2024[13] e 13

[6] https://tinyurl.com/2r5t9k3r.
[7] https://tinyurl.com/2ah6a6he.
[8] https://tinyurl.com/t89a4xtt.
[9] https://tinyurl.com/c3pswe5f.
[10] https://tinyurl.com/432ufewz.
[11] https://tinyurl.com/3nh2rctz.
[12] https://tinyurl.com/mua4hd6p.
[13] https://w.wiki/EfQd.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-679

marzo 2025[14]) e un corso relativo a WD e il controllo di autorità presso il Servizio Bibliotecario di Ateneo dell'Università di Pisa (9-12 giugno 2025[15]); la riconciliazione tra SBN e WD è stata anche l'esercizio svolto da 14 studenti tra il novembre 2024 e l'aprile 2025 nell'ambito del corso di *Archivi digitali, digital curation and preservation* tenuto da Chiara Storti presso l'Università Telematica Internazionale UNINETTUNO, con un totale di 444 VID aggiunti a WD.[16]

Si è anche cercato di facilitare ai volontari di WD il processo di comunicazione di problemi nei VID: per questo nel marzo 2024 è stata creata in WD una pagina per segnalazioni di errore,[17] che si affianca all'indirizzo mail reso disponibile da ICCU per le segnalazioni di errore (ic-cu.AFnomi@cultura.gov.it) in modo da meglio distribuire il carico di segnalazioni.

Al tempo stesso, per facilitare ai bibliotecari che possono modificare i VID l'uso di WD per trovare in essi errori da correggere, sono state allestite delle query per verificare potenziali errori[18] e sono state redatte delle istruzioni per individuare tramite WD e risolvere i casi di VID duplicati da fondere.[19]

Tra fine 2024 e inizio 2025, nell'ambito della sua tesi magistrale di informatica umanistica (De Monaco 2025), è stato svolto da Sara De Monaco un progetto di creazione e revisione in WD delle quasi 500 persone descritte in (Frattarolo 1975) e di sistemazione dei loro principali omonimi, aggiungendo in tutti gli elementi ove possibile un VID; inoltre, tutti i 253 errori incontrati in SBN (soprattutto record bibliografici male attribuiti, VID erroneamente privi di qualificazione e VID duplicati da unire) sono stati raccolti in una lista di segnalazioni (De Monaco e Pellizzari di San Girolamo 2025), poi verificate e risolte.

[14] https://w.wiki/EfQb.
[15] https://w.wiki/EfQa.
[16] https://w.wiki/EfQZ.
[17] https://w.wiki/EfQX.
[18] https://w.wiki/EfQU.
[19] https://w.wiki/EfQV.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-679

3.3 Dati statistici

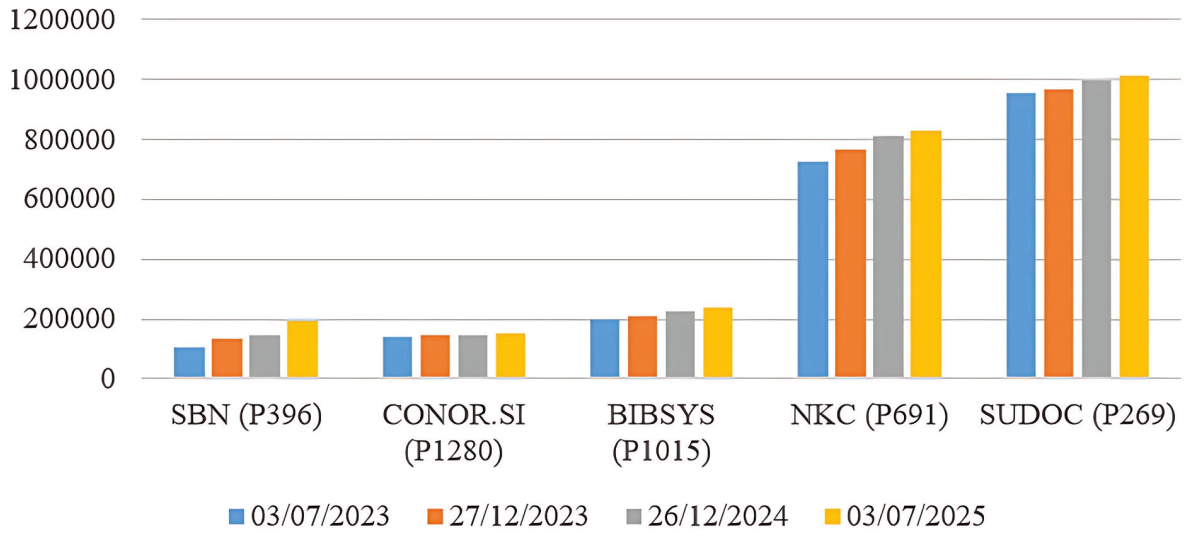


Tabella 1: numero di valori presenti in WD (come dichiarazioni) per cinque proprietà di *authority file* europei membri del VIAF[20]

Dalla Tabella 1 si nota come in due anni i valori di P396 siano cresciuti molto di più (+80%) rispetto a quelli di P691 (+14,4%) e di P269 (+5,7%); una parte significativa di questa crescita, soprattutto nei primi mesi del 2025, è dovuta all'efficacia del metodo VID-Q che sarà descritto nel par. 4.
La crescita nel numero di utenti coinvolti in WD nell'aggiunta di P396 è stata meno significativa: il numero di utenti che hanno aggiunto almeno 100 valori di P396 è cresciuto solo lievemente, da 51 a 60 (di cui, in entrambi i casi, 4 bot, cioè programmi automatici), tra il 1° aprile 2023[21] e il 1° luglio 2025;[22] al 1° luglio 2025 sono 1161 gli utenti totali che hanno aggiunto almeno un valore di P396 da quando la proprietà esiste (per confronto, sono 1420 per P691[23] e 3302 per P269[24]).
Se, quindi, il progresso negli strumenti a disposizione per l'abbinamento di P396 ha consentito di accelerare molto il ritmo di crescita dei valori di P396, sia grazie a metodi semiautomatici sia rendendo più efficiente il lavoro manuale degli utenti, è ancora necessario aumentare il numero di utenti coinvolti in questa attività.

4. Strumenti di riconciliazione e visualizzazione dati specifici per SBN

Il controllo di autorità svolto in SBN è un'attività complessa e articolata e saranno necessari tempi ancora lunghi per giungere alla pubblicazione completa dei dati di autorità prodotti in Italia, ma

[20] I dati sono tratti dalla cronologia di https://w.wiki/EfQj.
[21] https://tinyurl.com/4y4u7rw6.
[22] https://tinyurl.com/2a35ysu8.
[23] https://tinyurl.com/4r5xntd8.
[24] https://tinyurl.com/4jzzm8d3.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-679

è indispensabile che il bene comune rappresentato dai VID sia reso disponibile al meglio, integrandosi con quanto viene realizzato all'estero e incrementando il loro riuso nel web semantico in forma di Linked Open Data (Bianchini, Bargioni, e Pellizzari di San Girolamo 2021; 2022).
La soluzione più praticabile è quella di pubblicare e diffondere i VID attraverso gli hub del web semantico specialistici come VIAF o generalisti come WD, secondo il quarto principio dei Linked Data (Berners-Lee 2006). Tuttavia, il processo è molto differente dal punto di vista della fattibilità. Nel caso del VIAF, solo ICCU è in grado di svolgere questo compito, in quanto tra le fonti primarie di VIAF,[25] attraverso l'invio dei dati con cadenza all'incirca annuale (nell'aprile 2024 VIAF conteneva 200041 VID).[26]
Nel caso di WD, invece, la sua apertura alle modifiche e soprattutto l'uso della proprietà P396 per registrare i VID permettono a chiunque di contribuire all'identificazione e al collegamento dei VID con qualsiasi altra realtà produttrice di dati culturali che abbia identificatori della stessa entità.
In WD la proprietà P396 si può aggiungere con una modifica manuale,[27] ma, al momento esistono anche tre strumenti che ne facilitano l'aggiunta: WikiPlayground 2.0 (WPG), WikiLinker per SBN + AuthorityBox SBN (WikiLinker) e VID-Q reconciliator (VID-Q), le cui caratteristiche sono riassunte nella Tabella 2 (Appendice 1).

4.1 WikiPlayground

WikiPlayground (WPG) è un'applicazione web[28] e mobile friendly realizzata da Lorenzo Gobbo, bibliotecario addetto ai Servizi Digitali per le biblioteche dell'Università della Svizzera Italiana. Dopo aver aperto una sessione in WD nel browser, si esegue una ricerca su WD tramite una query SPARQL allo scopo di individuare un gruppo di autori presenti in WD che abbiano cittadinanza italiana,[29] identificativo VIAF e siano privi di P396. La query è personalizzabile, ovvero si possono modificare i criteri di ricerca degli elementi privi di P396. A partire dal gruppo di elementi estratti da WD mediante la query, viene effettuata una ricerca per nome su SBN e vengono presentati i possibili abbinamenti (Figura 1). Tra questi l'operatore, confrontando altri dati o le opere associate in VIAF e in SBN, può decidere di confermare la corrispondenza tra l'elemento WD e il VID o di scartarla. In caso di conferma, WPG aggiunge il VID come valore di P396 in WD, con il qualificatore "soggetto indicato come" (P1810) che registra la forma preferita in SBN.

[25] https://viaf.org/en/contributors?id=ICCU.
[26] Secondo il dump VIAF del 01/04/2024, file viaf-20240401-links.txt.gz; l'ultimo dump VIAF disponibile è di agosto 2024 (https://viaf.org/en/viaf/data).
[27] Il processo è descritto in questo video del GWMAB: https://www.youtube.com/watch?v=wndO45DWcpw.
[28] WPG è presentato in https://github.com/labaib/WikiPlayground2.0 ed è utilizzabile tramite https://labaib.github.io/WikiPlayground2.0/.
[29] Cittadinanza italiana include Regno d'Italia (pre 1946) e il caso di P396 "nessun valore" equivale alla presenza di P396.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-679

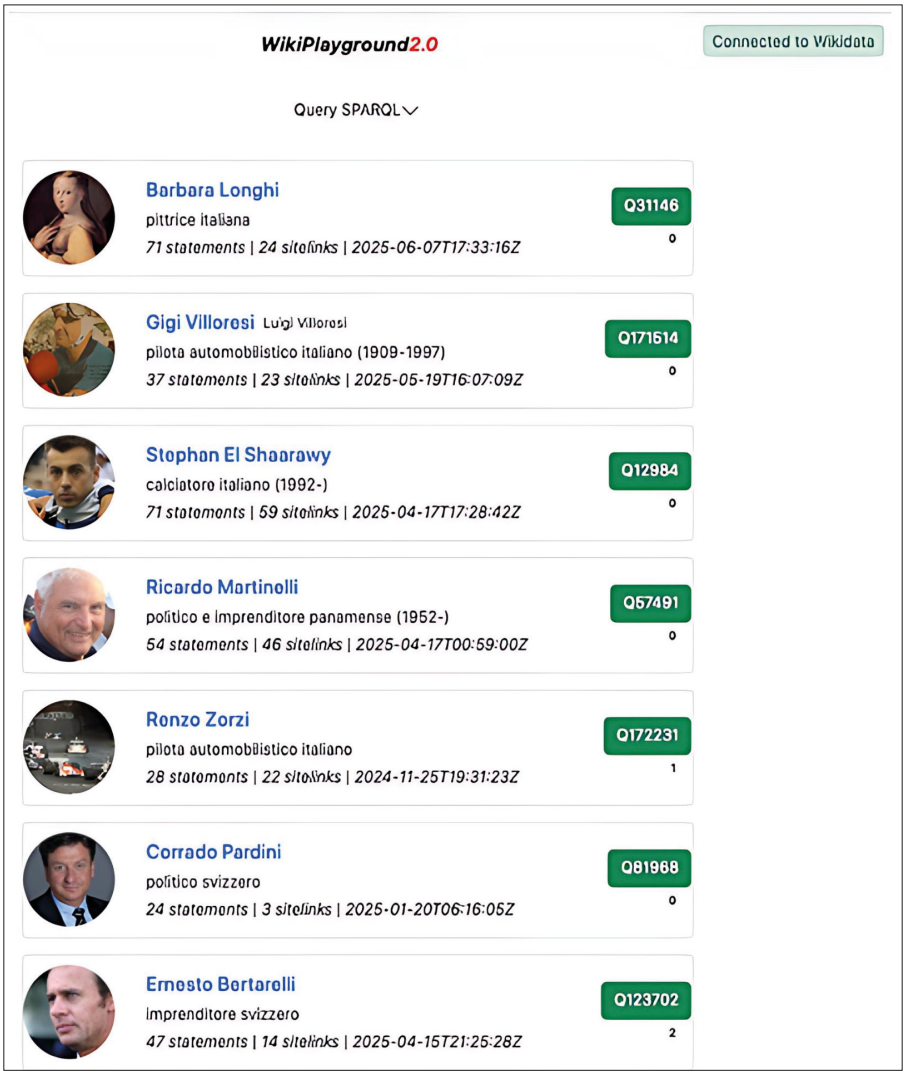


Figura 1. Abbinamenti potenziali proposti da WPG 2.0.

4.2 WikiLinker

WikiLinker è costituito da una coppia di programmi[30] richiamati dal proprio browser tramite un'estensione, come p.es. Code Injector,[31] che li esegue quando si sta visualizzando il risultato di una ricerca su Voci di autorità/Nomi nell'OPAC SBN.[32] La videata dei risultati include una prima colonna con il VID, copiabile con un click, una seconda con un link per mostrare i cluster VIAF

[30] WikiLinker per SBN (introduzione https://w.wiki/EfQN, codice https://w.wiki/EfQQ) e AuthorityBox SBN (codice https://w.wiki/EfQS).

[31] https://github.com/Lor-Saba/Code-Injector. Questa estensione non è più manutenuta; inoltre il browser Google Chrome, se aggiornato alle versioni più recenti, ne rifiuta l'installazione a causa di nuovi requisiti molto stringenti, motivati da ragioni di sicurezza; si può ancora installare in Firefox. Esistono altre estensioni simili a Code Injector, ma meno semplici da utilizzare.

[32] https://opac.sbn.it/voci-controllate-nomi.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-679

collegati alla forma preferita di SBN, e una terza riporta invece il link a WD nel caso in cui il VID sia già esistente in WD (Figura 2).[33]

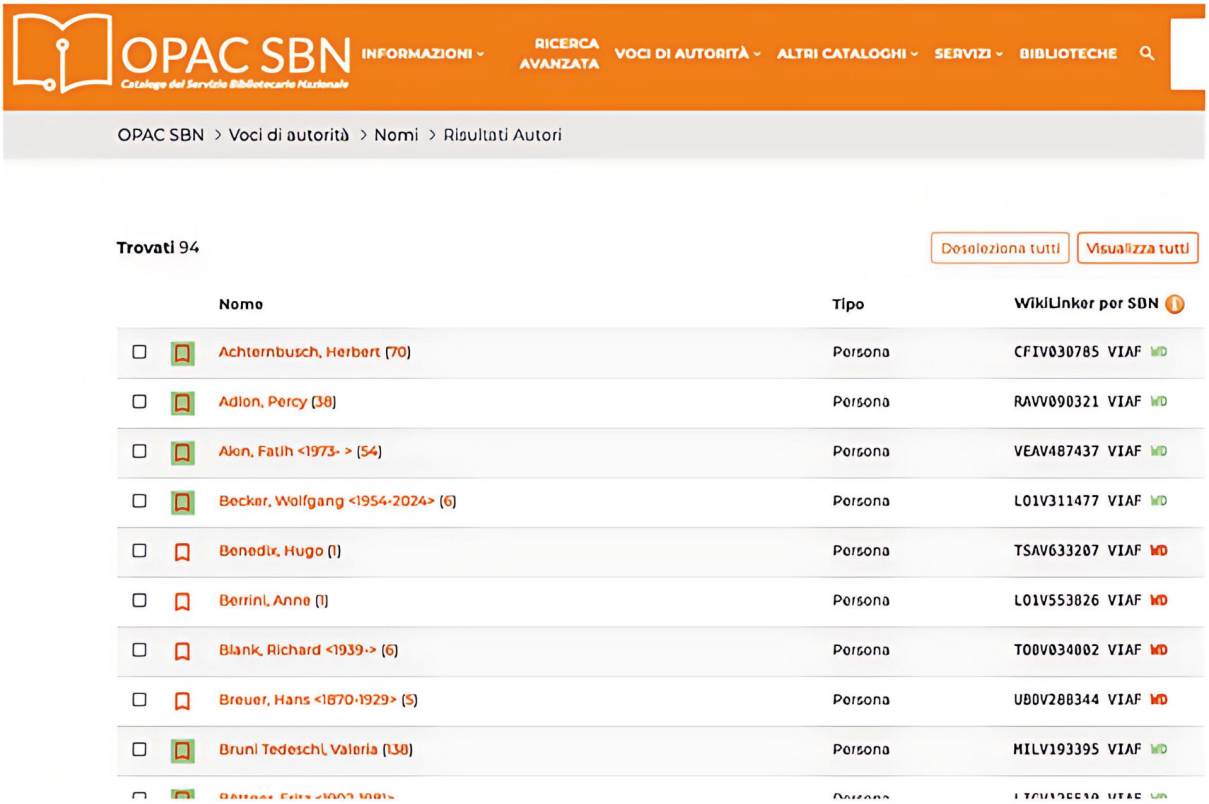


Figura 2. WikiLinker per SBN

L'utente può lavorare sulle voci non ancora associate a WD: il passaggio del mouse sulla voce visualizza il numero di pubblicazioni associate,[34] mentre il passaggio del mouse ma con il tasto "maiuscole" premuto mostra tutti i campi del VID in una finestra che include anche i titoli di 20 pubblicazioni.
In caso di corrispondenza tra VID e Item di WD, l'utente potrà aggiungere manualmente il valore del VID in WD (P396).
Se non c'è già una corrispondenza, l'utente potrà creare in modo guidato l'elemento WD mancante. Si procede con un click sulla forma preferita di SBN, che aprirà il VID corrispondente arricchito dall'AuthorityBox, cioè il secondo programma di WikiLinker.[35] Il riquadro dell'AuthorityBox ha diverse funzioni ma, in questo caso, sono disponibili due bottoni (Figura 3): uno per la ricerca

[33] Un tutorial di WikiLinker si trova in https://www.youtube.com/watch?v=YGMj8j6t8BE.
[34] Il numero viene visualizzato accanto alla voce e si può cliccare per mostrare tutte le opere associate.
[35] AuthorityBox è stato introdotto nel catalogo della Pontificia Università della Santa Croce nel 2019 e viene presentato in (Bargioni 2020) e discusso in (Bianchini 2024).

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-679

in WD tramite stringa del nome di un VID SBN,[36] e uno per la creazione dell'elemento WD una volta verificato che non esista, tramite l'utilizzo di QuickStatements (QS).[37]

Figura 3. Il VID SBLV223806 (Anna Scalera) con le aggiunte dell'AuthorityBox evidenziate nei riquadri blu

Premendo sul bottone "Crealo", viene composto il gruppo di istruzioni (o comandi) per la creazione dell'elemento (Figura 4). Si noti che i dati di luogo di nascita (Napoli) e occupazione/i (poetessa, docente) non possono fare parte delle istruzioni perché non facilmente isolabili e soprattutto perché non riconciliati con i corrispondenti elementi WD. Questa operazione sarà possibile solo manualmente in WD, una volta creato l'elemento. La composizione di P569 per la data di nascita segue il formato indicato da WD. Il riferimento indica che il dato proviene dall'OPAC SBN (P248 = Q105086090), in particolare dal VID SBLV223806, consultato (P813) in data 7 maggio 2025.[38]

[36] http://id.sbn.it/bid/SBLV223806.
[37] https://quickstatements.toolforge.org/.
[38] Inoltre P31 indica istanza di umano (Q5) e P21 genere femminile (Q6581072), tramite uno dei bottoni in basso a sinistra. La descrizione in italiano (Dit) è ottenuta modificando manualmente la nota informativa.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-679

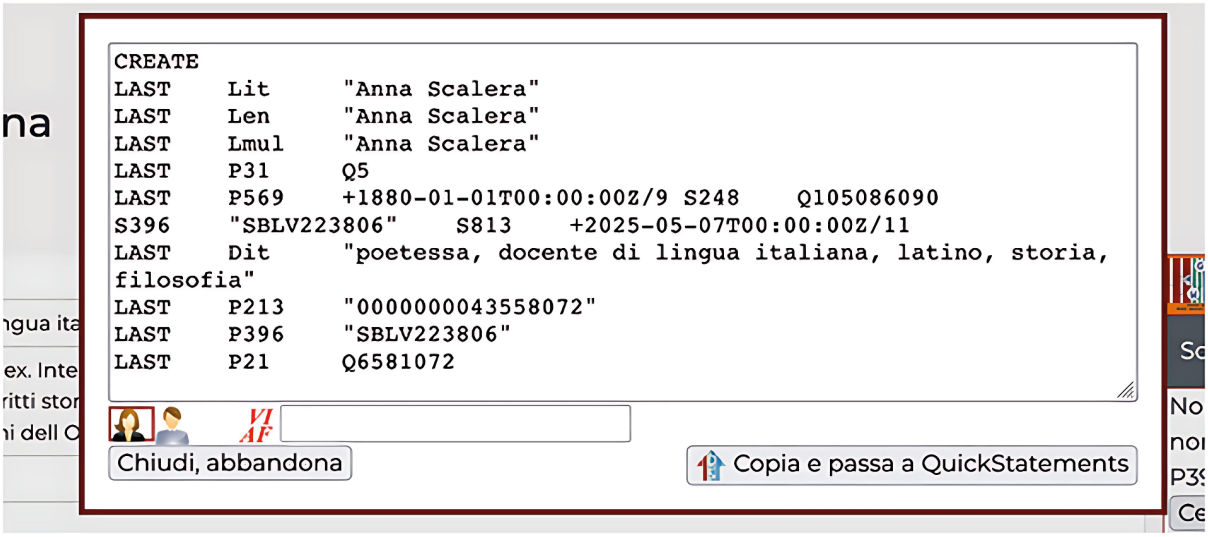


Figura 4. Le istruzioni di QuickStatements per la creazione di un nuovo elemento WD

4.3 VID-Q reconciliator

WPG e WikiLinker facilitano molto il lavoro di riconciliazione, ma la velocità operativa resta bassa perché trattano un caso alla volta. Considerata la quantità di lavoro necessaria, nemmeno una squadra di numerosi esperti può portare grandi risultati, se non in termini di qualità. D'altra parte, non è prudente delegare l'aggiunta di P396 ad operazioni del tutto automatizzate: per la riconciliazione dei nomi di autori devono essere valide due condizioni di corrispondenza: uguaglianza della forma preferita o di una forma variante con l'etichetta in italiano in WD e date di nascita e morte (anno o giorno). Se c'è corrispondenza solo per l'anno di nascita o di morte, occorre un ulteriore criterio di coincidenza.[39] Perciò, è stato creato un insieme di script in linguaggio Perl[40] – basati sulle API di ICCU-SBN[41] – ognuno dei quali associa un VID a un elemento Q di WD, da cui il nome VID-Q reconciliator o VID-Q sotto forma di comandi QS del tipo:

```
Qnnn P396 "VID"
```

Le tre API di Nomi sono *search*, *vid* e *browse* e hanno l'endpoint comune https://api.iccu.sbn.it/nomi/1.0.0/. In VID-Q sono utilizzate le prime due, soprattutto *search*, che ricalca la ricerca per campi disponibile in Voci di autorità/Nomi. Per ottenere abbinamenti sicuri, in base ai criteri già citati, sono stati preparati 5 script, uno per ogni tipo di ricerca. Nelle Tabella 3 e 4 (Appendice

[39] Questo tipo di riconciliazioni per associazione di attributi comuni tra entità presenti in database diversi è utilizzato anche in (Tillman e Oaxaca 2025).

[40] Su Zenodo è stato pubblicato lo script principale, che permette la ricerca di VID via API in modo analogo alla ricerca in Voci di autorità/Nomi: (Bargioni, Bianchini, e Pellizzari di San Girolamo 2025).

[41] https://api.iccu.sbn.it/devportal/apis. Le API richiedono un token di autenticazione (di tipo OAuth2; https://www.rfc-editor.org/rfc/rfc6749) basato su un account personale ottenibile gratuitamente, richiesto con una chiamata specifica, rinnovato ogni 3600 secondi e associato ad ogni chiamata a una API. Non vi è un limite di utilizzo delle API, né quantitativo né basato sulla frequenza delle chiamate. È buona pratica lasciare almeno un secondo tra una chiamata e l'altra, a tutela del lavoro ordinario dei server di SBN.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-679

1) sono riportati rispettivamente i metodi di ricerca utilizzati in VID-Q e, per ogni metodo, un esempio di abbinamento ottenuto, ma sono possibili altri tipi di abbinamento.
Va comunque notato che le parole chiave e soprattutto le date utilizzate nelle ricerche delle voci SBN si possono trovare ovunque nel VID e perciò non sempre sono pertinenti. Lo script Perl infatti verifica se il VID trovato rispetta il criterio di abbinamento ma controlla anche in quali campi del record SBN si trova il dato; perciò le voci cercate su WD sono solo una parte, e di questa gli abbinamenti sono una parte ridotta, se non a volte nulla.

4.4 Dati quantitativi sugli abbinamenti

Gli abbinamenti del tipo 1 si basano su ricerche per coppie di decenni, di nascita e di morte, dal 1700 in poi, per un totale di 1587 abbinamenti (alcuni esempi in Tabella 5, Appendice 1).
Gli abbinamenti del tipo 2 si basano sulle decadi di nascita da 1700 fino a 1799 abbinate a decadi di morte da 1720 a 1899, sulle decadi di nascita da 1800 fino a 1899 abbinate alle decadi di morte da 1820 a 1999, e sulle decadi di nascita da 1900 fino a 1999 abbinate alle decadi di morte da 1920 a 2025, per un totale di 1587 abbinamenti (rispettivamente 186, 1016 e 385).
Gli abbinamenti del tipo 3, basati su nome e data con precisione superiore ad anno, sono stati applicati a 106 mesi, da gennaio 1900 a dicembre 2005, e sono raggruppati per decadi nella Tabella 6 (Appendice 1).
Per il tipo 4, basato su nome, anno di nascita e parola chiave dell'occupazione, ma solo per alcuni casi (colonna 1 della Tabella 7, Appendice 1), viene presentata una statistica più articolata, che permette di comprendere meglio il peso delle operazioni svolte da ogni passaggio.
La ricerca su SBN, a volte eseguita per troncamento (storic* per storico o storica), restituisce il numero di voci riportate in colonna 4. Il filtro permette di effettuare la ricerca su WD su un numero un po' inferiore di casi (colonna 5). Gli abbinamenti ottenuti sono invece quasi sempre molto inferiori (colonna 6), per varie cause: elemento non presente in WD, nonostante la ricerca per uguaglianza della forma scelta in SBN con l'etichetta in italiano "it", o inglese "en" o multilingua "mul",[42] o con i possibili alias;[43] assenza in WD o mancata coincidenza delle date o dell'occupazione; abbinamento già presente in WD. Il totale degli abbinamenti ottenuti in questa fase è 9513; sono state trovate 173855 voci di SBN ed esaminati 151084 elementi WD.
Sempre per gli abbinamenti di tipo 4, ma con il confronto con la proprietà P611, sono state eseguite anche alcune ricerche per verificarne l'efficacia, riportate nella Tabella 8 (Appendice 1).
Benché la casistica delle prime tre colonne non sia certo esaustiva, la Tabella 8 mostra come si ottengano meno abbinamenti con questo ulteriore passaggio. Ma va comunque tenuto presente che ogni ricerca era seguita dalla registrazione in WD degli abbinamenti ottenuti, e quindi diminuiva via via la possibilità che una ricerca apparisse come efficace se l'abbinamento era già stato registrato.
Il criterio di abbinamento 5, basato sulla presenza di un identificativo di ambito italiano (IDita), ha dato risultati più significativi rispetto ai precedenti (Tabella 9, Appendice 1). Anche in questo caso, ogni estrazione poteva vedere ridotti gli elementi abbinati in funzione dei risultati delle precedenti.

[42] https://w.wiki/CjFz.
[43] Non sono state esaminate le eventuali varianti riportate in SBN, per non moltiplicare il lavoro. Potrebbe essere un ulteriore metodo di ricerca, se si riesce a individuare facilmente le voci di SBN con varianti.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-679

5. Conclusioni e prospettive future

Le diverse procedure di abbinamento tra VID di SBN ed elementi WD illustrate suggeriscono alcuni accorgimenti per migliorare ulteriormente il procedimento di riconciliazione, tanto relativi all'accessibilità e alla riusabilità dei dati di SBN che le procedure di creazione e modifica dei VID all'interno di SBN.

La prima necessità è che anche gli URI dei VID di livello 05 siano resi visibili nella sezione Voci di autorità; attualmente, infatti, per i VID di livello 05 WD è costretta a linkare a https://opac.sbn.it/risultati-autori/-/opac-autori/detail/**VID**?core=autoriall anziché a https://opac.sbn.it/nome/**VID** proprio per poter creare link.

Un secondo problema è che in Interfaccia Diretta (di seguito InD) mancano alcune funzioni che sono disponibili nell'interfaccia di ricerca Voci di autorità/Nomi e che sono molto utili per trovare VID da bonificare (es. cercare stringhe in tutti i campi del VID consente di trovare VID aventi, nella nota informativa, informazioni da spostare in altri campi, come l'ISNI[44]).

Un altro miglioramento del processo riguarda l'aspetto legale del riuso dei dati contenuti nei VID; è necessario che venga indicata la licenza con cui sono rilasciati. Se si adottasse la licenza CC0, come per l'Anagrafe delle Biblioteche Italiane,[45] sarebbero possibili importazioni massive di dati in WD.

Infine, il riuso dei dati e la realizzazione di analisi statistiche sarebbero molto avvantaggiate dal rilascio dei dump periodici dei VID, che, se contenessero anche il livello di autorità di ciascun VID (al momento non ottenibile da OPAC SBN), potrebbero fornire informazioni utili anche sulla distribuzione dei VID per livello e su come rendere la gestione dei livelli più efficiente. Se i dump contenessero anche i VID reindirizzati, si potrebbero usare periodicamente per aggiornare i VID in altri database.

Faciliterebbe il riuso dei dati anche la creazione di un endpoint SPARQL che permettesse di interrogare l'OPAC SBN, o almeno i VID, sul modello di BNF[46] e IDREF.[47]

Riguardo alle procedure di creazione e modifica dei VID (tramite InD e i gestionali di polo), sarebbe necessario implementare alcuni campi UNIMARC. Per esempio, per migliorare l'identificabilità delle persone descritte nei VID, sarebbe utile introdurre un campo analogo al 670 ("source data found") del MARC21 usato dalla Library of Congress per indicare nelle sue voci di autorità quale è stato il record bibliografico per il quale la voce è stata originariamente creata; stante il manuale UNIMARC Authorities (IFLA 2024, s.v. 810), il campo "source data found" in UNIMARC è l'810 (utilizzato ad es. nelle voci di autorità della BNF); tuttavia in SBN il campo 810 è usabile solo per indicare la presenza della persona nei repertori; l'indicazione del record bibliografico originariamente usato si potrebbe quindi inserire o nella nota del catalogatore (campo UNIMARC 830) o in un nuovo campo appositamente introdotto.

Per rendere possibile strutturare identificativi diversi dall'ISNI (che è inserito nel campo UNIMARC 010), sarebbe da introdurre il campo UNIMARC 017 ("other identifier") – invece che

[44] https://opac.sbn.it/risultati-autori?core=autori&item%3A1016%3AKeywords=ISNI.
[45] https://anagrafe.iccu.sbn.it/it/open-data/.
[46] https://data.bnf.fr/sparql/.
[47] https://data.idref.fr/sparql.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-679

utilizzare il campo 810 Fonti o 830 Nota del catalogatore –, ripetibile in modo da ospitare più identificativi (es. nello stesso VID un ID WD, un ID VIAF e un ID ORCID) facilitandone il riuso. Sempre in un'ottica di miglior riuso dei dati, sarebbe utile disporre di campi più granulari per suddividere le informazioni che al momento vengono inserite in forma testuale nel campo 300 Nota informativa (eventualmente separate da "//"), come il campo 104 Principali date dell'entità per la datazione. Sarebbe utile prevedere inoltre: 1) uno o due campi appositi per luogo di nascita e luogo di morte (p. es. il campo 301$a per luogo di nascita e 301$b per luogo di morte), registrando i luoghi come entità identificate in database/vocabolari controllati (es. Geonames), così da garantirne l'identificazione e da migliorarne la traducibilità e l'interoperabilità; 2) un campo apposito, ripetibile, per le occupazioni/professioni, anch'esso preferibilmente con valori attestati in vocabolari controllati (es. nel Thesaurus del Nuovo Soggettario).

Vista l'importanza dell'ISNI in quanto standard ISO e anche unico identificativo inseribile nei VID in modo strutturato, sarebbe indispensabile poterlo utilizzare per specifiche funzioni di mantenimento dei dati: cercare un VID a partire da un ISNI; estrarre periodicamente ISNI da WD, come già vengono annualmente estratti dal VIAF; aggiornare periodicamente in modo massivo gli ISNI reindirizzati o cancellati; indagare la possibilità per i catalogatori di SBN, o almeno a una parte di essi, di creare direttamente gli ISNI mancanti, visto che ICCU è tra le ISNI Registration Agencies,[48] così da dotare soprattutto gli autori italiani di un ISNI in modo più sistematico.

L'uso attuale dei livelli di autorità per i VID dovrebbe essere ripensato affinché essi possano più efficacemente svolgere il loro ruolo positivo di proteggere i VID di buona qualità da possibili modifiche improprie, senza invece svolgere un ruolo negativo di ostacolo al miglioramento di VID di scarsa qualità o semplicemente 'vuoti' (ossia contenenti solo il campo 200 compilato). Purtroppo attualmente non è tecnicamente possibile abbassare di livello gruppi di VID (es. quelli di livello 90 e 95 'vuoti'); anche per questo, tuttavia, sarebbe importante stabilire norme sull'uso dei livelli in modo che, almeno per i VID lavorati in futuro, il livello scelto dal catalogatore sia proporzionale all'insieme dei campi compilati, e VID che sicuramente avranno bisogno di modifiche future (es. quelli di persone viventi) vengano posti a livelli che consentano a un adeguato numero di persone di apportare tali modifiche, riducendo così i possibili colli di bottiglia.

Si suggerisce infine di creare una procedura ufficiale tramite la quale coloro che già catalogano in SBN possano richiedere accesso in visualizzazione, e soprattutto in modifica, a InD, e al contempo di fornire dei corsi di formazione all'uso di InD con periodicità regolare, in modo da coinvolgere nella bonifica dei VID un maggior numero di persone dopo adeguata formazione.

Le prospettive future per la collaborazione tra WD e l'*authority file* di SBN si collocano nei tre ambiti fondamentali che si sono consolidati in questi anni: l'abbinamento reciproco, l'arricchimento reciproco, la verifica incrociata.

L'abbinamento tra elementi WD e VID ha visto un'accelerazione negli ultimi due anni grazie soprattutto a efficaci metodi di abbinamento automatico (VID-Q) e agli strumenti forniti agli utenti per rendere il lavoro manuale più efficiente (WPG, WikiLinker). Per aumentarne ulteriormente la velocità è necessario un approccio con metodi automatici sull'intero insieme dei VID – che sarebbe facilitato dalla disponibilità di dump periodici – e convogliare più forze verso il lavoro

[48] https://isni.org/page/isni-registration-agencies/.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-679

di abbinamento manuale, pressoché indispensabile per i VID privi di informazioni identificative. WD e SBN possono beneficiare molto dai dati l'una dell'altro: per SBN, i risultati del primo arricchimento di VID 'vuoti' (concluso nel 2024) e del secondo (in corso nel 2025) sono buoni e si sta stabilizzando il *workflow* per farlo diventare una procedura periodica; per WD sarebbe possibile creare massivamente nuovi elementi a partire dai VID di buona qualità ma per farlo sarebbero necessari dump periodici e soprattutto l'attribuzione della licenza CC0 ai VID, sul modello di quanto fatto dalla Biblioteca Nazionale della Repubblica Ceca a partire dal 2019 (Jansová, Maixnerová, e Šťastná 2024, 8–10).

I dati di WD e di SBN hanno ovviamente delle divergenze: analizzarle può permettere di rintracciare errori di abbinamento, nonché errori fattuali da entrambe le parti. Tuttavia, è difficile per ragioni di tempo aggiungere l'analisi di queste divergenze al flusso di lavoro ordinario dei bibliotecari impegnati quotidianamente nella creazione e manutenzione dei VID e sarebbe più fattibile inserire queste attività di controllo della qualità dei dati nell'ambito di progetti tematici che riguardino un insieme coerente di persone da descriversi in WD, come p. es. gli autori del *Dizionario degli scrittori italiani contemporanei pseudonimi* di Renzo Frattarolo (De Monaco 2025); rendere l'identificazione di una persona in SBN parte strutturale di ogni progetto di ricerca che descriva in WD un corpus di persone legate all'Italia consentirebbe di coinvolgere anche i ricercatori nella manutenzione dei VID SBN, rendendoli quindi patrimonio comune non solo al mondo delle biblioteche e iniziando una loro possibile apertura anche ad altre istituzioni culturali e al mondo della ricerca, secondo il modello del GND. In generale, tanto più l'*authority file* di SBN riuscirà a diventare un punto di riferimento nell'ambito dei Linked Open Data anche al di fuori del mondo delle biblioteche, tanto più aumenterà il numero e la diversità dei suoi utenti, e di conseguenza saranno sempre più individuati e corretti errori e imprecisioni, tanto più ne aumenterà la qualità, instaurando così un circolo virtuoso.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-679

Appendice 1

Nome completo	Nome breve	Operatività	Ambito di partenza	Interfaccia	Linguaggio
WikiPlayground	WPG	manuale	WD	web	HTML e JavaScript
WikiLinker per SBN + AuthorityBox SBN	WikiLinker	manuale	SBN	web	JavaScript
VID-Q reconciliator	VID-Q	batch	SBN	linea di comando da terminale	Perl

Tabella 2 - Metodi di riconciliazione SBN-WD disponibili

Criterio di abbinamento	Esempio di ricerca	Note	
1. nome e due anni	191* 199*	nati tra gli anni 1910 e 1999	anni inclusi nel qualificatore <an-am>
2. nome e due anni	191* 199*	nati tra gli anni 1910 e 1999	anni ricavati dal campo Datazione
3. nome e data con precisione superiore ad anno	1910mm*	nati nel mese “mm” dell’anno 1910	anno mese e giorno nel campo Datazione
4. nome, anno di nascita e parola chiave dell’occupazione	19* chimic*	chimico/a del XX secolo	anno e occupazione ovunque
5. presenza di identificativo di ambito italiano (IDita)	query WD che estrae elementi con IDita ma senza SBN	il risultato della query viene analizzato da uno script apposito	le fonti degli identificativi sono indicate in Tabella 9

Tabella 3. Tipi di ricerca utilizzati in VID-Q

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-679

Criterio di abbinamento	Comando utilizzato	SBN	Wikidata
1. nome e due anni da qualificatore	perl cerca_range_date.pl 193 200	RMSV963070 Arnold, Hans Joachim <1932-2006>	Q17322103 Hans-Joachim Arnold P569 31 mar 1932 P570 20 feb 2006
2. nome e due anni da campo Datazione	perl cerca_data_datazione.pl 193 200	UBOV397234 Advis, Luis Datazione: 1935-2004	Q3038827 Luis Advis P569 10 feb 1935 P570 9 set 2004
3. nome e data di nascita con precisione superiore ad anno, da campo Datazione	perl cerca_data_datazione.pl 19300*	UBOV624476 Fain, Veniamin Moiseevič Varianti Fain, Benjamin Datazione: 1930.02.17-2013.04.15	Q2897209 Benjamin Fain P569 17 feb 1930
4. nome, anno di nascita e parola chiave dell'occupazione	perl cerca_data_occup.pl chimic Q593644 19	SBLV286830 Asinger, Friedrich Datazione: 1907-1999 Nota informativa: Chimico …	Q84786 Friedrich Asinger P569 26 giu 1907 P106 chimico (Q593644)
5. presenza di identificativo di ambito italiano (IDita)	perl q_IDita.pl file_elementi_estratti_dalla_query	UM1V037628 Carolina Santoni 1808 1878	Q100149180 Carolina Santoni

Tabella 4. Esempi di abbinamenti ottenuti con VID-Q

Decade anno nascita	Decade anno morte	Abbinamenti ottenuti
170*	173*	0
185*	192*	29
191*	199*	25
195*	202*	2

Tabella 5. Alcuni abbinamenti tra nome, date nascita e morte per decadi presenti nel qualificatore

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-679

Decade	Abbinamenti ottenuti
1900-1909	23
1910-1919	19
1920-1929	61
1930-1939	110
1940-1949	115
1950-1959	87
1960-1969	45
1970-1979	25
1980-1989	10
1990-1999	5
2000-2005	1
TOTALI	**501**

Tabella 6. Conteggi degli abbinamenti di tipo 3

Parola occupazione cercata in SBN	Q occupazione (valore di P106)	Data nascita troncata (secolo)	Trovati in SBN	Cercati in WD	Abbinamenti ottenuti	Durata minuti
pittore	Q1028181	19	698	660	69	19
pittrice	Q1028181	19	1006	871	85	20
sacerdote	Q42603	18	2001	1646	15	45
sacerdote	Q42603	19	5113	4144	46	110
filolog*	Q13418253	18	771	668	20	19
filolog*	Q13418253	19	2807	2481	49	77
giornalista	Q1930187	19	10986	9781	971	314
giornalista	Q1930187	18	2115	1806	42	51
traduttore	Q333634	19	3008	2628	203	73
traduttrice	Q333634	19	1537	1323	93	38

Tabella 7. Conteggi degli abbinamenti di tipo 4

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-679

Parola relativa all'ordine religioso	Q ordine religioso (valore di P611)	Data nascita troncata (secolo)	Trovati in SBN	Cercati in WD	Abbinamenti ottenuti	Durata minuti
gesuita	Q36380	15	251	203	11	6
gesuita	Q36380	16	543	431	13	13
gesuita	Q36380	17	402	325	4	9
gesuita	Q36380	18	437	378	31	10
gesuita	Q36380	19	898	770	49	21
domenican*	Q131479	14	66	47	1	1
domenican*	Q131479	15	118	74	1	2
domenican*	Q131479	16	104	61	1	3
domenican*	Q131479	17	86	61	1	3
domenican*	Q131479	18	141	104	7	3
domenican*	Q131479	19	323	260	22	10
francescano	Q913972	15	127	90	1	3
francescano	Q913972	16	184	120	2	3
francescano	Q913972	17	192	141	1	4
francescano	Q913972	18	451	401	6	11
francescano	Q913972	19	770	706	13	18
cappuccino	Q913972	15	42	30	0	1
cappuccino	Q913972	16	97	77	0	2
cappuccino	Q913972	17	97	73	0	2
cappuccino	Q913972	18	166	138	0	3
cappuccino	Q913972	19	319	277	0	7
TOTALI			**5814**	**4767**	**164**	**135**

Tabella 8. Conteggi degli abbinamenti di tipo 4, in base alla proprietà P611

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-679

Proprietà	Nome	Estratti con query	Abbinamenti ottenuti	% ottenuti
P8290	Archivio Storico Ricordi	2465	568	23,04%
P8982	Pontificio Istituto Archeologia Cristiana	1985	413	20,81%
P6715	Sistema Informativo Unificato per le Soprintendenze Archivistiche	558	56	10,04%
P7203	Dizionario biografico dei Friulani	433	50	11,55%
P9625	Dizionario Biografico dell'Educazione (1800–2000)	178	26	14,61%
P10844	Pontificia facoltà teologica Teresianum	787	157	19,95%
P8985	Dizionario bio-bibliografico dei bibliotecari italiani del XX secolo	88	0	0,00%
P5619	Fondazione Franco Fossati	70	10	14,29%
P5739	Pont. Univ. S. Croce	17345	3012	17,37%
P8034	Biblioteca apostolica vaticana	71609	8760	12,23%
P12458	Parsifal (URBE)	76894	4862	6,32%
P5731	Pontificia Università S. Tommaso	14089	17	0,12%
P8153	Accademia delle Scienze di Torino	759	119	15,68%
P1986	Dizionario Biografico degli Italiani	11871	635	5,35%
TOTALI		**199131**	**18685**	**9,38%**

Tabella 9. Fonti e conteggi degli abbinamenti di tipo 5

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-679

Riferimenti bibliografici

Anderson, Dorothy. 1974. *Universal Bibliographic Control. A long term Policy — A Plan for action*. München: Verlag Dokumentation.

Anderson, Dorothy. 2000. «IFLA's Programme of Universal Bibliographic Control: Origins and Early Years.» *IFLA Journal* 26 (3): 209-14. https://doi.org/10.1177/034003520002600309.

Atturo, Valentina, e Elena Ravelli. 2024. «The evolution of authority work in SBN. From origins to Alphabetica and future prospects.» *JLIS.it* 15 (1): 45–58. https://doi.org/10.36253/jlis.it-572.

Bargioni, Stefano. 2020. «From Authority Enrichment to AuthorityBox: Applying RDA in a Koha environment.» *JLIS.it* 11 (1): 175–89. https://doi.org/10.4403/jlis.it-12595.

Bargioni, Stefano, Elisa Bellistri, Giovanni Bergamin, Carlo Bianchini, Giulio Blasi, Alessandra Boccone, Marilena Daquino, Roberto Delle Donne, Pierluigi Feliciati, Claudio Forziati, Alberto Gambardella, Stefano Gambari, Maurizio Lana, Valeria Lo Castro, Federico Meschini, Alessandra Moi, Rossana Morriello, Stefano Parise, Valdo Pasqui, Tiziana Possemato, Roberto Raieli, Cristina Silvani, Francesca Tomasi, Federico Valacchi, Chiara Veninata, Maurizio Vivarelli. 2022. «Ventisei bibliotecari e professori di biblioteconomia rispondono a dieci domande poste da Mauro Guerrini su temi legati alla metadatazione». In *Metadatazione*, 107–272. Milano: Editrice Bibliografica.

Bargioni, Stefano, Carlo Bianchini, e Camillo Carlo Pellizzari di San Girolamo. «Dati e script dell'articolo "Wikidata e SBN. Un bilancio di due anni di lavoro (2023-2025)"». Zenodo, 7 luglio 2025. https://doi.org/10.5281/zenodo.15829966.

Berners-Lee, Tim. 2006. *Linked Data – Design Issues*. https://www.w3.org/DesignIssues/LinkedData.html.

Bianchini, Carlo. 2017. «Osservazioni sul modello IFLA Library Reference Model.» *JLIS.it* 8 (3): 86–99. https://doi.org/10.4403/jlis.it-12416.

Bianchini, Carlo. 2024. «Nessun catalogo è un isola». In *Parsifal. Un modello di collaborazione bibliotecaria per condividere la conoscenza registrata*, a cura di Silvano Danieli, 57–72. Firenze: Firenze University Press. https://doi.org/10.36253/979-12-215-0356-2.10.

Bianchini, Carlo, Stefano Bargioni, e Camillo Pellizzari di San Girolamo. 2021. «Beyond VIAF. Wikidata as a complementary tool for authority control in libraries.» *Information Technology and Libraries* 40 (2): 1–31. https://doi.org/10.6017/ital.v40i2.12959.

Bianchini, Carlo, Stefano Bargioni, e Camillo Carlo Pellizzari di San Girolamo. 2022. «Le voci di autorità dei nomi di persona in SBN e Alphabetica: problemi e prospettive.» *Bibliothecae.it* 11 (1): 1–67. https://doi.org/10.6092/issn.2283-9364/15078.

Bianchini, Carlo, e Mauro Guerrini. 2014. *Introduzione a RDA. Linee guida per rappresentare e scoprire le risorse*. Milano: Editrice Bibliografica.

Bianchini, Carlo, e Lucia Sardo. 2022. «Wikidata: a new perspective towards universal bibliographic control.» *JLIS.it* 13 (1): 291–311. https://doi.org/10.4403/jlis.it-12725.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-679

Cencetti, Elena, Camillo Carlo Pellizzari di San Girolamo, e Elisabetta Viti. 2025. «Termini, dati e collegamenti: 'conversazioni' tra il Thesaurus del Nuovo soggettario e Wikidata.» *Imagines* 12: 85–94.

De Monaco, Sara, «Il Dizionario degli scrittori italiani contemporanei pseudonimi in Wikidata. Metodologia e risultati» (Tesi di laurea magistrale, Università di Pisa, 2025). https://etd.adm.unipi.it/t/etd-05132025-160428.

De Monaco, Sara, e Camillo Carlo Pellizzari di San Girolamo. 2025. «Il Dizionario degli scrittori italiani contemporanei pseudonimi in Wikidata. Metodologia e risultati (dati)». Zenodo. https://doi.org/10.5281/ZENODO.15474929.

Frattarolo, Renzo. 1975. *Dizionario degli scrittori italiani contemporanei pseudonimi. 1900-1975. Con un repertorio delle bibliografie nazionali di opere anonime e pseudonime*. Ravenna: Longo.

Guerrini, Mauro. 2020. *Dalla catalogazione alla metadatazione: tracce di un percorso*. Roma: AIB.

Guerrini, Mauro. 2022. *Metadatazione: la catalogazione in era digitale*. Milano: Editrice Bibliografica.

Guerrini, Mauro, e Lucia Sardo. 2003. *Authority control*. Roma: Associazione italiana biblioteche.

Guerrini, Mauro. 2018. *IFLA Library Reference Model (LRM). Un modello concettuale per le biblioteche del XXI secolo*. Milano: Editrice Bibliografica.

IFLA (International Federation of Library Associations and Institutions). 2016. *Overview of differences between IFLA LRM and the FRBR-FRAD-FRSAD models*. Draft. https://www.ifla.org/files/assets/cataloguing/frbr-lrm/transitionmapping_overview_20161207.pdf.

IFLA (International Federation of Library Associations and Institutions). 2017. *IFLA Library Reference Model. A Conceptual Model for Bibliographic Information*. A cura di Pat Riva, Patrick Le Boeuf, e Maja Zumer. Den Haag: IFLA. https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla_lrm_2017-03.pdf.

IFLA (International Federation of Library Associations and Institutions). 2024. *UNIMARC Authorities Format Manual*. Online edition. Version: 1.1.0, 2024. Den Haag: IFLA. https://www.ifla.org/unimarc-updates/unimarc-authorities-format-manual-online-ed/.

IFLA (International Federation of Library Associations and Institutions). Study Group on the Functional Requirements for Bibliographic Records. 2009. *Functional Requirements for Bibliographic Records: Final Report*, approved by the Standing Committee of the IFLA Section on cataloguing. September 1997; as amended and corrected through February 2009. München: K.G. Saur. http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf.

IFLA (International Federation of Library Associations and Institutions). Working Group on Functional Requirements and Numbering of Authority Records (FRANAR). 2009. *Functional Requirements for Authority Data: a Conceptual Model. Final report*. München: K.G. Saur.

IFLA (International Federation of Library Associations and Institutions). Working Group on Functional Requirements for Subject Authority Records (FRSAR). 2010. *Functional Requirements for Subject Authority Data (FRSAD). A conceptual model*. München: K. G. Saur. https://www.ifla.org/files/assets/classification-and-indexing/functional-requirements-for-subject-authority-data/frsad-final-report.pdf.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-679

Jansová, Linda, Lenka Maixnerová, e Petra Šťastná. 2024. «Library Data in Wikimedia Projects: Case Study from the Czech Republic.» *O-Bib. Das Offene Bibliotheksjournal* 11 (4): 1–19. https://doi.org/10.5282/O-BIB/6081.

JSC (Joint Steering Committee for Development of RDA), a c. di. 2010. *Resource Description & Access. RDA*. American library association; Canadian library association; CILIP.

Linked Data for Production, *Wikidata as a hub for identifiers*. 11 giugno 2020, https://t.ly/677QP.

Manzoni, Laura. 2022. *Identificatori*. Roma: Associazione italiana biblioteche.

Pellizzari di San Girolamo, Camillo Carlo. 2024a. «Storia della collaborazione tra Wikidata e le biblioteche della Rete URBE nel controllo di autorità». In *Parsifal. Un modello di collaborazione bibliotecaria per condividere la conoscenza registrata*, a cura di Silvano Danieli, 147–62. Firenze: Firenze University Press. https://doi.org/10.36253/979-12-215-0356-2.17.

Pellizzari di San Girolamo, Camillo Carlo. 2024b. «The reconciliation of SBN authority records with Wikidata. Progresses and perspectives after a decade of work (2013-2023).» *JLIS.it* 15 (1): 33–44. https://doi.org/10.36253/jlis.it-573.

Possemato, Tiziana. 2024. *Entity modeling: la terza generazione della catalogazione*. Firenze: Firenze University Press. https://doi.org/10.36253/979-12-215-0393-7.

Ravelli, Elena. 2024. «Authority file di SBN e Wikidata. Un progetto per l'arricchimento dei dati del Catalogo nazionale.» (Intervento presentato ai Wikidata Days, Bologna, novembre 8, 2024). https://commons.wikimedia.org/wiki/File:Authority_file_di_SBN_e_Wikidata_-_Wikidata_Days_Bologna_2024.pdf.

Solimine, Giovanni. 1995. *Controllo bibliografico universale*. Roma: Associazione italiana biblioteche.

Tillett, Barbara B., Khaled Mohamed Reyad, e Ana Lupe Cristán, a c. di. 2006. «A Virtual International Authority File (VIAF)». In *IFLA Cataloguing Principles: Steps towards an International Cataloguing Code, 3*, 90–116. Berlin, Walter de Gruyter – K. G. Saur. https://doi.org/10.1515/9783598440335.1.90.

Tillman, Ruth Kitchin, e Gala Campos Oaxaca. 2025. «Crossing Silos: Assessing the Utility of Identity Attributes in Name Reconciliation.» *portal: Libraries and the Academy* 25 (2): 299–319. https://dx.doi.org/10.1353/pla.2025.a955947.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-677

JLIS.it

Wikidata-Enhanced Authority Records: A Project for Personal Names in SBN

Elena Ravelli[(a)]

a) Central Institute for the Union Catalog of Italian Libraries and for Bibliographic Information (ICCU), https://orcid.org/0000-0001-6402-2039

Contact: Elena Ravelli, elena.ravelli@cultura.gov.it
Received: 16 July 2025; **Accepted:** 18 October 2025; **First Published:** 15 January 2026

ABSTRACT
The catalogue of the National Library Service (SBN), developed and expanded through cooperative cataloguing, has experienced exponential growth over time. Currently, the titles archive contains over 21 million records, while the authors archive includes around 4,200,000 entries. However, not all authority records in SBN meet the level of completeness and accuracy required of a national catalogue. As part of the collaboration between Wikimedia Italy and ICCU, a project was launched to enrich a portion of the authority records related to personal names in SBN that lacked information. This enrichment was made possible thanks to data available on Wikidata, where many records include the SBN identifier. The number of entries in Wikidata associated with SBN identifiers is expected to grow further, also thanks to ICCU's recent decision to make all records related to personal and corporate names in the Index visible in the SBN OPAC, regardless of their authority level.

KEYWORDS
Wikidata; Opac SBN; Authority control; Interoperability.

Il progetto di arricchimento dei nomi di persona di SBN con i dati di Wikidata

ABSTRACT
Il catalogo del Servizio bibliotecario nazionale (SBN), sviluppato e ampliato attraverso la catalogazione partecipata, ha registrato nel tempo una crescita esponenziale. Attualmente, l'archivio titoli conta oltre 21 milioni di record, mentre l'archivio autori raccoglie circa 4.200.000 voci. Tuttavia, non tutte le registrazioni di autorità presenti in SBN possiedono il livello di completezza e accuratezza richiesto da un catalogo nazionale. Nell'ambito della collaborazione tra Wikimedia Italia e l'ICCU è stato avviato un progetto volto ad arricchire una parte delle voci di autorità relative ai nomi di persona che in SBN risultavano prive di informazioni. L'arricchimento è stato possibile grazie ai dati disponibili su Wikidata, dove molti record contengono l'identificativo SBN. Il numero di voci in Wikidata associate a identificativi SBN è destinato a crescere, anche grazie alla recente decisione dell'ICCU di rendere visibili nell'OPAC SBN tutte le registrazioni relative ai nomi di persona e di ente presenti nell'Indice, a prescindere dal loro livello di autorità.

PAROLE CHIAVE
Wikidata; Opac SBN; Authority control; Interoperabilità.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-677

Introduzione

Il progetto di arricchimento dei record relativi ai nomi di persona presenti nel Catalogo del Servizio Bibliotecario Nazionale (SBN) con dati di Wikidata si inserisce nel quadro della convenzione fra l'Istituto centrale per il catalogo unico delle biblioteche italiane (ICCU) e Wikimedia Italia (WMI).[1] L'accordo, oltre a sancire un impegno condiviso nella diffusione dei principi dell'open access e dell'open culture, promuove l'integrazione dei dati e dei materiali dei progetti dell'ICCU con quelli di WMI. Tra le prime azioni realizzate, vi è stata l'integrazione dei dati dell'Anagrafe delle Biblioteche Italiane (ABI)[2] con Wikidata e OpenStreetMap.[3] Un'altra significativa attività, sviluppata tra il 2015 e il 2019, durante la collaborazione di un wikipediano in residenza presso l'ICCU, ha riguardato il collegamento tra le entità di Wikidata e i record relativi ai nomi di persona presenti nell'OPAC di SBN, relazione che è basilare per il progetto qui descritto. A quegli anni risale anche la decisione dell'ICCU di utilizzare una piattaforma Mediawiki per la pubblicazione delle normative per la catalogazione in SBN,[4] uno strumento che consente una più agevole modalità di aggiornamento dei contenuti e, aspetto più importante, l'accesso libero e gratuito alle normative da parte di tutti i catalogatori e bibliotecari italiani. Uno degli impieghi più recenti dei dati di Wikidata riguarda Alphabetica,[5] l'ultimo portale sviluppato dall'ICCU. Al suo interno, il percorso "Protagonisti" offre brevi profili degli autori, tratti dalla Nota informativa del record SBN. Quando questa nota non è presente, le informazioni vengono recuperate da Wikidata. Da Wikidata provengono sempre – tramite query SPARQL sull'endpoint LOD e senza alcun sistema di caching – anche il luogo e le date di nascita e di morte, oltre all'immagine del protagonista.[6]
Negli ultimi anni si è stabilito un dialogo attivo e proficuo tra l'ICCU e alcuni membri della comunità Wiki, con un focus particolare sull'Authority file di SBN. Da questa collaborazione sono nate segnalazioni sistematiche da parte dei collaboratori di Wikidata[7] all'Ufficio Authority di SBN in merito a errori e duplicazioni presenti nel Catalogo nazionale. Questo confronto si è concretizzato in un percorso di formazione da parte dell'ICCU sulle principali funzionalità dell'applicativo di Indice, in modo da consentire a un membro del gruppo Wikidata di intervenire direttamente sulle registrazioni, correggendo le voci di autorità presenti nel Catalogo.

---

[1] La Convenzione ICCU Wikimedia Italia è stata sottoscritta per la prima volta nel 2015 (ICCU e Wikimedia Italia 2015) poi è stata rinnovata nel 2018 (ICCU, e Wikimedia Italia 2018). Attualmente è in vigore la Convenzione del 2022 (ICCU e Wikimedia Italia 2015).

[2] ABI: https://anagrafe.iccu.sbn.it/it/

[3] Sull'evoluzione del progetto cfr. (Rolleri 2024).

[4] Norme catalografiche per SBN: https://norme.iccu.sbn.it/index.php?title=Normative_catalografiche. La piattaforma, utilizzata per la prima volta nel 2016, ospita anche il Codice nazionale REICAT, e dal dicembre 2024 la guida a Manus Online.

[5] Alphabetica: https://alphabetica.it/web/alphabetica/.

[6] L'algoritmo per l'acquisizione delle informazioni si attiva quando l'utente avvia una ricerca sul Protagonista e procede seguendo una gerarchia di controlli. Per prima cosa verifica la presenza della proprietà P396 (identificativo SBN): se disponibile, la ricerca si conclude qui. Se nel record di authority SBN è presente anche il codice ISNI, viene effettuata una query sulla proprietà P213 (ISNI) in Wikidata e, se trovata, il processo termina. In assenza sia del VID sia dell'ISNI in Wikidata, l'algoritmo effettua la ricerca basandosi sulla somiglianza della stringa del Nome, applicando un eventuale filtro per data quando disponibile. In quest'ultimo caso, in presenza di omonimie, può presentarsi la possibilità di errori di associazione in presenza di omonimie.

[7] Si ringraziano nello specifico Stefano Bargioni e Camillo Carlo Pellizzari di San Girolamo per le liste di segnalazioni inviate nel corso degli ultimi due anni all'Ufficio Authority dell'ICCU.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-677

Oggetto del presente intervento è la descrizione del progetto che consiste nell'utilizzo dei dati di Wikidata per arricchire i nomi di persona che nel Catalogo di SBN risultino privi di informazioni identificative fondamentali, come la Nota informativa, le Datazioni e i codici Lingua e Paese.

Authority File di SBN

Fra le attività qualificanti del Catalogo di SBN, una delle più rilevanti riguarda la gestione e il monitoraggio dell'Authority file. Attualmente, in SBN vengono gestite voci d'autorità relative a nomi di persona, nomi di ente, titoli dell'opera, titoli dell'opera musicale, soggetti e classi. Inoltre, per le registrazioni bibliografiche anteriori al 1831, sono previsti ulteriori punti di accesso controllati: luogo di pubblicazione, editori e tipografi, e marche tipografiche. Questa attività è coordinata dall'Ufficio Authority dell'ICCU, il cui obiettivo principale è quello di garantire la qualità e il controllo dei punti di accesso ai record bibliografici presenti nel Catalogo nazionale. A tal fine, l'ICCU ha definito specifiche normative per la registrazione delle voci di autorità,[8] in accordo con gli standard internazionali, e ha promosso percorsi formativi per i catalogatori della rete nazionale,[9] anche in collaborazione con la Fondazione Scuola nazionale del patrimonio e delle attività culturali, nell'ambito del progetto Dicolab.[10]

A supporto del lavoro di authority, è stato costituito un *Gruppo di lavoro per la gestione e la manutenzione dell'Authority file di SBN. Nomi*[11] composto da esperti bibliotecari dei Poli di SBN che, per competenza, dislocazione territoriale, specificità di funzioni ed utenza, assicurassero al contesto cooperativo la professionalità e l'impegno indispensabili al compito delicato ed oneroso loro assegnato. Più recentemente è stato costituito un *Comitato di coordinamento per la gestione e la manutenzione dell'Authority file di SBN*[12] con il compito di fornire indirizzo e raccordo ai diversi Gruppi di lavoro operanti in attività di authority in SBN.

Lo sforzo impiegato dall'ICCU e dall'intera comunità di SBN per garantire la correttezza e l'uniformità formale dei punti di accesso ha dovuto misurarsi, nel tempo, con la crescita esponenziale dei dati del catalogo: i soli nomi di persona superano ormai i 3,7 milioni di record. Parallelamente, la progressiva riduzione delle risorse umane ed economiche destinate alla catalogazione, rende sempre più difficile gestire l'Authority file di SBN affidandosi esclusivamente al lavoro umano. Per questo motivo, nei prossimi anni, sarà necessario potenziare gli strumenti di monitoraggio del

[8] Per le normative per la registrazione di authority in SBN cfr. https://norme.iccu.sbn.it/index.php?title=Norme_comuni/Authority_file. Per i soggetti e le classi, cfr. https://norme.iccu.sbn.it/index.php?title=Norme_comuni/Linee_guida_sull%27indicizzazione.

[9] L'ultimo percorso formativo realizzato dall'ICCU si è svolto nel maggio 2025. Per il programma cfr. https://www.iccu.sbn.it/it/eventi-novita/novita/Corsi-di-formazione-e-aggiornamento-professionale-2025/.

[10] Il progetto Dicolab si colloca all'interno del Piano Nazionale per la Digitalizzazione del patrimonio culturale e ha come obiettivo quello di aggiornare e migliorare le competenze digitali dei professionisti della cultura. «Catalogare in SBN» è uno dei percorsi realizzati in collaborazione con ICCU, cfr. Catalogare in SBN: un nuovo percorso Dicolab dedicato ai processi di catalogazione digitale - Dicolab.

[11] Il gruppo di lavoro è stato più volte aggiornate nel corso degli anni, l'ultimo aggiornamento risale al 4 dicembre 2024. https://www.iccu.sbn.it/it/attivita-servizi/gruppi-di-lavoro-e-commissioni/gruppo-di-lavoro-per-la-gestione-e-la-manutenzione-dellauthority-file-di-sbn-/index.html.

[12] https://www.iccu.sbn.it/it/attivita-servizi/gruppi-di-lavoro-e-commissioni/comitato-di-coordinamento-per-la-gestione-e-manutenzione-dellauthority-file-di-sbn/.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-677

Catalogo e realizzare interventi correttivi massivi e centralizzati. In questa direzione, sarà fondamentale anche l'adozione di procedure automatiche per la validazione e l'arricchimento dei dati, basate sul confronto con altre banche dati, sia nazionali che internazionali. L'obiettivo è rafforzare l'interoperabilità del Catalogo nazionale, promuovendo lo scambio e il riutilizzo dei dati, così da migliorare la qualità dell'Authority file, sia in termini di completezza che di affidabilità dei record. Del resto, l'idea stessa di Controllo Bibliografico Universale si basa sulla possibilità di condividere i dati bibliografici e le voci di autorità prodotti dalle agenzie bibliografiche nazionali (Bianchini, Bargioni, e Pellizzari di San Girolamo 2022, 248).[13] In quest'ottica, tra gli sviluppi previsti con il progetto Indice 3 (Atturo e Ravelli 2024, 56), si prevede un uso più esteso del campo 017 di UNIMARC, utile per collegare i record SBN con identificatori esterni, diversi da ISNI, come ad esempio VIAF e Wikidata. Questi collegamenti potranno contribuire in modo significativo al controllo e al miglioramento della qualità dei dati di authority presenti nel nostro Catalogo.

La prima fase del progetto

L'arricchimento massivo dell'Authority file di SBN si basa sul riutilizzo dei dati di un progetto, come quello di Wikidata, basato sull'archiviazione di una grande quantità di dati strutturati. Alla base di questo progetto c'è un'attività, ormai consolidata nel tempo che consiste nel collegamento delle entità di Wikidata con le voci dell'Authority file presenti nell'OPAC di SBN, mediante la registrazione nella scheda Wiki della proprietà P396, ovvero dell'identificativo del corrispettivo record di SBN.[14] A tale proposito, va ricordato che, fino a giugno del 2023, nell'OPAC di SBN venivano pubblicate solo le voci con un livello di autorità compreso tra 90 e 97, cioè meno del 10% del totale delle voci registrate nell'Indice SBN.[15] Nel giugno del 2023, l'ICCU decideva di includere anche le voci con livello di autorità 51 e 71, che fino a quel momento non erano state esportate nell'OPAC, nella consapevolezza che in quella porzione di record fossero concentrate le maggiori criticità dell'Authority file di SBN: voci duplicate, voci non controllate e prive di forme di rinvio, voci per le quali spesso non è stata verificata la coerenza dei titoli collegati. Inoltre, la maggior parte dei record con livello 51 e 71 contiene informazioni identificative scarse o addirittura assenti. Tuttavia, in un'ottica di maggiore trasparenza e apertura dei dati delle proprie piattaforme, l'ICCU ha scelto di rendere pubblici anche questi dati, che costituiscono la parte numericamente più consistente. Di conseguenza, nell'OPAC di SBN sono ora presenti tutti i record relativi a nomi di persona e di ente, con esclusione di quelli con livello di autorità 05.[16] L'auspicio è che questa maggiore apertura dei dati favorisca una migliore interoperabilità con altre piattaforme, anche non necessariamente bibliografiche, che gestiscono le medesime entità.

[13] Sulla condivisione e il riuso dei dati bibliografici cfr. anche (Guerrini 2020, 38-42); sull' importanza della combinazione di processi manuali e automatici di validazione e arricchimento dei dati, cfr. anche (Casalini 2022).

[14] Per una storia della proprietà 396 di Wikidata, cfr. (Pellizzari di San Girolamo 2024).

[15] Si trattava delle stesse voci che l'ICCU invia anche a VIAF. A partire dal 2009 l'ICCU ha inviato annualmente al VIAF le voci di autorità con un alto grado di completezza ed esaustività, corrispondenti al livello 97 (Authority) e 95 (Super). Dal 2023, nell'ottica di ampliare la visibilità dei record di SBN, ha iniziato a inviare anche le voci con livello 90 (Massimo).

[16] "Il livello 05 è predisposto per le retroconversioni da cataloghi o da repertori, i dati previsti sono quelli del trattamento del nome a livello minimo (51), se presenti e desumibili dalla scheda o dal repertorio di riferimento, in particolare i dati da inserire, se possibile, sono quelli relativi ai codici di qualificazione e i dati necessari alla disambiguazione del nome", cfr. (Atturo e Ravelli 2024, 48-9).

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-677

Il progetto di arricchimento ha avuto avvio nel maggio 2023, quando gli identificativi di SBN collegati a entità di Wikidata erano 103.148. Da questo elenco sono stati selezionati solamente i record relativi ai nomi di persona[17] che, nell'Indice SBN, risultavano privi di Nota informativa, Datazioni, codice Lingua e codice Paese. In questa prima fase l'attenzione è stata rivolta solo ai record che non contenevano alcuna informazione, se non i campi ISNI o Fonte. I record con queste caratteristiche erano 17.145. Per queste voci sono state selezionate alcune proprietà di Wikidata destinate a completare, nel record di Indice, i campi Datazioni, Nota informativa e, quando assente, il campo ISNI. In particolare, le proprietà selezionate sono state le seguenti: P19 (Luogo di nascita), P20 (Luogo di Morte), P569 (Data di nascita), P570 (Data di morte), P106 (Occupazione), P611 (Ordine religioso), P123 (ISNI). Un requisito fondamentale è stato che ciascuna proprietà fosse supportata da almeno una fonte di riferimento.
Con l'esperienza e la consapevolezza che le operazioni massive sulle basi dati comportano necessariamente un margine di errore, è stata condotta una verifica estremamente approfondita sulle proprietà estratte, al fine di individuare tutte le possibili tipologie di errore e valutarne l'impatto sul Catalogo. La situazione più critica si è verificata in caso di errata associazione fra entità, errore riscontrato in un numero residuale di casi. Più frequente è stata invece l'associazione in Wikidata con identificativi che in SBN potevano fare riferimento a entità duplicate. Tuttavia, anche questo errore non era tale da inficiare il buon risultato del progetto. In ogni caso, tutti gli errori rilevati durante il test sono stati immediatamente corretti: in particolare, si è intervenuti sulle entità duplicate in SBN, attività che ha consentito di estendere le correzioni anche a Wikidata. Inoltre, sono stati necessari alcuni interventi per adattare le informazioni provenienti da Wikidata ai formalismi richiesti per la registrazione nei campi dei record di SBN.
Le principali problematiche che hanno richiesto un intervento di ICCU sono illustrate nei punti seguenti.

- **Luogo di nascita e il luogo di morte**
  Poiché i luoghi di nascita e morte registrati in Wikidata possono indicare siti specifici, come un ospedale o un castello, e non necessariamente un'unità amministrativa, è stata necessaria una revisione degli elenchi dei luoghi. Questo ha permesso di inserire nella Nota informativa di SBN l'entità amministrativa corrispondente alla nascita e/o alla morte, generalmente il comune o, in mancanza di questo, la regione o il Paese.

- **Occupazione**
  Poiché in Wikidata possono essere registrate più occupazioni, non tutte sempre accompagnate da una fonte, capita talvolta che l'unica occupazione referenziata sia quella meno qualificante o comunque insufficiente a identificare con precisione la persona. Ad esempio, molti nomi risultavano qualificati solo con l'occupazione "politico", un'indicazione poco significativa per persone che hanno ricoperto ruoli pubblici di rilievo, come presidente del consiglio, presidente della repubblica, senatore, e simili. Tale informazione è registrata nella proprietà P39 (Carica ricoperta), che in una prima fase non era stata considerata.

[17] Sono stati quindi esclusi i nomi di ente, compresi quelli di tipografi, editori, librai, identificati tramite la forma inversa del nome (cognome, nome), poiché in SBN il record relativo all'autore è distinto da quello riferito al tipografo, anche quando si tratta della stessa entità.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-677

- **Ordine religioso**
  Analogamente al caso dell'occupazione, anche l'indicazione dell'ordine religioso risultava spesso insufficiente a identificare correttamente una persona che, nel corso della sua vita, aveva ricoperto incarichi di rilievo, come quello cardinalizio o papale. Tali informazioni sono registrate nella proprietà P39 (Carica ricoperta). Inoltre, nel corso di questa prima fase di verifica, è emerso che per i religiosi è particolarmente importante estrarre anche la proprietà P411 (Stato di canonizzazione), inizialmente non presa in considerazione.

- **Data**
  In Wikidata, le date sono espresse secondo lo standard ISO 8601 che definisce i formati per la rappresentazione di date e orari. Oltre alla stringa che indica anno-mese-giorno, viene associato un valore che ne specifica il grado di precisione: ad esempio il valore 11 indica che sono noti giorno, mese e anno; il valore 10 indica che sono noti solo il mese e l'anno; il valore 9 che è noto solo l'anno; il valore 8 che è noto il decennio e così via.[18] Un ulteriore indicatore segnala se la data è incerta ("ca."= circa), o solo ipotizzata ("?"). L'informazione relativa alla data andava riportata sia nel campo Datazioni, sia nel campo Nota informativa di SBN: nel campo Datazioni la data è stata trascritta utilizzando solamente caratteri numerici, secondo la normativa, recentemente aggiornata, per la registrazione delle Datazioni nei record di authority di SBN.[19] Nella Nota informativa, invece, i dati sono stati rielaborati e trascritti in forma discorsiva, adottando locuzioni diverse a seconda delle differenti modalità in cui la data si poteva presentare, ad esempio: "nato il 10 novembre 1824, nato nell'ottobre del 1994, vissuto nel XVI secolo", etc.

- **ISNI**
  A eccezione dei casi di associazione errata di codici ISNI in Wikidata, difficili da individuare, la principale criticità riscontrata riguardava la presenza di entità duplicate in ISNI. Tuttavia, grazie al controllo presente in Interfaccia Diretta, che impedisce l'inserimento di uno stesso ISNI in più record, durante la fase di importazione degli ISNI da Wikidata sono stati rilevati diversi casi di duplicazione, che sono stati tempestivamente corretti.

Al termine della prima fase di verifica delle informazioni estratte da Wikidata, è emerso che, oltre alla Carica ricoperta e allo Stato di beatificazione, utili a completare le informazioni sulle entità presenti nel Catalogo, erano importanti anche altre tre proprietà inizialmente non prese in considerazione: P21 (Sesso o genere), P27 (Paese di cittadinanza), P6886 (Lingua di scrittura).

[18] Sulla modalità di registrazione delle date in Wikidata vedi https://www.wikidata.org/wiki/Help:Dates/it.

[19] https://norme.iccu.sbn.it/index.php?title=Norme_comuni/Authority_file/Nomi_di_persona/Registrazione_di_authority/Datazioni. La normativa per la registrazione delle datazioni delle voci di autority in SBN prevede l'utilizzo di soli caratteri numerici nella forma: anno.mese.giorno, quando tutti e tre conosciuti, altrimenti solo l'anno o il decennio o il secolo, secondo diverse casistiche. Una modalità che va incontro a un progressivo adeguamento agli standard internazionali, cfr. (Atturo e Ravelli 2024, 51-2).

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-677

- **Sesso o genere**
  Pur non essendo previsto un codice di genere in SBN, questa informazione risulta fondamentale per poter declinare correttamente al maschile o al femminile le occupazioni e le espressioni "nato/nata" e "morto/morta" all'interno della Nota informativa.

- **Paese di cittadinanza**
  Poiché il record SBN consente l'associazione di un solo codice Paese, si è deciso di utilizzare la proprietà di Wikidata solo nei casi in cui fosse presente una sola occorrenza, in assenza di un principio di gerarchia che definisse un paese principale di cittadinanza.

- **Lingua di scrittura**
  Secondo le norme SBN per i Nomi di persona, la lingua da indicare è quella utilizzata dall'autore nelle sue pubblicazioni. Si tratta tuttavia di una proprietà non molto rappresentata in Wikidata. Come per il codice Paese, anche questa informazione è stata acquisita solo quando risultava una singola occorrenza.

Una volta ricomposte in un unico file tutte le proprietà estratte separatamente da Wikidata, è stato necessario concatenare le informazioni per elaborare le stringhe di carattere discorsivo da inserire nella Nota informativa del record di SBN. Le espressioni sono state adattate e declinate in modo diverso a seconda delle tipologie di dati da ricomporre.

Di seguito alcuni esempi di formulazioni da inserire nel campo Nota informativa:
- Agronomo, giornalista. Nato a Cerro Maggiore, 31.12.1878, morto a Roma, 18.04.1960,
- Omeopata. Nato nel 1840, morto nel 1901,
- Archivista, filosofo, scrittore. Nato in Cina, presumibilmente intorno al 571 a.C., morto presumibilmente intorno al 490 a.C.

Al termine del lavoro, è emerso che per circa 2.000 autori mancavano informazioni relative all'attività svolta, alla carica ricoperta o all'eventuale stato di beatificazione. Nella maggior parte dei casi, tali dati potevano essere ricavati dalla descrizione presente nella voce di Wikidata, di seguito all'etichetta con il nome. Questa descrizione viene generata automaticamente sulla base di informazioni non sempre contenute nei campi formalizzati o, se presenti, non sempre accompagnate da fonti. Un'indagine a campione ha confermato che queste descrizioni risultavano, nel complesso, attendibili. Si è quindi proceduto a estrarle ed eventualmente a tradurle in italiano nei casi in cui comparissero in lingua inglese, tedesca o francese.
Nel luglio 2024, completati tutti i test e apportate le necessarie correzioni e integrazioni, sono stati inseriti nell'Indice SBN 16.102 record arricchiti con nuove informazioni identificanti. Tali record sono riconoscibili dalla Nota del catalogatore, che riporta un'espressione analoga alla seguente: "Informazioni da Wikidata 01.07.2023–31.03.2024 wikidata.org/wiki/Q94409375".

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-677

La prosecuzione del progetto

Nell'autunno del 2024 è stata effettuata una nuova estrazione di dati da Wikidata. Rispetto alla precedente selezione, le entità con proprietà P396 (Identificativo SBN dell'autore) risultavano pari a 144.347,[20] inclusi i 16.102 che erano stati già arricchiti con informazioni provenienti da Wikidata. Anche in questa fase, l'obiettivo era individuare i record di SBN privi di informazioni identificanti. L'attenzione si è concentrata non solo sui record con Nota informativa vuota, ma anche su quelli in cui la Nota risultava compilata con contenuti non pertinenti, quali la data, le fonti, o altre informazioni di servizio.[21] Per questi casi si è proceduto a una bonifica dei record, riportando le informazioni nei campi corretti oppure eliminando le informazioni non appropriate. Di conseguenza, il numero di record presenti in Indice privi di Nota informativa è cresciuto in modo significativo, raggiungendo circa 36.000 unità.

A differenza della prima fase, si è ritenuto opportuno intervenire anche sui record di SBN che presentavano alcuni campi già compilati, come Datazione, Lingua o Paese. Si è infatti ritenuto che l'aggiunta di elementi come il luogo di nascita o di morte, l'occupazione, le cariche ricoperte o l'eventuale stato di canonizzazione, potesse contribuire a migliorare sensibilmente la qualità complessiva del record.

In questa seconda fase, è stata sviluppata una procedura specifica per l'estrazione dei dati di Wikidata, progettata per recepire tutte le esigenze emerse durante la fase iniziale del progetto. Le specifiche della procedura sono illustrate nelle 'Tabelle per estrazione dati per ICCU' in Wikidata.[22]

Il progetto è tuttora in corso e procede attraverso successivi affinamenti ma ha già superato la fase sperimentale iniziale, restituendo risultati decisamente incoraggianti. La procedura messa a punto ha permesso non solo di arricchire oltre 51.000 record di SBN precedentemente privi di informazioni identificanti, ma anche di individuare e correggere molteplici criticità: voci duplicate, altre che necessitavano di essere disambiguate, note informative improprie, datazioni non aggiornate e altre incongruenze presenti nell'Indice di SBN. Il miglioramento della qualità dei record non ha solo un impatto sull'affidabilità complessiva del catalogo, ma apporta benefici su più livelli. Per gli utenti, significa disporre di informazioni più chiare, dettagliate e coerenti; per i catalogatori, rappresenta uno strumento prezioso per facilitare le attività di *authority control*, agevolando l'univocità delle intestazioni e la coerenza nell'organizzazione dei dati bibliografici.

Guardando al futuro, l'auspicio è quello di ampliare ulteriormente il dialogo e le possibilità di interoperabilità con altre banche dati, sia nazionali che internazionali. Questo approccio mira a rafforzare l'integrazione tra piattaforme di dati diverse e a promuovere un ecosistema informativo più aperto, dinamico e collaborativo, in grado di valorizzare al meglio l'Authority file di SBN e, più in generale, il patrimonio documentario e culturale a cui esso dà accesso.

[20] Al momento dell'elaborazione del presente contributo (luglio 2025), gli elementi in Wikidata con proprietà P396 hanno superato le 190.000 unità.

[21] In molti casi queste informazioni derivano da lavorazioni o importazioni massive precedenti a Indice 2 quando non esisteva il campo *Nota informativa* ma solo una generica *Nota* che poteva essere compilato con contenuti vari.

[22] https://www.wikidata.org/w/index.php?title=Property_talk:P396&oldid=2341991629#Tabelle_per_estrazione_dati_per_ICCU.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-677

Riferimenti bibliografici*

Atturo, Valentina, e Elena Ravelli. 2024. «The evolution of authority work in SBN. From origins to Alphabetica and future prospects.» *JLIS.it* 15 (1): 45-58. https://doi.org/10.36253/jlis.it-572.

Bianchini, Carlo, Stefano Bargioni, Camillo Carlo Pellizzari di San Girolamo. 2022. «Le voci di autorità dei nomi di persona in SBN e Alphabetica: problemi e prospettive.» *Bibliothecae.it* 11 (1): 247-314. https://doi.org/10.6092/ISSN.2283-9364/15078.

Casalini, Michele. 2022 «The future of bibliographic services in light of new concepts of authority control.» *JLIS.it* 13 (1):107-15. https://doi.org/10.4403/jlis.it-12766.

Guerrini, Mauro. 2020 «Dalla catalogazione alla metadatazione. Tracce di un percorso.» Roma: Associazione italiana biblioteche.

ICCU (Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche), e Wikimedia Italia. 2015 *Convenzione tra ICCU e Associazione Wikimedia Italia - As-sociazione per la diffusione della conoscenza libera*. https://www.iccu.sbn.it/export/sites/iccu/documenti/2016/convenzione_Wikimedia.pdf.

ICCU (Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche), e Wikimedia Italia. 2018 *Convenzione tra ICCU e Associazione Wikimedia Italia - As-sociazione per la diffusione della conoscenza libera*. https://www.iccu.sbn.it/export/sites/iccu/documenti/2020/ICCU-Wikimedia.pdf.

ICCU (Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche), e Wikimedia Italia. 2022 *Convenzione tra ICCU e Associazione Wikimedia Italia - As-sociazione per la diffusione della conoscenza libera*. https://www.iccu.sbn.it/export/sites/iccu/documenti/2022/ICCU_Wikimedia_2022.pdf.

Pellizzari di San Girolamo, Camillo Carlo. 2024. «The riconciliation of SBN Authority records with Wikidata.Progresses and perspectives after a decade of work (2013-2023).» *JLIS.it* 15 (1):v33-44. https://doi.org/10.36253/jlis.it-573.

Rolleri, Davide. 2024. «Importazione dell'Anagrafe delle Biblioteche Italiane in Wikidata.» (slide e testo dell'intervento ai Wikidata Days, Bologna, novembre 8-9, 2024).

* Ultima data di consultazione dei siti web: 11 novembre 2025.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-693

JLIS.it

Cataloguing, Metadata, and Generative AI. Early Experiences and Future Perspectives

Gino Roncaglia[(a)]

a) Roma Tre University, https://orcid.org/0000-0003-4632-3342

Contact: Gino Roncaglia, gino.roncaglia@uniroma3.it
Received: 30 July 2025; **Accepted:** 05 November 2025; **First Published:** 15 January 2026

ABSTRACT
The article deals with the intersection of generative artificial intelligence (AI) and bibliographic/metadata practices, assessing how large language models (LLMs) can support cataloguing and metadata creation while navigating the constraints of formal knowledge architectures. In the first section, the article discusses the evolution of cataloguing paradigms from MARC to Linked Open Data (LOD), emphasizing the shift from rigid records to semantic, entity-based models like FRBR, RDA, and BIBFRAME. The second section deals with the epistemological clash between deterministic, rule-based metadata standards (the "architect") and probabilistic, generative AI systems (the "oracle").
Three strategies are discussed for integrating AI into bibliographic workflows:
1) Specialized AI systems trained exclusively on controlled, high-quality datasets.
2) Retrieval-Augmented Generation (RAG), blending LLMs with authoritative knowledge bases.
3) Next-generation LLMs enhanced via reasoning models, multimodal inputs, expanded context windows, and small/medium-scale local models to align generative outputs with metadata standards.
Key challenges include hallucinations, data sparsity in bibliographic corpora, and the obsolescence of MARC-centric experiments. The article argues for caution against retrofitting AI onto outdated data models, urging alignment with LOD and IFLA's Library Reference Model (LRM). Ethical considerations (bias, transparency, AI literacy) and the potential of local SLMs/MSLMs for privacy-sensitive applications are highlighted.

KEYWORDS
Artificial intelligence; Metadata creation/management; LLM; Bibliographic description; Cataloguing standards.

Catalogazione, metadati e IA generativa. Prime esperienze e prospettive future

ABSTRACT
L'articolo affronta l'intersezione tra intelligenza artificiale generativa (AI) e l'ambito catalografico e di gestione dei metadati, valutando in che modo i modelli linguistici di grandi dimensioni (LLM) possano supportare la catalogazione e la creazione di metadati, rispettando i vincoli posti dalle architetture formali della conoscenza.
Nella prima sezione, l'articolo analizza l'evoluzione dei paradigmi di catalogazione, dal formato MARC ai Linked Open Data (LOD), sottolineando il passaggio da record rigidi a modelli semantici basati su entità, come FRBR, RDA e BIBFRAME. La seconda sezione tratta del conflitto epistemologico tra gli standard di metadatazione formalizzati e basati su regole (l'"architetto") e i sistemi di intelligenza artificiale generativa, per loro natura prevalentemente non deterministici e probabilistici (l'"oracolo"). Per risolvere tale conflitto, vengono prese in esame tre strategie per integrare l'AI nei flussi di lavoro bibliografici:
1) sistemi di AI specializzati, addestrati esclusivamente su dataset controllati e di alta qualità;
2) modelli di tipo Retrieval-Augmented Generation (RAG), che combinano LLM e basi di conoscenza autorevoli;
3) LLM di nuova generazione potenziati da modelli di ragionamento, input multimodali, finestre contestuali estese e modelli locali di piccola/media scala con fine tuning, per allineare gli output generativi agli standard di metadatazione.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-693

Tra le principali criticità vengono evidenziati fenomeni di allucinazione, scarsità di dati nei corpora bibliografici e obsolescenza degli esperimenti basati su MARC. L'articolo invita alla cautela rispetto ai tentativi di adattare l'AI a modelli di dati superati, promuovendo invece l'allineamento con i LOD e con il Library Reference Model (LRM) dell'IFLA. Vengono infine discusse le implicazioni etiche (bias, trasparenza, alfabetizzazione all'AI) e il potenziale dei modelli SLM/MSLM locali per applicazioni che richiedono maggiore tutela della privacy.

PAROLE CHIAVE
Intelligenza artificiale; Metadatazione; LLM; Descrizioni bibliografiche; Standard di catalogazione.

0. Premessa

Questo articolo ha l'obiettivo di proporre alcune considerazioni sul ruolo che può avere l'intelligenza artificiale generativa in campo bibliotecario e documentale, in particolare per quanto riguarda le attività di catalogazione e metadatazione. Si tratterà di considerazioni non solo sintetiche ma anche inevitabilmente provvisorie, considerata la velocità di evoluzione del settore e in particolare dei grandi modelli linguistici, che rappresentano – almeno nel momento in cui scrivo – lo strumento più avanzato utilizzabile in questo contesto.

Prima di illustrare la struttura del testo, una premessa indispensabile riguarda l'uso dell'IA generativa nel lavoro di scrittura dell'articolo. Considerato il tema trattato, infatti, mi è sembrato non solo utile ma anche opportuno interagire in maniera abbastanza stretta con alcuni fra i grandi modelli linguistici (d'ora in poi LLM – *Large Language Model*) che rappresentano riferimenti obbligati per la trattazione. Nello specifico, gli strumenti utilizzati sono stati ChatGPT versioni 4o e o3 (di norma con Deep Research), Manus (una piattaforma basata su Claude Sonnet 4, con funzionalità avanzate di Deep Research) e il modello cinese Kimi K2 con funzionalità Researcher attiva, che nel momento in cui scrivo rappresenta una novità nel campo dei LLM e le cui prestazioni sembrano decisamente interessanti.[1]

L'interazione con questi modelli ha riguardato tre fasi diverse del lavoro: 1. la richiesta al modello di fornire proposte su temi rilevanti da trattare, sulla base di un prompt che spiegava gli obiettivi dell'articolo e le sezioni principali in cui mi proponevo di suddividerlo; 2. la richiesta di approfondimenti specifici, riferimenti ulteriori e suggerimenti su casi rilevanti; 3. l'esame separato dei testi delle varie sezioni dell'articolo, nella mia stesura, con un prompt che chiedeva di individuare errori, punti deboli o riferimenti mancanti.

Le indicazioni fornite sono risultate in diversi casi utili (e in altri casi a mio avviso sbagliate o fuorvianti), ma tanto la stesura quanto la responsabilità del testo sono ovviamente del tutto mie, con una sola eccezione interessante che segnalerò in nota (v. oltre, nota 15). L'obiettivo principale nell'uso di questi sistemi era verificare se e in che misura il ricorso all'IA generativa in forma di *sparring partner* durante la stesura di un lavoro di ricerca potesse risultare utile. La valutazione del risultato spetta ovviamente a chi legge, ma la mia impressione come autore è che l'utilità ci sia (soprattutto nella segnalazione di alcune esperienze e risorse rilevanti e di alcune imprecisioni). Non

[1] Sulle differenze – in particolare in termini di allucinazioni e capacità di risalire a letteratura secondaria rilevante – fra modelli tradizionali e modelli Deep Research si veda (Roncaglia 2025a). Una sintetica spiegazione è fornita anche in chiusura di questo articolo.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-693

vi è invece – né mi aspettavo ci fosse, considerato che non intendevo in alcun modo rinunciare al controllo autoriale – un risparmio di tempo nella stesura: direi semmai che il lavoro ha richiesto più tempo di quello che sarebbe stato necessario senza l'uso di questi strumenti[2].
Oltre a questa premessa, l'articolo comprende due sezioni principali:[3] la prima è dedicata all'evoluzione dei paradigmi catalografici e di metadatazione, con l'obiettivo di rendere chiari i problemi legati alla loro standardizzazione ed evoluzione nel tempo, e il tipo di vincoli che la loro natura

[2] In sede di revisione anonima dell'articolo, pur apprezzando le indicazioni fornite nei paragrafi precedenti sulle modalità d'uso di LLM nel lavoro di stesura dell'articolo, è stato osservato che la disponibilità integrale delle interazioni avute con i LLM utilizzati – prompt usati e risposte ricevute – garantirebbe una trasparenza ancor maggiore. Si tratta di una osservazione mossa da una esigenza largamente condivisibile, quella di massima trasparenza nell'uso di strumenti di rete, e particolarmente di strumenti delicati e potenzialmente fuorvianti come indubbiamente sono i sistemi di intelligenza artificiale generativa, ma (non solo nel caso del mio articolo) anche assai problematica dal punto di vista pratico. Ho provato a fare una valutazione necessariamente parziale (dato che alcuni dei sistemi usati non conservano nel tempo tutte le interazioni avute, o le conservano solo per un periodo di tempo limitato o solo mantenendo attivo un abbonamento a pagamento), e posso dire che complessivamente nel lavorare sul tema i miei prompt e le risposte dei vari sistemi usati corrispondono ad almeno 230.000 (e probabilmente a più di 250.000) battute di testo: ovviamente non sarebbe possibile includerle tutte in questa sede (l'articolo, già in questa forma, supera largamente il limite di 50.000 battute che era stato fissato inizialmente dalla redazione). Inoltre, si tratterebbe di materiali potenzialmente fuorvianti, dato che includono alcune allucinazioni e diverse indicazioni a mio avviso sbagliate o problematiche, ciascuna delle quali richiederebbe una discussione specifica. Astrattamente, se fosse stato usato un singolo sistema capace di offrire una buona garanzia di permanenza nel tempo delle interazioni, con un link persistente e affidabile, ci si potrebbe limitare a fornire il link; ma i risultati andrebbero a quel punto comunque commentati, per spiegarne l'uso specifico; inoltre ho appunto usato più sistemi, e al momento direi che nessuno dei LLM disponibili offra garanzie sull'affidabilità e la persistenza dei link alle interazioni (che, al contrario, spesso spariscono). D'altro canto l'esigenza, certo condivisibile, della massima trasparenza relativamente a fonti e strumenti usati deve sempre – e non solo nel caso dei LLM – fare i conti con la ragionevolezza pratica: per fare solo qualche esempio, nel lavoro di ricerca tutte e tutti noi interagiamo sempre e comunque con colleghe e colleghi con cui discutiamo tesi o parti di un lavoro, ed è ottima pratica riconoscerlo nei ringraziamenti, ove possibile anche in maniera abbastanza specifica; ma – tranne in casi particolari – non è concretamente possibile (e appesantirebbe moltissimo il testo) citare *verbatim* tutte le interazioni avute, anche se queste hanno in alcuni casi contribuito a influenzare o indirizzare il lavoro svolto. Analogamente, tutte e tutti noi usiamo strumenti di ricerca e discovery, e sappiamo bene che ogni strumento di questo tipo utilizza algoritmi di ranking non sempre neutrali o trasparenti; ma di nuovo non sarebbe né possibile né probabilmente desiderabile citare ogni ricerca fatta su Google o Google Scholar (o su altri database bibliografici o discovery tools) nel lavorare su un articolo, o fornire l'insieme delle risposte di volta in volta ottenute. Infine, anche le indicazioni ottenute attraverso il meccanismo di peer review – come quella che sto discutendo in questa sede – indirizzano evidentemente l'elaborazione di un articolo (in questo caso, mi hanno spinto ad aggiungere questa nota e a modificare – spero utilmente – alcuni altri passi del testo): in linea di principio dovrebbero forse essere rese integralmente pubbliche e – dopo la fase di peer review – correttamente e individualmente attribuite. Essendo ormai gestite attraverso piattaforme di mediazione editoriale, la cosa è non solo possibile, ma anche facilmente realizzabile. Tuttavia, di nuovo, quanta parte di questo lavoro di elaborazione è sensato documentare integralmente, considerata l'esigenza di focalizzare l'attenzione del lettore su quel che è centrale, cioè l'articolo come prodotto finale dell'attività di ricerca?
Questo non significa, naturalmente, che il problema della trasparenza nell'uso dei LLM – come di qualunque altra fonte diretta o indiretta, ma ancor più in questo caso, per i caratteri specifici di questo tipo di fonti – non si ponga, e che non sia importante discuterlo, come ho cercato di fare sia nel testo sia in questa nota (anche se una discussione soddisfacente richiederebbe molto più spazio: il tema è assai rilevante, e meriterebbe una sede specifica). Resta fermo, ovviamente, il principio fondamentale di correttezza che impone di riportare ogni fonte diretta utilizzata, e in particolare tutti i testi citati: così, qui come in tutti i miei lavori, ogni citazione diretta o comunque sostanziale dalle risposte di un LLM è sempre chiaramente indicata e attribuita (in questo articolo, come già ricordato, l'unico caso di questo tipo è discusso nella nota 15), e impostazione, tesi generali e contenuto non virgolettato del testo sono miei, e non prodotti da sistemi di IA generativa.

[3] Un'ulteriore tematica di grande interesse è quella della catalogazione e metadatazione di contenuti prodotti dall'IA generativa o risultato di una collaborazione con l'IA generativa: ma discuterla in questa sede richiederebbe ancora più spazio di quello, già notevole, richiesto dalla trattazione dei temi su cui ho scelto di concentrare il lavoro.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-693

pone all'uso di sistemi di IA generativa. La seconda sezione propone la distinzione fra tre diverse strategie d'uso dell'IA generativa in ambito bibliotecario e documentale, con il riferimento ad alcune esperienze concrete in cui sistemi di IA generativa sono stati utilizzati in particolare in contesti di catalogazione e metadatazione.

1. Evoluzione dei paradigmi catalografici e di metadatazione: Da MARC ai nuovi framework nel contesto Linked Open Data

Le origini della standardizzazione in campo catalografico risalgono al XIX secolo, con la definizione di regole precise e uniformi per la realizzazione delle schede cartacee, che rappresentavano l'unità fondamentale di descrizione bibliografica. Regole catalografiche ragionevolmente standardizzate vengono progressivamente individuate nel corso dell'800 tanto nel contesto europeo quanto in quello nord-americano, a partire dalle famose 91 regole attribuite ad Antonio Panizzi[4] e alle indicazioni contenute nei primi 'manuali' bibliotecari, come Namiur (1834). In Italia, il manuale di Giuseppe Fumagalli sottolinea il rapporto fra catalografia e bibliografia descrittiva (Fumagalli 1887, xiii) e nota già chiaramente il carattere semantico del rapporto fra scheda e libro catalogato: "…schede, che sono come i segni rappresentativi dei libri…" (xv).

Il lavoro di affinamento delle pratiche catalografiche su scheda prosegue fino ai Principi definiti nella Conferenza di Parigi del 1961 (Chaplin e Anderson 1963), ma già nel corso della prima metà degli anni '60 diviene chiaro che l'avvento dell'informatica comporta una rivoluzione nel campo della catalogazione: rivoluzione concretizzata nel 1966 con la nascita del formato MARC (Machine-Readable Cataloguing). Sviluppato dalla Library of Congress soprattutto ad opera di Henriette Avram (1975), il formato MARC permette di rappresentare informazioni bibliografiche in maniera non solo standardizzata ma direttamente leggibile e utilizzabile da parte dei sistemi informatici dell'epoca. MARC rappresenta dunque il passaggio dalla scheda cartacea al record digitale, pur mantenendo, almeno inizialmente, una stretta corrispondenza con la struttura e la logica della scheda catalografica tradizionale. I record MARC sono infatti composti da campi, sottocampi e indicatori che codificano i diversi elementi descrittivi (autore, titolo, editore, soggetto…), consentendo la creazione di cataloghi elettronici (OPAC – Online Public Access Catalog) e lo scambio di dati bibliografici tra biblioteche.

Nonostante l'enorme passo in avanti rappresentato da MARC e il suo successo internazionale (con la nascita di diversi formati nazionali, poi parzialmente unificati in UNIMARC), la sua struttura, ancora legata al concetto tradizionale di record catalografico, mostra ben presto i suoi limiti. La rigidità del formato, la difficoltà di estrarre e collegare informazioni specifiche presenti all'interno del record e la mancanza di una semantica esplicita, con la conseguente confusione fra componenti sintattiche e componenti semantiche nella definizione dei record (e a volte l'interferenza fra componenti diverse dello stesso record), rende progressivamente chiara la necessità di nuovi modelli e nuovi standard.[5] È in primo luogo la geniale tesi di dottorato di Barbara Tillett ad avviare nel 1987

[4] Le *Rules for the Compilation of the Catalogue* sono del 1839 e aprono il primo volume del *Catalogue of printed books in the British Museum* pubblicato nel 1841 (Panizzi 1841). Sulla genesi e sul contesto storico e teorico delle regole di Panizzi si vedano (Guerrini e Neri 2020; Guerrini e Gambari 2023).

[5] Per una ricostruzione storica di questo processo, sintetizzato in (Tennant 2002) con lo slogan 'MARC must die', si veda

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-693

una riflessione di tipo nuovo, legata allo studio delle relazioni concettuali fra entità bibliografiche (Tillett 1987).[6] Riflessione che nel decennio successivo si sviluppa soprattutto nell'ambito dell'IFLA Section on Cataloguing e porta all'elaborazione del modello concettuale FRBR, *Functional Requirements for Bibliographic Records*, approvato dalla 63ª Conferenza IFLA di Copenaghen nel 1997 e pubblicato, nella sua prima versione ufficiale, nel 1998 (IFLA 1998).[7]

Nel frattempo, la rapida diffusione di Internet e la nascita del web spingono verso una crescente richiesta di interoperabilità e riuso dei dati, mentre l'idea del semantic web, avanzata alla fine degli anni '90 da Tim Berners Lee, contribuisce a rafforzare la consapevolezza che la standardizzazione delle pratiche descrittive deve basarsi su sistemi classificatori (ontologie) a loro volta concettualmente rigorosi e ben definiti anche dal punto di vista formale. In particolare, dal lavoro sul semantic web nasce il *Resource Description Framework* (RDF), un modello adottato come raccomandazione dal W3C nel 1999 e pubblicato nella sua prima versione ufficiale nel 2004 (la versione 1.1 sarà pubblicata nel 2014).

RDF consente di formalizzare l'attribuzione di metadati descrittivi attraverso l'uso di triple soggetto-predicato-oggetto, in cui il soggetto rappresenta la risorsa da descrivere, il predicato il tipo di proprietà o relazione e l'oggetto il valore assunto da tale proprietà o relazione quando attribuita al soggetto. In un uso documentalmente solido di RDF, soggetto, predicato e oggetto devono essere espressi in maniera rigorosa: ad esempio, se il soggetto fosse l'autore di un libro, lo si dovrebbe identificare univocamente (di norma attraverso il ricorso a un authority file), la proprietà 'autore di' dovrebbe essere presente in una ontologia formale (che deve essere dichiarata), e l'oggetto dovrebbe essere a sua volta espresso in maniera univoca e formalmente corretta, cosa che viene fatta di norma attraverso un Uniform Resource Identifier (URI).

Come si vede, dunque, il framework RDF da solo non basta: si tratta solo di uno dei livelli necessari per costruire un complesso edificio di rappresentazioni formali della conoscenza. E in effetti il lavoro su RDF ha rappresentato la base per la nascita di ontologie formali come il *Web Ontology Language* (OWL), che permette di costruire in maniera rigorosa tassonomie e – in generale – sistemi di classificazione. Nel frattempo, il lavoro sul web semantico si è orientato in direzione dei *Linked Open Data* (LOD), che rappresentano una versione in parte semplificata ma più realisticamente gestibile dell'idea originaria di semantic web.[8] Il nome scelto sottolinea da un lato l'elemento fortemente relazionale di qualunque attività di descrizione e metadatazione,[9] dall'altro il requisito di apertura dei dati, che ne permette il riuso e quindi l'integrazione di ontologie e basi di conoscenze parziali e settoriali in costruzioni di più alto livello.

(Thomale 2010). Nonostante l'auspicio di Tennant, MARC sembra comunque ancora lontano dall'essere definitivamente abbandonato: è interessante notare come alcune delle sperimentazioni – anche recenti – di uso dell'IA di cui parleremo nella seconda sezione avvengano proprio su dati in uno dei molti formati derivati da MARC.

[6] Si vedano (Norouzi 2012) e (Ghiringhelli e Guerrini 2019).

[7] Non provo a sintetizzare la sterminata letteratura sul tema, ma per una breve introduzione al modello FRBR si può vedere (Galeffi e Sardo 2013), per una valutazione aggiornata del modello e del suo impatto è utile (Coyle 2022), mentre per il quadro concettuale complessivo dell'evoluzione qui delineata punti di riferimento utilissimi sono (Guerrini, Crupi, e Peruginelli 2013; Guerrini e Sardo 2018; Guerrini 2022a; 2022b).

[8] Per una descrizione introduttiva di queste tematiche e la bibliografia essenziale al riguardo si veda (Roncaglia 2023, 57-66 e, molto più approfondito, (Tomasi 2022).

[9] Riferimento obbligato in tema è (Zeng e Qin 2022).

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-693

Il lavoro di costruzione di livelli descrittivi man mano più rigorosi e formalizzati riguarda anche l'ambito strettamente bibliografico e documentale. Nonostante il deciso passo avanti comportato dall'adozione del modello entità-relazioni, infatti, il riferimento ai *bibliographic records* presente nello stesso acronimo FRBR lasciava ancora trasparire un'idea di dato bibliografico legata al concetto tradizionale di record. Occorreva dunque costruire anche per i dati (e metadati) bibliografici quel complesso edificio di livelli descrittivi formalizzati necessario alla piena adozione del paradigma degli Open Linked Data e di RDF, superando definitivamente la concezione del record come entità principale e arrivando a una rappresentazione basata su dati di livello più basso, interconnessi e semanticamente ricchi. E occorreva definire formalmente, per ciascuno dei diversi 'piani' di questo edificio, il passaggio dal generale modello entità-relazioni su cui è basato FRBR al più specifico modello delle triple RDF. Infatti, se è vero che una tripla RDF viene molto naturalmente concepita come la dichiarazione di una relazione fra due entità, è anche vero che la centralità delle triple – e dunque degli *statement* – RDF rappresenta in un certo senso una specificazione ulteriore. Questo percorso passa attraverso la definizione degli *International Cataloguing Principles* (ICP), pubblicati in una prima versione nel 2009 dopo sei anni di discussioni negli IFLA Meetings of Experts on an International Cataloguing Code, e dello standard *Resource Description and Access* (RDA), del 2010, che fornisce elementi, linee guida e indicazioni specifiche per la creazione di metadati relativi al patrimonio culturale, non solo (ma anche) in ambito bibliografico e documentale.[10] Passa attraverso la progressiva creazione di una vera e propria 'famiglia' di standard FR, in cui a FRBR si affiancano i *Functional Requirements for Authority Data* (FRAD) e i *Functional Requirements for Subject Authority Data* (FRSAD), coordinati poi attraverso il *Library Reference Model* (LRM) sviluppato dall'IFLA FRBR Review Group e pubblicato nel 2017 (IFLA 2017).[11] Per quanto riguarda specificamente l'adozione in ambito bibliotecario del paradigma Linked Open data, fondamentale è anche il *Final Report* del W3C Library Linked Data Incubator Group, del 2011.[12]

È in questo contesto – che resta, come si vede, in continuo e progressivo affinamento concettuale – che si sviluppano oggi anche gli standard più strettamente bibliografici e catalografici. ISBD (*International Standard Bibliographic Description*), nato per promuovere il controllo bibliografico universale ma inizialmente assai condizionato dalla sua natura di 'mediazione' fra le diverse pratiche descrittive nazionali, ha visto così la sua portata ampliata per "migliorare la portabilità dei dati bibliografici nell'ambiente del Web Semantico e l'interoperabilità dell'ISBD con altri standard di contenuto" (IFLA 2011). Questo obiettivo è stato perseguito attraverso la pubblicazione di un insieme di elementi ISBD in RDF, inizialmente nell'Open Metadata Registry nel 2015 e successivamente, dal 2020, come Vocabolari ISBD nei Namespaces IFLA.

Fra gli standard catalografici prodotti dalla 'rivoluzione FRBR' un ruolo di primo piano spetta sicuramente a BIBFRAME (*Bibliographic Framework*), sviluppato dalla Library of Congress come potenziale sostituto di MARC. In Italia, sono influenzate da FRBR le *Regole italiane di catalogazione* (REICAT, ripensamento abbastanza radicale delle precedenti RICA). In entrambi i casi, la scel-

[10] Su RDA una utile rassegna è fornita da (Bianchini e Guerrini 2016).

[11] Si vedano (Bianchini 2017; Guerrini e Sardo 2018). L'ultima versione consolidata è (IFLA 2024).

[12] W3C Library Linked Data Incubator Group 2011. Sull'uso dei linked data in ambito bibliotecario e documentale si vedano (Baker 2013; Guerrini, Crupi, e Peruginelli 2013; Guerrini e Possemato 2015).

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-693

ta di FRBR come punto di riferimento permette una progressiva evoluzione verso l'uso di RDF e LOD e (auspicabilmente) verso il contesto LRM. Il risultato è un cambiamento radicale nel modo di concepire e utilizzare gli standard catalografici: non più regole per la compilazione di record, ma veri e propri modelli concettuali espressi in linguaggi formali che definiscono classi, proprietà, vincoli e relazioni semantiche tra gli elementi descrittivi. Un dato, questo, che – come si vedrà – è estremamente rilevante volendo progettare e utilizzare sistemi di intelligenza artificiale utilizzabili in ambito bibliotecario e documentale.
Va notato peraltro come questa evoluzione renda sempre più angusta l'idea di standard catalografici nazionali o legati a singole istituzioni: strumenti come il *Virtual International Authority File* (VIAF) per gli autori, *GeoNames* per i luoghi, o le ultime versioni del *Library of Congress Subject Headings* (LCSH) e i thesauri LOD per i soggetti, indicano una strada di internazionalizzazione che sembra imposta dalla natura stessa dei sistemi adottati e – soprattutto – dal passaggio dalla catalogazione locale alla metadatazione. Passaggio che deve avvenire all'interno di uno spazio concettuale condiviso e collocato a un più alto livello di astrazione. Una evoluzione di cui dovranno evidentemente tenere conto gli sviluppi futuri di SBN e delle nostre pratiche catalografiche, in particolare nel processo verso l'Indice 3 SBN: ma queste tematiche sono sviluppate, assai meglio di quanto non potrei fare io, in altri contributi a questo fascicolo.

2. Metadati e IA generativa: un rapporto problematico ma promettente

La rapidissima sintesi proposta nella sezione precedente può sembrare forse superflua in un intervento centrato in primo luogo sull'uso di strumenti di IA generativa, ma risponde a una funzione specifica: quella di ricordare che nel mondo della metadatazione bibliotecaria si manifesta ormai da tempo una duplice tendenza. Da un lato, la tendenza alla standardizzazione e alla formalizzazione, nell'ambito di modelli basati su architetture della conoscenza a livello di astrazione sempre più alto. Dall'altro, una tendenza alla granularizzazione dei dati. È bene notare che, anche se queste tendenze sono associate, l'una non implica necessariamente l'altra. E che, nell'incontro con l'intelligenza artificiale generativa, possono avere ruoli almeno in parte diversi.
Per illustrare il rapporto fra strumenti di organizzazione strutturata delle informazioni e delle conoscenze e i nuovi paradigmi dell'IA generativa, ho usato altrove (Roncaglia 2023) la metafora dell'architetto e dell'oracolo. L'architetto corrisponde all'idea di creare sistemi descrittivi e organizzativi rigorosi, formalizzati, architettonicamente solidi. L'oracolo corrisponde invece alla generazione statistico-probabilistica di contenuti propria dei sistemi di intelligenza artificiale generativa basati su reti neurali profonde, in cui l'enorme ragnatela numerica di pesi, valori, funzioni di attivazione e rappresentazione vettoriale di unità minime informative (i token) produce risultati sorprendenti ma è almeno parzialmente oscura, sia per la sua complessità, sia per il suo carattere intrinsecamente non deterministico.
Semplificando in maniera estrema, anche a rischio di una certa imprecisione, nella storia dell'intelligenza artificiale si possono distinguere almeno due grandi fasi. Quella iniziale, fra gi anni '50 e gli anni '90 del secolo scorso, in cui si è lavorato soprattutto partendo da una concezione logico-simbolica e linguistica dell'intelligenza, e durante la quale si è cercato di costruire sistemi artificiali (inizialmente con ambizioni maggiori; man mano – dopo il fallimento del programma

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-693

'forte' iniziale, divenuto chiaro a metà degli anni '70 – in ambiti più ristretti e limitati) guidati per lo più da regole deterministiche a loro volta di natura prevalentemente linguistica e logico-simbolica. Poi, negli ultimi decenni e con una qualche sovrapposizione dei due approcci negli anni '80 e '90, la fase legata all'uso di reti neurali profonde e alla loro trasformazione in senso statistico-probabilistico, che ha portato ai sistemi di intelligenza artificiale generativa di oggi.[13] La metafora dell'architetto e dell'oracolo può essere applicata, almeno a grandi linee, anche a questa evoluzione e alle due fasi che la caratterizzano. Ovviamente, come ogni metafora, anche questa va presa *cum grano salis*: il riferimento all'oracolo vuole sottolineare il meccanismo predittivo e non deterministico di generazione, ma non implica affatto che i sistemi generativi siano infallibili, né che siano totalmente inesplicabili. D'altro canto, l'IA logico-simbolica, pur avendo deluso almeno in parte le aspettative iniziali, non è affatto morta, ed è abbastanza probabile che in futuro le due metodologie si intrecceranno sempre più spesso (la cosiddetta strada 'neurosimbolica', che molti osservatori vedono come la più probabile evoluzione futura del settore)

In ogni caso, la prima delle tendenze sopra ricordate relativamente all'organizzazione delle informazioni e delle conoscenze, quella alla standardizzazione e alla formalizzazione, è chiaramente architettonica. Potrebbe dunque sembrare, almeno in prima istanza, assai lontana dall'IA generativa e semmai più direttamente compatibile con l'IA logico-simbolica classica, che era basata su regole esplicite e su risultati prevedibili e sicuri. Come pensare, per fare solo l'esempio più ovvio, di affidare una qualsiasi forma di metadatazione standardizzata di contenuti a sistemi che possono produrre – e di fatto producono abbastanza spesso – risposte totalmente errate (allucinazioni), anche e particolarmente in campo bibliografico?[14]

D'altro canto, i sistemi di IA basati sul *machine learning* (ML), compresi quelli generativi, amano l'abbondanza di dati granulari, e ancor più di dati granulari affidabili, che risultano particolarmente adatti al loro addestramento. L'esplosione dei big data – di cui anche i LOD fanno a loro modo parte – costituisce anzi una delle precondizioni che spiegano i progressi realizzati negli ultimi anni in intelligenza artificiale. Il problema, però, è che i grandi modelli linguistici (che rappresentano la tipologia di IA generativa più avanzata e – almeno al momento – di più diretto interesse per il contesto discusso in questa sede) non sono addestrati solo o principalmente su dati granulari formalizzati, ma su enormi corpora testuali eterogenei e poco strutturati. Nell'universo di contenuti che alimenta l'IA generativa, i dati bibliografici e in generale i sistemi di dati formalizzati sono una piccola isola, per molti versi felice (si pensi al ruolo dei LOD in Wikipedia e al particolare valore di Wikipedia rispetto alla grande maggioranza delle risorse web), ma del tutto minoritaria.

In questa situazione, un incontro proficuo fra metadatazione bibliotecaria e IA generativa richiede l'adozione di strategie specifiche. Suggerisco di distinguerne tre: 1) creazione di sistemi IA specializzati, addestrati (spesso in forma di *machine learning* tradizionale, più che attraverso le nuove architetture proprie dell'IA generativa attuale) unicamente attraverso dati ben costruiti e standardizzati. 2) collaborazione fra sistemi architettonici tradizionali (ad esempio, una base di dati

[13] Non è ovviamente possibile ripercorrere in questa sede la storia dell'evoluzione dell'intelligenza artificiale; mi limito a ricordare, fra i testi di riferimento al riguardo, il classico (Russell e Norvig 2020). Ricostruzioni di questo processo più dettagliate di quella possibile in questa sede sono in (Roncaglia 2023) e in (Roncaglia 2025b).

[14] Sulle allucinazioni bibliografiche rimando a (Roncaglia 2025a).

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-693

strutturata e affidabile) e sistemi generativi. È la strada, in particolare, della *Retrieval Augmented Generation* (RAG); 3) costruzione di sistemi generativi che, pur conservando la flessibilità caratteristica dei grandi modelli linguistici, siano indirizzati, attraverso una o più fra le molte strade esplorate negli ultimissimi anni (e talvolta negli ultimi mesi: uso di tecniche avanzate di prompting e di ragionamento, selezione di un contesto adeguato, *deep research, fine tuning* specifico, capacità agentiche…), verso la produzione e/o la gestione di metadati standardizzati e affidabili.
In estrema sintesi, le tre strategie sopra delineate *rappresentano altrettanti tentativi di far dialogare i due approcci apparentemente antitetici dell'architetto e dell'oracolo. La prima cerca di 'addomesticare' l'IA confinandola in un ambito strutturato: per garantire l'obiettivo del rigore formale depotenzia o fa scomparire l'oracolo. La seconda cerca di integrare IA generativa e sistemi strutturati, e rappresenta una forma di collaborazione diretta fra l'architetto e l'oracolo, lasciando all'uno e all'altro le proprie caratteristiche. La terza cerca di guidare l'IA generativa verso la produzione di strutture: in sostanza cerca di spingere l'oracolo ad acquisire anche le competenze dell'architetto.*[15]
La distinzione di queste tre strategie va intesa, è bene chiarire subito, come assolutamente provvisoria e funzionale unicamente a fornire un primo schema espositivo a una tematica in rapidissima evoluzione. È del tutto possibile (e per certi versi anche probabile) che fra queste strategie siano possibili ibridazioni e contaminazioni, così come è possibile (e anzi probabile) la nascita di prospettive del tutto nuove. Inoltre, come si vedrà in conclusione, le prospettive aperte dai modelli locali di piccole e medie dimensioni possono contribuire a cambiare il quadro di riferimento e suggerire nuove e diverse sperimentazioni. Trattandosi di sviluppi assai recenti, è dunque probabilmente prematuro tentare un bilancio: vale la pena notare che un testo collettaneo e recentissimo che si propone come punto di riferimento sul tema (Balnaves et al. 2025) include contributi certo interessanti ma corrispondenti a strategie e casi di ricerca abbastanza eterogenei, dai quali non emergono ancora né un consenso su modelli e situazioni d'uso, né una chiara visione dell'IA bibliotecaria nella nuova era dell'IA generativa. Tuttavia, siccome un qualche ordine – pur se provvisorio e parzialmente arbitrario – è comunque preferibile al disordine totale, proverò a seguire la tripartizione sopra delineata, fornendo per ciascuna delle tipologie qualche indicazione aggiuntiva e qualche esempio.

2.1. IA specializzate e non-oracolari

La prima strategia si collega per certi versi all'eredità dell'intelligenza artificiale classica, e non a caso corrisponde in primo luogo a sperimentazioni che precedono l'introduzione dell'intelligenza artificiale generativa nel senso contemporaneo del termine, basata su architetture a Transformer, e la sua rapida diffusione: uno sviluppo avviato nel 2017 e accelerato dopo il rilascio pubblico di ChatGPT nel 2022.[16]

[15] Il contenuto di questo specifico paragrafo di testo mi è stato suggerito il 2/8/2025 da Kimi K2 come possibile sintesi delle tre strategie, sulla base di un prompt in cui chiedevo al sistema di proporre critiche e controesempi alla tripartizione da me proposta. Mi è sembrata una sintesi particolarmente felice, e – anche se ho parzialmente riformulato o modificato il testo prodotto dall'IA (che per questo motivo ho marcato in corsivo anziché con l'uso delle virgolette) – si tratta dell'unico caso in cui un passo di questo articolo è stato effettivamente prodotto, nelle sue linee portanti, da un sistema di IA generativa.
[16] Per una ricostruzione storica di questi sviluppi si veda (Roncaglia 2025b).

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-693

Ricorderò qui solo tre casi interessanti, che corrispondono a tre epoche diverse dello sviluppo dell'IA pre-Transformer. Il primo è rappresentato dall'idea di classificazione automatica dei documenti come naturale prosecuzione del lavoro puramente algoritmico di sviluppo di concordanze e indici: un lavoro avviato assai presto, a partire dalla pionieristica esperienza dell'*Index Thomisticus* di Padre Busa.[17] Oggi sembra difficile considerare lavori di questo tipo come pertinenti all'intelligenza artificiale, ma all'epoca si riteneva che il passaggio dal semplice ordinamento di lemmi e occorrenze a forme di indicizzazione più astratta potesse avvenire attraverso algoritmi relativamente semplici di analisi statistica e fosse quasi a portata di mano. Lo mostra chiaramente una citazione del lontano 1963: "Raw data, i. e. , unedited natural language text, can be processed statistically so as to automatically assign index terms to each document and to classify the document into a subject category; this has been demonstrated" (Borko e Bernick 1963, 28).[18] Nonostante il notevole lavoro svolto sul tema negli anni '60 e '70 del secolo scorso,[19] anche in questo caso – come in quello più generale dell'AI logico-simbolica – l'ottimismo iniziale non sembra portare a strumenti effettivamente efficaci ed usabili: non a caso, anche questo indirizzo di ricerca sarà di fatto ripreso solo nel nuovo millennio attraverso le nuove metodologie offerte dal *machine learning*.[20]

In sostanziale coincidenza con il periodo in cui l'IA logico-simbolica abbandona l'iniziale fiducia sulla possibilità di costruire sistemi 'generalisti' e si concentra su ambiti più limitati, anche il focus relativo alle sue applicazioni bibliotecarie si sposta su sistemi esperti specifici, in grado di facilitare le ricerche degli utenti: i primi sviluppi in questo campo sono degli anni '80 e '90 del secolo scorso.[21] Un esempio emblematico è l'*Intelligent Reference Information System* (IRIS) *Project* sviluppato dalle biblioteche dell'Università di Houston tra il 1988 e il 1993. Il progetto IRIS ha portato a un prototipo e successivamente a un sistema di reference che assisteva gli utenti nella selezione degli indici e nell'accesso a reti LAN (ad esempio per l'accesso a CD-ROM),[22] il cui codice aperto, in Prolog, è stato distribuito anche ad altre biblioteche.

Con il nuovo millennio l'attenzione si sposta nuovamente, passando dai sistemi esperti settoriali a soluzioni basate su basi di dati accessibili via web. In questo caso, un buon esempio è il sistema esperto per la selezione di database sviluppato nella biblioteca dell'università dell'Illinois da Wei Ma e Timothy Cole fra il 2000 e il 2002.[23] Il sistema si basava su un *selector* che, assumendo il ruolo di un bibliotecario di reference, proponeva all'utente i database più adeguati alla ricerca che si voleva svolgere. L'idea era quella di "analyze and then emulate in software the techniques most

[17] Sulla figura di Padre Busa e sul progetto dell'*Index Thomisticus* si veda (Jones 2018).

[18] Harold Borko, in particolare, ha fornito negli anni '60 e '70 del secolo scorso moltissimi contributi sul tema. Per rendersi conto di quanto poco questo lavoro pionieristico sia ricordato oggi, merita notare che nel momento in cui scrivo solo le versioni francese, spagnola e catalana di Wikipedia gli dedicano una voce, assente invece nella versione inglese (nonostante Borko sia statunitense, abbia lavorato tutta la vita negli Stati Uniti e sia stato presidente della American Society for Information Science).

[19] Ne è buona testimonianza (Stevens 1965), assolutamente impressionante per la mole di riferimenti a linee di ricerca affascinanti e oggi quasi dimenticate, alcune delle quali basate peraltro su metodologie statistiche che conservano un notevole interesse. Per un'idea di queste metodologie, si po' vedere ad es. (Heaps 1973).

[20] Un esempio che rappresenta un po' un ponte fra le esperienze precedenti e gli sviluppi successivi è fornito da (Sebastiani 2002).

[21] Per una bibliografia in merito si veda (Bailey 2018).

[22] Per una descrizione, si veda (Bailey 1990).

[23] Su questa esperienza cfr. (Ma e Cole 2001; Ma 2002).

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-693

commonly used by reference librarians when helping library users identify databases to search. A key component of this approach has been the inclusion in the system's index of database characteristics the controlled vocabulary terms used by most of the databases characterized" (Ma e Cole 2001, 282). Un sistema, dunque, basato sull'uso guidato di vocabolari controllati.

L'epoca successiva è quella del *machine learning*, che porta all'IA generativa attuale. La meta-ricerca di Islam et al. (2025) mostra come i riferimenti a progetti di IA bibliotecaria all'interno della letteratura specializzata crescano notevolmente nel periodo 2009-2018, prima dunque della rivoluzione dei Transformer, delineando un panorama in cui la presenza della Cina diventa maggioritaria rispetto alle pubblicazioni e tende progressivamente a crescere anche rispetto alle citazioni. Una buona rassegna, basata sulla situazione pre-Transformer, è fornita dal report *Machine Learning + Libraries* di Ryan Cordell (2020).

In questa fase si cominciano a usare intensivamente le reti neurali profonde, ed emergono per la prima volta i problemi legati alle allucinazioni: si affaccia dunque l'oracolo – anche se non ancora con le sue capacità attuali – e per limitarne gli effetti si lavora soprattutto con reti addestrate attraverso corpora controllati e con meccanismi di verifica umana (human-in-the-loop[24]). Inizialmente, però, l'obiettivo non è tanto la produzione di metadati o la generazione di testi quanto il lavoro in ambiti in cui le reti neurali pre-Transformer (e dunque pre-LLM) sembrano eccellere, come il riconoscimento dei caratteri e delle grafie, inclusi il recupero di annotazioni manoscritte e l'uso di OCR su testi 'difficili' (molti programmi e progetti di trascrizione automatica o assistita da manoscritti – ad es. il software Transkribus[25] – si basano su sistemi di questo tipo).

A questa fase risale anche l'*IFLA Statement on Libraries and Artificial Intelligence* (IFLA 2020). Un documento di grande interesse, che per certi versi prefigura tematiche assolutamente attuali, a partire dal lavoro sull'AI Literacy, che costituisce oggi la parte probabilmente più importante e delicata dell'information literacy. Corretta pare anche l'identificazione delle possibilità di estrema personalizzazione dei contenuti come una delle conseguenze principali dei nuovi sistemi di IA, così come la sottolineatura del rilievo delle considerazioni etiche e della consapevolezza dei bias che possono essere presenti (e di norma sono presenti) nei corpora di addestramento e nell'output di questi sistemi. D'altro canto, diversi aspetti dell'analisi sono già datati, riferendosi alla situazione pre-LLM.[26]

Sistemi più direttamente orientati all'ambito classificatorio e alla generazione automatizzata di metadati – ma non ancora basati su architetture generative a Transformer – si affacciano fra il 2020 e il 2022. Un buon esempio è Annif, toolkit realizzato dalla biblioteca nazionale finlandese e distribuito in accesso aperto.[27] Annif è un sistema di indicizzazione e classificazione bibliotecaria che riprende e affida al nuovo paradigma del *machine learning* la vecchia idea di indicizzazione automatica. Nel frattempo, si approfondisce l'uso del LOD come dataset per realizzare in *machine learning* applicazioni di generazione di metadati e annotazioni: uno sviluppo ben esemplificato dal

[24] L'espressione è usata in questo caso con riferimento alla partecipazione di operatori umani nella fase di verifica dei risultati forniti dal sistema. In un senso più generale, l'espressione 'human-in-the-loop' è molto usata anche nelle discussioni di etica dell'IA: al riguardo, si veda (De Caro e Giovanola 2025).

[25] Il sito di riferimento è https://www.transkribus.org/.

[26] Un aspetto giustamente sottolineato da (Lana 2023).

[27] Per maggiori informazioni si veda (Suominen et al. 2022).

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-693

progetto SAGE, che lavora sui dati di Europeana usando metodologie di *Natural Language Processing* (NLP) e *machine learning*, con l'obiettivo di "generating, enriching, validating, publishing, and searching RDF data" (Kaldeli et al. 2024, 358).
Progetti basati sul *machine learning* a partire da dati controllati sono anche quelli dell'OCLC (2025), focalizzato sulla scoperta di duplicati catalografici e addestrato anche attraverso segnalazioni degli utenti, standardizzate attraverso la maschera di segnalazione, o alcuni fra i progetti della British Library, come Languid, basato sull'uso di sistemi ML per l'identificazione automatica della lingua dei documenti, finalizzato all'inserimento del dato nei record catalografici MARC (Morris 2019).
Un altro esempio di iniziativa 'ponte' verso la situazione attuale, in cui si affaccia la generazione di metadati, è il progetto *Exploring Computational Description* (ECD) della Library of Congress, avviato nel 2022.[28] L'obiettivo è valutare l'efficacia dei modelli di *machine learning* nella generazione di record bibliografici di alta qualità su larga scala. Il progetto si concentra sull'utilizzo di reti neurali addestrate su un vasto corpus di dati bibliotecari, composto da alcune decine di migliaia di libri ed ebook e dai relativi record MARC, che fungono da *ground truth* per l'addestramento e la valutazione dei modelli (sono stati testati 5 diversi modelli, presumibilmente basati su architetture di rete neurale diverse, anche se la documentazione che ho potuto reperire non fornisce indicazioni specifiche al riguardo).
Le metodologie impiegate nel progetto ECD includono in primo luogo la classificazione del testo, utilizzata per suggerire i *Library of Congress Subject Headings* (LCSH). In questo caso, un sistema di *machine learning* analizza il testo degli ebook per identificare gli argomenti principali e li traduce in termini di soggettazione LCSH. Questo approccio consente di automatizzare una parte complessa e dispendiosa in termini di tempo del processo di catalogazione, fornendo suggerimenti basati su un vocabolario controllato e che i catalogatori possono poi rivedere e affinare.
In secondo luogo, viene sperimentato un sistema di classificazione dei token che ha lo scopo di suggerire i nomi degli autori, come stabiliti nel *Library of Congress Name Authority File* (NAF). In questo caso, il modello di ML è addestrato per riconoscere e prevedere con precisione le stringhe di caratteri che costituiscono il nome dell'autore all'interno delle pagine del titolo, facilitando l'associazione corretta con le forme standardizzate dell'authority file.
I limiti di questa prima strategia, legata all'uso di AI classica o di *machine learning* basato su corpora controllati, sono fondamentalmente due. In primo luogo, anche se i corpora di partenza sono strettamente standardizzati, nei modelli più recenti basati su reti neurali profonde il rischio di allucinazioni non è totalmente evitato: si tratterà di allucinazioni che il sistema avrà cura di inserire nella struttura formale utilizzata – ad esempio, l'uso di un soggetto sbagliato, ma presente nel vocabolario controllato adottato e collocato nel luogo corretto di un record catalografico – e che però proprio per questo possono essere particolarmente difficili da individuare. In secondo luogo, in sistemi di questo tipo manca l'enorme flessibilità garantita dall'uso del linguaggio naturale: per quanto desiderabile e preziosa sia la formalizzazione, la costruzione di una sorta di recinto chiuso attorno a dati e metadati bibliografici potrebbe non essere particolarmente desiderabile, soprattutto in una prospettiva di integrazione dell'AI in uno spettro più vasto di funzionalità bibliotecarie.

[28] La descrizione che segue è basata su (Saccucci e Potter 2025) e su altri materiali – incluso il video di una conferenza di presentazione – disponibili in (Saccucci e Potter 2024a; 2024b).

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-693

Il primo problema è tipicamente affrontato attraverso l'enfasi sul controllo umano (*Human-In-The-Loop*), che però è spesso assai 'costoso' in termini di impegno delle risorse umane, e non è particolarmente gratificante: passare il proprio tempo a correggere errori nei metadati prodotti dall'IA sembra essere un compito assai più faticoso e alienante della produzione diretta degli stessi metadati da parte del bibliotecario, anche se alla fine i risultati sono quantitativamente di più e qualitativamente più ricchi. Il secondo problema non è superabile senza il ricorso a un LLM. Non sorprende, dunque, che negli ultimissimi anni – tendenzialmente, dopo il 2022 – l'attenzione si sia spostata sui risultati spettacolari ottenuti dalle architetture a Transformer e dai grandi modelli linguistici.

2.2. Tentativi di collaborazione: la RAG

Arriviamo così alla seconda fra le strategie sopra delineate, che prevede una collaborazione fra LLM (ma eventualmente anche sistemi generativi multimodali) e basi di conoscenze esterne affidabili e controllate. Tipicamente, questa collaborazione sfrutta la base di dati 'tradizionale' – e dunque l'architetto – per svolgere ricerche ed estrarre risultati, mentre usa l'IA generativa da un lato per trasformare l'input in linguaggio naturale in una query strutturata, e dall'altro per riformulare in linguaggio naturale e se necessario tradurre e personalizzare la risposta. Si tratta del modello corrispondente alla cosiddetta *Retrieval Augmented Generation*.[29]

La RAG opera attraverso due fasi principali: il recupero (Retrieval) e la generazione (Generation). Nella fase di recupero, una query dell'utente viene utilizzata per interrogare una base di conoscenza esterna (ad esempio un database di documenti, un catalogo bibliotecario, un archivio digitale) al fine di identificare e recuperare i contenuti informativi più pertinenti. Questi contenuti divengono 'documenti contestuali' e vengono passati, insieme alla query originale, a un LLM. Nella fase di generazione, il LLM utilizza sia la query che i documenti recuperati per formulare una risposta coerente, informativa e basata sulle evidenze. Questo processo assicura che la risposta generata sia non solo fluida e naturale, ma anche accurata e supportata da fonti verificabili. Inoltre, l'uso di un LLM garantisce enormi possibilità di personalizzazione: il sistema può rispondere automaticamente nella lingua dell'utente, può adattarsi a registri linguistici e competenze diverse, può fornire sintesi, può usare – nel caso di sistemi multimodali – anche codici comunicativi diversi, ad esempio producendo una risposta in audio, una sintesi in forma di podcast e, in un futuro non troppo lontano, anche di video (ad esempio un documentario).

L'applicazione più naturale di sistemi RAG in ambito bibliotecario è evidentemente per il reference: un esempio tutto italiano è Alphy, il chatbot realizzato per il portale Alphabetica (IPaC 2024), ma le applicazioni di questo tipo, nonostante la novità dell'impostazione, sono già numerose, non solo in ambito bibliotecario: un buon esempio relativo all'ambito archivistico è ad es. Archibot 3.0, il chatbot per l'interrogazione in linguaggio naturale degli archivi storici e legislativi del parlamento europeo (Kimaid 2024). Strumenti simili, basati su chatbot e RAG, sono stati realizzati o sono in via di realizzazione in diverse realtà bibliotecarie: fra i molti esempi che si potrebbero fare, il progetto "Enriching Collection Access and Use with Generative AI" delle Northwestern

[29] Una rassegna recente sulle metodologie RAG, con ulteriore bibliografia, è (Gan 2025).

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-693

University Libraries (Northwestern University Libraries 2025, con un video dimostrativo). Una rassegna utilissima, accompagnata da una discussione più approfondita degli usi della RAG in ambito bibliotecario, è in (La Gorga, Testa, e Verna 2025)[30].

Progressivamente, l'uso di LLM integrati da basi di conoscenze affidabili si è esteso negli ultimi anni anche alla generazione di metadati. I primi esperimenti affidati unicamente a LLM (in particolare nel biennio 2023-2024) si erano infatti scontrati subito con il problema delle allucinazioni,[31] e il ricorso alla RAG – assieme all'enfasi sul ruolo del controllo umano, con il già ricordato paradigma Human-In-The-Loop – si è ben presto imposto come una delle strade migliori per limitare il problema. Così, ad esempio, il progetto "Improving Access to Digital Collections Using GenAI in Libraries" dell'Ontario Council of University Libraries (OCUL) e da Scholars Portal, che mira a migliorare la qualità dei metadati e dunque la reperibilità di quasi 50.000 documenti governativi distribuiti in forma aperta sull'Internet Archive (molti in forma di semplici scansioni), utilizzando OCR integrati da LLM per la trascrizione dei testi e la generazione dei metadati tratti da thesauri controllati.

Architetture RAG e/o sistemi di ML addestrati su dati specifici sono molto probabilmente alla base anche di molti fra i servizi AI-based offerti al momento dalle grandi aziende attive nel settore della mediazione informativa e dell'editoria accademica, anche se spesso i progetti di questo tipo non forniscono informazioni complete sui modelli utilizzati. Molte applicazioni di questo tipo sono legate alla fase di reference e discovery, ma non mancano quelle legate alla generazione automatica o assistita di metadati, come l'AI assistant presente nell'editor di metadati ALMA di Clarivate – Ex Libris.[32] Una rassegna completa delle offerte di questo tipo esula però dai limiti di questo articolo e si scontrerebbe probabilmente con notevoli problemi di trasparenza, fattore su cui la comunità bibliotecaria deve naturalmente riflettere: la selezione dei servizi e degli strumenti di intelligenza artificiale da usare in questo campo non dovrebbe infatti prescindere da una informazione ragionevolmente corretta e completa sulle metodologie, sulle architetture e sui corpora di addestramenti utilizzati.

2.3. Nuovi sviluppi

Anche se l'architettura RAG sembra per molti versi unire il meglio dei due mondi, non è affatto scontato che rappresenti l'unica strada, o la strada necessariamente migliore, per l'incontro fra l'IA generativa e le esigenze di standardizzazione e validazione proprie del contesto bibliotecario. La terza fra le strategie qui delineate propone una strada diversa: costruire sistemi di IA generativa capaci di imparare a rispondere alle esigenze dell'architetto molto meglio di quelli attuali. Certo, i primi LLM presentavano (e continuano a presentare) enormi problemi di allucinazioni e di ragio-

[30] Ringrazio un revisore anonimo per la segnalazione di questo articolo, che non era ancora disponibile al momento della prima stesura di questo lavoro.

[31] Diversi esempi curiosi – molti dei quali, anche in questo caso, legati al lavoro con record catalografici in formati MARC – sono in (Deng 2023).

[32] La pagina di riferimento per le applicazioni AI offerte da Clarivate-Ex Libris è https://exlibrisgroup.com/ai-we-can-use-and-trust/, che però non fornisce informazioni specifiche sui diversi sistemi (così come non le offre, se non in minima parte, il report ExLibris 2023).

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-693

namento: problemi, come si è detto, legati ai meccanismi statistico-predittivi di generazione, che dipendono da una modellizzazione generale degli usi linguistici e non dal recupero e dalla verifica puntuale di informazioni validate. Ma gli sviluppi più recenti suggeriscono strade nuove per 'agganciare' alla realtà le predizioni dell'oracolo. Anche il progressivo affinamento delle capacità di programmazione (*coding*) – compito evidentemente 'architettonico' e basato sul rispetto rigoroso di regole – di molti modelli generativi attuali mostra che si tratta di una strada in linea di principio praticabile. Vediamo dunque sinteticamente alcune (per l'esattezza, sei) fra le innovazioni che possono risultare più rilevanti rispetto a questa prospettiva.

In primo luogo, va notato il progressivo allargamento della 'finestra di contesto' (*context window*) con cui possono lavorare i modelli. La finestra di contesto è la quantità di informazione che un modello può tenere presente contemporaneamente nel generare contenuti: in sostanza, lo spazio informativo al cui interno può distribuire la sua attenzione. Inizialmente, questa finestra era piuttosto ristretta: le poche righe di un prompt. Progressivamente, con il miglioramento dei modelli, la finestra si è allargata, e questo permette, oltre a prompt più lunghi e articolati, di 'allegare' al prompt materiali specifici che vogliamo siano tenuti presenti nel generare la risposta (ad esempio, testi da tradurre o sintetizzare). Nel momento in cui scrivo, Gemini 2.5 Pro – il modello di Google e uno dei sistemi con la finestra più ampia – permette ad esempio di utilizzare un contesto di oltre un milione di token, e dichiara che presto il contesto disponibile salirà a due milioni di token. Finestre di contesto ampie, accompagnate da metaprompt adeguati, permettono di usare il contesto stesso come strumento di 'ancoraggio' (*grounding*) ai fatti. È una strategia già praticabile (Roncaglia 2024), ma ha diversi limiti: per quanto ampi, i contesti sono comunque piuttosto angusti per istituzioni con un ampio patrimonio, la gestione di contesti con aggiornamento dinamico non è banale, e le allucinazioni – anche se assai più rare – restano comunque possibili soprattutto su interazioni non direttamente legate al contenuto del contesto (contenuto che l'utente di norma non ha modo di conoscere in dettaglio).

In secondo luogo, una prospettiva potenzialmente interessante è legata allo sviluppo dei sistemi multimodali. In questo testo mi sono concentrato soprattutto sui modelli linguistici, i più interessanti dal punto di vista bibliotecario, ma come sappiamo vi sono anche sistemi generativi che producono immagini, suoni, video… I sistemi multimodali di ultima generazione non si limitano a far lavorare insieme più sistemi monomodali, ma sono addestrati direttamente su corpora multimodali. Normalmente si pensa ai 'modi' in termini di codici comunicativi diversi, ma esistono già sistemi che lavorano su modi meno ovvi, ad esempio le temperature o il grado di inquinamento di un ambiente. È possibile considerare come 'modi' anche alcune tipologie di dati standardizzati, e usare le tecnologie proprie dei sistemi multimodali per integrare modi con maggior grado di libertà (come il linguaggio naturale) e modi con minor grado di libertà, come i dati bibliografici? Sullo sfondo, si parla anche di *world models*: modelli del mondo. In teoria, se riuscissimo ad addestrare un modello su molte tipologie diverse di dati, che corrispondono alle diverse tipologie di input percettivi umani (e non solo), potremmo pensare di produrre modelli con la capacità di comprendere il mondo in maniera simile alla nostra, o addirittura più ricca, e quindi con una capacità simile alla nostra nella distinzione fra contesti liberi e creativi e contesti vincolati e strutturati. Creare *world models* non è certo facile, e probabilmente non sarebbe banale neanche mettersi d'accordo su cosa esattamente dovrebbe conoscere e comprendere un modello per diventare 'modello del mondo'. Ma anche questa linea di ricerca ha un ovvio, ed enorme, interesse.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-693

Una terza categoria di innovazioni è legata allo sviluppo dei cosiddetti *reasoning models*. Questi modelli funzionano generando, prima della risposta vera e propria, una descrizione del processo di ragionamento che viene seguito per produrla. E sono sottoposti a un addestramento specifico che riguarda proprio la qualità di questo ragionamento: valutatori umani (e a volte reti generative avversarie) 'premiano' i ragionamenti più funzionali. Il testo del ragionamento aiuta poi a sua volta il modello a produrre risposte di qualità migliore. Già oggi, quasi tutti i LLM di livello più alto permettono di usare – accanto alle risposte secche (modalità *one shot*) – anche la modalità *reasoning*. Anche in questo caso, si può senz'altro pensare di addestrare un modello all'uso di un *reasoning* orientato alla precisione, alla struttura, alla validazione… insomma all'insieme delle caratteristiche proprie di quello che potremmo definire 'reasoning documentale'.
La quarta innovazione da ricordare è rappresentata dalla *deep research*, anche in questo caso una caratteristica già presente nei sistemi più avanzati. Si tratta di un passo ulteriore rispetto ai sistemi che cercano il *grounding* attraverso una semplice ricerca in rete. Sappiamo che le informazioni in rete non sono certo un database certificato e affidabile, anzi, possono contenere a loro volta errori marchiani; per questo la RAG usa basi di conoscenze affidabili e non una semplice ricerca in rete. Tuttavia, i sistemi più recenti non si limitano a 'consultare' la rete ma operano quella che in gergo è detta appunto 'ricerca profonda' (*deep research*): prendono in esame siti diversi, 'ragionando' sulla loro maggiore o minore affidabilità e confrontando fra loro i risultati ottenuti. I sistemi di *deep research* sono abbastanza costosi in termini computazionali, perché devono costruirsi un contesto composto da molte pagine di informazioni, selezionandole e valutandole man mano. Anziché rispondere immediatamente, richiedono di norma una decina o anche una ventina di minuti di lavoro. Ma i risultati che producono sono di qualità nettamente superiore, e in certe situazioni potrebbero avvicinarsi alla qualità della RAG, senza la necessità di costruire architetture esterne complesse. In particolare, ciò potrebbe avvenire (ed è questa la quinta delle sei innovazioni qui discusse) attraverso l'uso di sistemi con capacità agentiche, capaci cioè non solo di rispondere a prompt, ma di compiere autonomamente sequenze di azioni complesse. Sistemi di questo tipo possono svolgere ricerche su basi di conoscenze affidabili (ad esempio si un OPAC) senza bisogno di un'architettura RAG specifica.
Un sesto e ultimo sviluppo da ricordare è il miglioramento dei modelli (inclusi i modelli linguistici) di dimensioni piccole (*Small Language Models*: SLM) e medie (*Medium Size Language Models*: MSLM). Di per sé, non si tratta di una novità: la prima versione di Chat GPT, sviluppata da OpenAI nel 2018, l'anno dopo il già ricordato articolo sull'attenzione, aveva 'solo' 117 milioni di parametri. Con i criteri di oggi, si tratterebbe di un modello piccolissimo. In linea di massima, nel caso dei modelli linguistici oggi parliamo di modelli piccoli (SLM) sotto i 20 miliardi di parametri, di modelli di medie dimensioni fra i 20 e i 200 miliardi di parametri, e di modelli di grandi dimensioni sopra i 200 miliardi di parametri; ma sono criteri relativi, che potrebbero mutare in futuro.
Abbiamo davvero bisogno di tutti i miliardi di parametri dei modelli più grandi? Poter usare modelli di piccole e medie dimensioni avrebbe il vantaggio di poterli installare su computer locali, senza doversi necessariamente affidare (e senza dover necessariamente affidare i nostri prompt, i nostri contenuti, le nostre ricerche) ai grandi fornitori su cloud. Tanto più che nel frattempo abbiamo cominciato a produrre computer più veloci, con processori ottimizzati per il tipo di calcoli richiesti dall'IA generativa. Negli ultimissimi anni, molto lavoro è andato proprio in questa direzione. Diverse aziende (fra i risultati migliori, quelli prodotti da alcune aziende cinesi) hanno rea-

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-693

lizzato e distribuito buoni modelli di medie e piccole dimensioni, con prestazioni a volte migliori dei modelli di grandi dimensioni delle generazioni precedenti. C'è però un elemento da tener presente: almeno finora, per avere un buon modello di piccole o medie dimensioni dobbiamo 'distillarlo' da un modello di grandi dimensioni, ad esempio riducendo il numero di cifre decimali dopo la virgola nei parametri, o con altre tecniche che permettono di 'potare' – *pruning* – il modello di partenza. I modelli di piccole e medie dimensioni rappresentano una strada molto interessante, ma per ora non permettono davvero di risparmiare le risorse – come si è visto, enormi – necessarie ad addestrare modelli di grandi dimensioni. Tuttavia, possono permettere di utilizzare in locale tecniche di *grounding* attraverso il cosiddetto *fine tuning*: al modello originale viene aggiunto una sorta di mini-modello locale costruito attraverso dati standardizzati e affidabili, che 'indirizza' le risposte complessive verso una maggiore precisione. Inoltre, l'uso di modelli locali avrebbe il notevole vantaggio di conservare in sede locale dati sensibili (ad esempio, dati sull'utenza o contenuti sotto copyright), aspetto sicuramente desiderabile in un contesto in cui i big player usano in maniera abbastanza disinvolta la loro posizione dominante per fare incetta di informazioni con politiche d'uso non sempre trasparenti (e, in prospettiva, non sempre rispondenti alle indicazioni dell'*AI Act* europeo).

Le innovazioni qui considerate sono troppo giovani per essere valutate attraverso concreti casi d'uso in ambito bibliotecario o documentale,[33] ma è probabile che nel prossimo futuro almeno alcune di queste possibilità (e magari altre) saranno esplorate anche nel nostro mondo.

Una considerazione conclusiva va fatta infine sulle tipologie di dati e metadati prodotti nelle prime sperimentazioni in ambito bibliotecario. In molti dei casi sopra ricordati, il lavoro riguarda ambiti specifici e, quando si lavora su dati catalografici, vengono spesso usati dati concettualmente 'vecchi': ho già sottolineato la frequente presenza nelle sperimentazioni fatte con IA generativa di dati catalografici in formati ricavati da MARC (UNIMARC. MARC 21 ecc.). Certo, MARC è indubbiamente un buon esempio di organizzazione strutturata delle informazioni, e certo, come si è detto, i formati MARC sono ancora assai diffusi; ma come si è visto nella prima parte dell'articolo, sono ormai da tempo concettualmente superati. Sarebbe bene che se ne tenesse conto, anche e soprattutto nel lavoro con l'IA generativa. Tanto più che questi sistemi, se utilizzati con il *reasoning*, sono probabilmente più adatti a gestire il livello di astrazione concettuale tipico dei modelli entità-relazioni e di RDF. Applicare modelli IA di punta su strutture di dati concettualmente obsolete potrebbe non essere – soprattutto nel medio e lungo periodo – una strategia particolarmente produttiva. A ulteriore conferma dello stretto legame che deve necessariamente esistere tra modelli concettuali di metadatazione e descrizione dell'universo informativo, e strumenti – anche generativi – per la produzione e gestione di tali dati.

[33] Includo però il riferimento a un esempio recentissimo che mi sembra particolarmente interessante: (D'Souza et al. 2025). In questo caso, un team di partecipanti ha lavorato presso il TIB Leibniz Information Centre for Science and Technology di Hannover su un progetto di classificazione automatica per il TIBKAT, il catalogo della Technische Informationsbibliothek, la biblioteca nazionale tedesca per i campi dell'ingegneria, tecnologia e scienze naturali. Il team si è diviso lo studio di possibili strategie diverse d'uso dei LLM nella generazione dei metadati, incluse diverse fra le strategie qui ricordate.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-693

Riferimenti bibliografici*

Avram, Henriette D. 1975. *MARC. Its History and Implications*. Washington, DC: Library of Congress.

Bailey, Charles W., Jr. 1990. "Building Knowledge-Based Systems for Public Use: The Intelligent Reference Systems Project at the University of Houston Libraries." In *Convergence: Proceedings of the Second National Conference of the Library and Information Technology Association, October 2-6, 1988,* a cura di Michael Gorman, 190-4. Boston: American Library Association.

Bailey, Charles W. 2018. "Artificial Intelligence in Libraries in the Late 1980's and Early 1990's." *DigitalKoans*. https://digital-scholarship.org/digitalkoans/2018/08/02/artificial-intelligence-in-libraries-in-the-late-1980s-and-early-1990s/.

Baker, Thomas, 2013. "Designing data for the open world of the web." *JLIS.it* 4 (1): 63-6. https://doi.org/10.4403/jlis.it-6308.

Balnaves, Edmund, Leda Bultrini, Andrew Cox, e Raymond Uzwyshyn, a c. di. 2025. *New Horizons in Artificial Intelligence in Libraries*. Berlin, Boston: De Gruyter Saur.

Bianchini, Carlo. 2017. "Remarks about IFLA Library Reference Model." *JLIS.It* 8 (3): 86-99. https://doi.org/10.4403/jlis.it-12416.

Bianchini, Carlo, e Mauro Guerrini, a c. di. 2016. "RDA, Resource Description and Access: The metamorphosis of cataloguing." *JLIS.it* 7 (2). http://jlis.it/index.php/jlis/issue/view/15.

Borko Harold, e Myrna Bernick. 1963. "Automatic document classification." *J. Assoc. Comput. Mach* 10: 151-62.

Cordell, Ryan. 2020. *Machine Learning + Libraries: A Report on the State of the Field*. Commissioned by LC Labs Library of Congress. https://labs.loc.gov/static/labs/work/reports/Cordell-LOC-ML-report.pdf.

Coyle, Karen. 2022. *FRBR, Before and After: A Look at Our Bibliographic Models*. Chicago: ALA.

De Caro, Mario, e Benedetta Giovanola. *Intelligenze. Etica e politica dell'IA*. Bologna: Il Mulino 2025.

Deng, Sui. 2023. "AI, Cataloging & Metadata." *STARS Faculty Scholarship and Creative Works*, University of Central Florida. https://stars.library.ucf.edu/ucfscholar/1251/.

D'Souza, Jennifer, Sameer Sadruddin, Holger Israel, Mathias Begoin, e Diana Slawig. 2025. *SemEval-2025 Task 5: LLMs4Subjects -- LLM-based Automated Subject Tagging for a National Technical Library's Open-Access Catalog*. https://arxiv.org/abs/2504.07199v1.

ExLibris. 2023. *The Impact of Generative AI on Libraries*. https://discover.clarivate.com/The_impact_of_Generative_AI_on_libraries.

Fumagalli, Giuseppe. 1887. *Cataloghi di biblioteche e indici bibliografici*. Firenze: G.C. Sansoni.

* Ultima data di consultazione dei siti web: luglio 2025.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-693

Galeffi, Agnese, e Lucia Sardo. 2013. *FRBR*. Roma: Associazione italiana biblioteche.

Gan, Aoran, Hao Yu, Kai Zhang, Qi Liu, Wenyu Yan, Zhenya Huang, Shiwei Tong, e Guoping Hu. 2025. "Retrieval Augmented Generation Evaluation in the Era of Large Language Models: A Comprehensive Survey." https://arxiv.org/abs/2504.14891.

Ghiringhelli, Lapo, e Mauro Guerrini. 2019. "Entità, attributi e relazioni bibliografiche: rileggendo la tesi PhD di Barbara B. Tillett trent'anni dopo." *AIB Studi 58* (3): 417-25. https://doi.org/10.2426/aibstudi-11868.

Guerrini, Mauro. 2022a. *Dalla catalogazione alla metadatazione. Tracce di un percorso*. Seconda edizione a cura di Denise Biagiotti e Laura Manzoni. Roma: Associazione italiana biblioteche.

Guerrini, Mauro. 2022b. *Metadatazione*. Milano: Editrice Bibliografica.

Guerrini, Mauro, Gianfranco Crupi, e Ginevra Peruginelli, a c. di. 2013. "Global Interoperability and Linked Data in Libraries: Special issue." *JLIS.it* 4 (1). https://www.jlis.it/index.php/jlis/issue/view/23.

Guerrini, Mauro, e Tiziana Possemato. 2015. *Linked data per biblioteche, archivi e musei*. Milano: Editrice Bibliografica.

Guerrini, Mauro, e Lucia Sardo. 2018. *IFLA Library Reference Model (LRM). Un modello concettuale per le biblioteche del XXI secolo*. Milano: Editrice Bibliografica.

Guerrini, Mauro, e Franco Neri. 2020. "La tormentata formulazione delle Regole del British Museum del 1839." In *Scaffali come segmenti di storia: studi in onore di Vincenzo Trombetta*, a cura di Rosa Parlavecchia e Paola Zito, 153-65. Roma: Edizioni Quasar.

Guerrini, Mauro, e Stefano Gambari. 2023. "'Definite cataloguing rules set down in writing': Antonio Panizzi's Rules and the catalogue's manifestations." *JLIS.it* 14 (2): 93-9. https://doi.org/10.36253/jlis.it-528.

Heaps Harold Stanley. 1973. "A theory of relevance for automatic document classification." *Information & Computation* 22 (3): 268–78. https://doi.org/10.1016/S0019-9958(73)90310-0.

IFLA (International Federation of Library Associations and Institutions). 1963. *International Conference on cataloguing principles. Paris, october 9, 1961. Report*. A cura di Arthur Hugh Chaplin e Dorothy Anderson. London: IFLA.

IFLA (International Federation of Library Associations and Institutions). 1998. *Functional Requirements for Bibliographic Records. Final Report*. Den Haag: IFLA.

IFLA (International Federation of Library Associations and Institutions). 2011. *ISBD International Standard Bibliographic Description Consolidated Edition*. Den Haag: IFLA.

IFLA (International Federation of Library Associations and Institutions). 2017. *IFLA Library Reference Model. A Conceptual Model for Bibliographic Information*. A cura di Pat Riva, Patrick Le Bœuf, e Maja Žumer. Den Haag: IFLA.

IFLA (International Federation of Library Associations and Institutions). 2020. *IFLA Statement on Libraries and Artificial Intelligence*. https://repository.ifla.org/handle/20.500.14598/1646.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-693

IFLA (International Federation of Library Associations and Institutions). 2022. *IFLA Library Reference Model. 2017-2022 Consolidation*. A cura di Pat Riva, Patrick Le Bœuf, e Maja Žumer. Den Haag: IFLA.

IPaC (Infrastruttura e servizi digitali per il Patrimonio culturale). 2024. Aggiornamenti di progetto. Alphy 1.9.0. 9 agosto 2024. https://ipac.cultura.gov.it/2024/08/09/alphy-1-9-0/.

Islam, Md Nurul, Shakil Ahmad, Mohammad Aqil, Guangwei Hu, Murtaza Ashiq, Majed Mohammed Abusharhah, e Sheikh Abu Toha Md Saky. 2025. "Application of artificial intelligence in academic libraries: a bibliometric analysis and knowledge mapping." *Discov Artif Intell* 5, 59. https://doi.org/10.1007/s44163-025-00295-9.

Jones, Steven E. 2018. *Roberto Busa, S. J., and the Emergence of Humanities Computing: The Priest and the Punched Cards*. London: Routledge.

Kaldeli, Eirini, Alexandros Chortaras, Vassilis Lyberatos, Jason Liartis, Spyridon Kantarelis, e Giorgos Stamou. 2024. "Combining Automatic Annotation with Human Validation for the Semantic Enrichment of Cultural Heritage Metadata." In *CHR 2024 Computational Humaniti. Proceedings of the Computational Humanities Research Conference 2024 Aarhus, Denmark, 4-6 dicembre, 2024*, a cura di Wouter Haverals, Marijn Koolen, e Laure Thompson, 353-68. https://ceur-ws.org/Vol-3834/.

Kimaid, Luís. 2024. "Artificial Intelligence-Driven Archibot: Transforming Access to European Union Parliament Archives." *LegisTech Library*. https://library.bussola-tech.co/p/artificial-intelligence-archibot-eu-parliament.

La Gorga, Angelo, Roberto Testa, e Lorenzo Verna. 2025. "LLMs and Retrieval Augmented Generation (RAG) for libraries." *DigitCult - Scientific Journal on Digital Cultures*,10 (1): 59-73. https://doi.org/10.36158/97912566920714.

Lana, Maurizio. 2023. "Leggere l'IFLA Statement on Libraries and Artificial Intelligence al tempo di ChatGPT." *Biblioteche oggi Trends* 9 (1). https://doi.org/10.3302/2421-3810-202301-004-1.

Ma, Wei. 2002. "A Database Selection Expert System Based on Reference Librarian's Database Selection Strategy: A Usability and Empirical Evaluation." *Journal of the American Society for Information Science & Technology* 53 (7): 567-80.

Ma, Wei, e Timothy W. Cole. 2001. "Test and Evaluation of an Electronic Database Selection Expert System." In *ACRL 10th National Conference, "Crossing the Divide"*. American Library Association. http://hdl.handle.net/11213/16793.

Morris, Victoria. 2019. "Automated Language Identification of Bibliographic Resources." *Cataloging & Classification Quarterly 58* (1): 1–27. https://doi.org/10.1080/01639374.2019.1700201.

Namur, Jean Pie. 1834. *Manuel du bibliothécaire, accompagné de notes critiques, historiques et littéraires*. Bruxelles: J.B. Tircher.

Noruzi, Alireza. 2012. "FRBR and Tillett's Taxonomy of Bibliographic Relationships." *Knowledge Organization* 39 (6): 409-16.

Northwestern University Libraries. 2025. *Enriching Collection Access and Use with Generative AI*. https://collections-and-ai.library.northwestern.edu/.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-693

OCLC (Online Computer Library Center). 2025. *Implementing AI to further scale and accelerate WorldCat de-duplication*, 4 febbraio, 2025. https://www.oclc.org/en/news/announcements/2025/ai-worldcat-deduplication.html.

Panizzi, Antonio, a c. di. 1841. *Catalogue of Printed Books in the British Museum: Volume I*. London: Printed by order of the Trustees of the British Museum.

Rodriguez, Elena Escolano. 2013. "ISBD adaptation to semantic web of bibliographic data in linked data." *Jlis.it* 4 (1): 119-37. https://www.jlis.it/index.php/jlis/article/view/259.

Roncaglia, Gino. 2023. *L'architetto e l'oracolo. Modelli digitali di organizzazione delle conoscenze da Wikipedia a ChatGPT.* Roma-Bari: Laterza.

Roncaglia, Gino. 2024. "IA generativa, system prompt e biblioteche." In *Un incontro di sguardi: biblioteche, libri e lettura come nodi di un reticolo di possibilità creative e generative: scritti in onore di Maurizio Vivarelli*, a cura di Sara Dinotola e Anna Maria Marras, 343-54. Roma: Associazione italiana biblioteche.

Roncaglia, Gino. 2025a. "Verso il reference generativo?." *Biblioteche Oggi Trends* 11 (1). https://www.bibliotecheoggitrends.it/it/articolo/4165/verso-il-reference-generativo.

Roncaglia, Gino. 2025b. *Filosofia dell'intelligenza artificiale*. Milano: Edizioni Corriere della Sera.

Russell, Stuart, e Peter Norvig. 2020. *Artificial Intelligence: A Modern Approach* (4th Edition). Boston: Pearson. https://aima.cs.berkeley.edu/.

Saccucci, Caroline, e Abigail Potter. 2024a. "Exploring Computational Description: LC Labs Planning Framework in action" (presentazione PowerPoint tenuta presso il *Works in Progress Webinar: LC Labs AI Planning Framework in action – Understand, experiment, and implement AI tools that support catalogers*). https://www.oclc.org/research/events/2024/ai-planning-framework-in-action.html.

Saccucci, Caroline, e Abigail Potter. 2024b. "Could Artificial Intelligence Help Catalog Thousands of Digital Library Books? An Interview with Abigail Potter and Caroline Saccucci." *The Signal* (blog). 19 novembre 2024. https://blogs.loc.gov/thesignal/2024/11/could-artificial-intelligence-help-catalog-thousands-of-digital-library-books-an-interview-with-abigail-potter-and-caroline-saccucci/.

Saccucci, Caroline, e Abigail Potter. 2025. "16 Assessing Machine Learning for Cataloging at the Library of Congress." In *New Horizons in Artificial Intelligence in Libraries*, a cura di Edmund Balnaves, Leda Bultrini, Andrew Cox, e Raymond Uzwyshyn, 227-38. Berlin-Boston: De Gruyter Saur. https://doi.org/10.1515/9783111336435-017.

Sebastiani, Fabrizio. 2002. "Machine learning in automated text categorization." *ACM Computing Surveys (CSUR)* 34 (1): 1-47. https://doi.org/10.1145/505282.505283.

Stevens, Mary Elizabeth. 1965. *Automatic Indexing: A State-of-the-Art Report*. Washington, D.C: United States Government Printing Office.

Suominen, Osma, Juho Inkinen, e Mona Lehtinen. 2022. "Annif and Finto AI: Developing and Implementing Automated Subject Indexing." *JLIS.It* 13 (1): 265-82. https://doi.org/10.4403/jlis.it-12740.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-693

Tennant, Roy. 2002. "MARC Must Die." *Library Journal* 127 (17): 26-7.

Thomale, Jason. 2010. "Interpreting MARC: Where's the Bibliographic Data?." *Code4Lib Journal* 11. https://journal.code4lib.org/articles/3832.

Tillett, Barbara B. "Bibliographic relationships: toward a conceptual structure of bibliographic information used in cataloging". Ph.D. dissertation, University of California, Los Angeles, 1987.

Tomasi, Francesca. 2022. *Organizzare la conoscenza. Digital Humanities e web semantico*. Milano: Editrice Bibliografica.

W3C Library Linked Data Incubator Group. 2011. *Final Report*. https://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/.

Zeng, Marcia L., e Qin Jian. 2022. *Metadata*. London: Facet Publishing.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-676

JLIS.it

Collective intelligence. AI for expanding the information content offered to users: the SBN Sommerso Project

Irene Piccari[(a)]

a) Central Institute for the Union Catalog of Italian Libraries and for Bibliographic Information (ICCU)

Contact: Irene Piccari, irene.piccari@cultura.gov.it
Received: 16 July 2025; **Accepted:** 21 October 2025; **First Published:** 15 January 2026

ABSTRACT
The SBN Sommerso project was developed with the aim of integrating bibliographic heritage not yet described in the centralised database into the collective catalog of the National Library Service (Servizio Bibliotecario Nazionale, SBN). To enrich the catalog, the project involves an automated process based on the use of artificial intelligence techniques for the processing of UNIMARC records which, when compared with the data in the SBN, are recognised as identical, similar or new, thus making new bibliographic data or new information relating to library holdings available, while ensuring that the uniqueness of the bibliographic record within the SBN Index is maintained. The automation of activities, which is necessary considering the amount of data managed, is achieved through the use of specific machine learning algorithms, designed and developed for the I.PaC infrastructure and adapted within SBN Sommerso through training on data and metadata specific to the bibliographic domain.
For the clustering and deduplication of entities represented in the records, namely Manifestation, Agent, Series, Work, Subject and Printer's Mark, the algorithms took into account all the specific metadata for each element. The output required from the system involves presenting the user with groups of identical entities and identifying new ones. The system also groups entities that are assessed as similar and presents them to the domain expert user for verification of the results and correct modulation of the evaluative choices made by the AI, for an increasingly rich, accurate and reliable collective catalog.

KEYWORDS
Collaborative cataloging; Artificial intelligence; National Library Service; Bibliographic information; Machine learning.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-676

The SBN Sommerso project: introduction

The SBN Sommerso project was defined and developed by the Central Institute for the Union Catalog of Italian Libraries and for Bibliographic Information (ICCU)[1] as part of a broader initiative devoted to enriching the SBN union catalog[2] within the evolutionary developments of the collective database of the Italian library network (SBN Index).[3]

SBN Sommerso's objective is to feed the SBN Index with UNIMARC[4] data not yet present in the union catalog, through an automated procedure dedicated to recognizing new, identical, and similar records to import into the SBN database, thereby making new bibliographic information or new holdings information from libraries available to its vast user base. The process was carried out with the support of comparison algorithms developed on the basis of rules defined by domain experts as necessary to make evaluative decisions about the degree of equality between distinct records, and with the development of machine learning models, based on statistical-probabilistic computation. As is well known, collaborative cataloging — on which the cooperative model of SBN is based — implies and requires the uniqueness of each record in the Index and its non-duplication, in order to optimize cataloging of the bibliographic heritage and thus, ultimately, to ensure the availability of cataloging records for users seeking to retrieve information and to enable the development of more efficient services. In SBN, this organizational model corresponds to the relational data architecture chosen from the very beginning of the network's development as the technological support for nationwide cooperation. The technological model made available for comparing SBN data and UNIMARC data provided for the synergistic use of rule-based algorithms and machine learning. Both models—the rule-based and the machine learning–based—are used in the process of identifying new, identical, and similar records, and they are applied to different entities: rule-based algorithms for authority entities and machine learning algorithms for entities of the Manifestation type.

How can we decide if data encoded in two different records refer to the same entity? What are the essential characteristics of the entity being compared? Where, in the specific data encoding format, is the information on which to base the evaluation? Which sources in the cataloging domain are most relevant for selecting reliable information? How do we interpret the specific weight of a certain data element in context? These are some of the pieces of information that the artificial intelligence must be able to learn during training, so that its computational power can effectively collaborate in building an ever more populated, participatory, and reliable national bibliographic

[1] ICCU is the central institute of the Ministry of Culture that, in compliance with the autonomy of libraries, coordinates, promotes and manages the SBN network, the SBN interlibrary loan service (ILL SBN), the Manus and Edit16 censuses, and the Registry of Italian Libraries. It develops standards and guidelines for cataloging and digitization. For information on the Institute's history and functions, see the ICCU website (dedicated page: https://www.iccu.sbn.it/it/istituto/).

[2] The Servizio Bibliotecario Nazionale (SBN) is the network of Italian libraries promoted by the Ministry of Culture with the cooperation of the Regions and Universities, coordinated by the ICCU. For more information on SBN, see the dedicated page on the ICCU website at: https://www.iccu.sbn.it/it/SBN/.

[3] The SBN Index is the central database of SBN, populated with the bibliographic records produced by the libraries that are part of the network.

[4] UNIMARC (Universal Machine Readable Cataloging) is a standard format in the international MARC family created to facilitate the exchange of bibliographic records in electronic format through the encoding of bibliographic elements and a defined logical and physical record structure.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-676

database. The SBN Sommerso project thus represents a further step in the long process of automating bibliographic data management that has characterized SBN's history, seeking the support of a powerful technological tool to optimize a procedure that is strategic for increasing the quantity of data in the SBN Index and, in the long term, for improving the overall quality of the data in the union catalog itself.

As of now, the National Library Service involves 103 hubs (Poli) with the participation of more than seven thousand libraries spread throughout the country, comprising a cataloged collection of about 21 million documents and 115 million location holdings. The cooperative cataloging project demonstrates its vitality through the entry of new hubs and libraries into the network. SBN Sommerso aims to meet the need for automation dictated by the volume of data managed by the Index and the necessity of maintaining control over the uniqueness of records, while at the same time making projects to import UNIMARC databases not yet included in the union catalog more feasible (or even possible). This applies both to data from hubs and libraries already participating in the Service and to entities not yet part of it but which could join, contributing their cataloged records to the Index from the initial stage of participation, ensuring no duplication of entities described in bibliographic records and populating the catalog with new, accurate, and reliable descriptions.

Current data management procedures in SBN already make use of algorithms to define rules capable of identifying and suggesting to the user lists of similar or identical records within the Index database to allow their potential merging. The advances over the current situation include, as will be discussed later, the ability to take into account UNIMARC records external to the SBN Index for comparison, the adoption of AI (artificial intelligence)–based algorithms in addition to rule-based ones,[5] and the use of the I.PaC graph[6] as part of the AI service applied to descriptive metadata, of which the SBN Sommerso project is a concrete application.

[5] Artificial intelligence (AI) is the ability of a machine to exhibit human capabilities such as reasoning, learning, planning, and creativity. The EU Artificial Intelligence Act provides, in Article 3, the following definition of an "AI system": "'AI system' means a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments." , as a field of study and as a set of computing technologies, has a long history whose origin can be chronologically placed in the mid-1950s. Traditionally, the beginning of AI as a scientific discipline is traced to the Dartmouth Conference, held in 1956 at Dartmouth College by a group of American mathematicians, among whom was John McCarthy—who coined the term "artificial intelligence" to refer to the research project defined at that event. This project originally aimed to create an artificial intelligence capable of solving problems in all domains, modeled on the abilities attributed to humans. That early ambition, for a so-called *artificial general intelligence* (AGI), was later scaled down, and research focused on systems dedicated to solving problems confined to a single or narrower domain. A rule-based algorithm is one of the earliest strategies used in AI systems, where the steps for solving a problem are first determined and then encoded into a set of rules and sequential steps that make up the algorithm. The limitations of this approach, along with studies conducted in other fields such as neuroscience and cognitive science, led to a shift in research methods with the development of neural networks, whose key feature—and game-changing aspect—is their capacity for *self-learning*. In this approach, the solution to a given problem is not specified a priori by an algorithm; instead, learning occurs on the basis of many examples provided of the correct solution. For this reason, techniques based on neural networks are called *machine learning*. For a clear and effective treatment accessible even to non-experts, see Francesca Rossi (2024), from which the information summarized here is drawn, and Gino Roncaglia (2023a), where, for the topics introduced above, reference is made to chapter 9.

[6] I.PaC, *Infrastruttura e servizi digitali per il Patrimonio Culturale*, is the national data space designed to archive, manage, and enrich digital cultural heritage.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-676

This paper therefore aims to report on the project activities carried out by ICCU in defining the SBN Sommerso project, which is tied both to the goal of achieving a quantitative expansion of the bibliographic heritage made available to users and—looking to the medium and long term, respectively—to increasing the qualitative level of the data present in the Index and to supporting experiments of potentially strategic value relating to training AI models on rules, criteria, and practices specific to the cataloging domain. These experiments are aimed at determining whether two bibliographic records describe the same entity and at presenting the probabilistic computation of similarity between entities, in a perspective of collaborative integration of AI techniques in the work of librarians.

AI in the Library Context: Theory and Practice – a Brief Overview of the State of the Art

Within the library community, there is a strong interest in technologies related to artificial intelligence, both in terms of analyzing their technical specificities and engaging in more theoretical reflections on the impact of their application in the field, as well as in terms of launching projects to experiment with AI technologies applied to library processes and services. In line with the general trend, attention to the topic received further momentum following the availability of generative AI technologies to the general public after the launch of ChatGPT by OpenAI in November 2022.[7]
In the international library landscape, the IFLA Advisory Committee on Freedom of Access to Information and Freedom of Expression (IFLA FAIFE) published in 2020 the *IFLA Statement on Libraries and Artificial Intelligence* (IFLA 2020) with the intent to outline key issues related to the use of AI technologies in libraries and to highlight the role libraries could assume in a society that was envisioned—indeed expected—to be increasingly permeated by these technologies.[8] The Statement opens by acknowledging the transformative power inherent in AI technologies and offers a positive outlook on their societal and innovation impacts. Moreover, the vision proposed indicates that AI technologies can be effectively employed by libraries precisely to strengthen their social mission (IFLA 2020, 1), not only by integrating artificial intelligence and machine learning techniques into daily systems and workflows, but also by actively working to educate users in the conscious use of AI. Libraries are encouraged to broaden their aims on this front by providing algorithmic literacy pathways in addition to digital literacy (IFLA 2020, 2), and by working in synergy with the research community to offer necessary expertise in the ethical and safe management and evaluation of data. IFLA thus seems to assign to libraries and their professionals a role that is not only active but also fundamentally important in particularly critical areas related to the use of AI. These areas range from the possibility of mitigating the worst social impacts on employment—through training and dissemination of the skills needed to face ongoing changes—to influencing the design process of AI technologies by intervening upstream on the quality of training data.

[7] November 22, 2022 is the date on which the U.S. company OpenAI made available an application interface (a chatbot) through which to interact with its generative AI language model, ChatGPT.
[8] For an in-depth reading of the IFLA Statement in light of current AI developments, see (Lana 2023).

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-676

The IFLA Statement therefore presents libraries with a challenging adaptive scenario, but overall a positive assessment of the mediating role that libraries have historically played in the information society. These assumptions are confirmed in a more recent IFLA initiative on AI with the publication of *IFLA Entry Point to Libraries and AI* (2025) by Cox and De Brasdefer (2025),[9] released as an appendix to the updated 2024 IFLA Internet Manifesto (IFLA 2024). The document opens with an overall positive view of the possibility that AI technologies can support and promote library values in terms of access to and creation of knowledge, and it presents a strong point of view on the role of libraries and on the responsibilities of librarians in safeguarding a responsible, equitable, distributed, and sustainable use of these tools. It is intended, primarily, as an agile support tool for librarians in evaluating the ethical use of AI and as guidance for professional debate (Cox and De Brasdefer 2025, 2). At least two specific points of interest can be highlighted from the document cited. The first concerns the confirmation of a key role assigned to library institutions regarding the spheres of influence they can exercise through activities such as training, advising on choices of different AI models, user education in AI use, and advising on and promoting the responsible use of AI. This strongly reaffirms libraries' mediating capacity in access to information and in the use of information tools, even in the era of generative AIs[10] accessible to a general audience. The second point of interest is identified (by the author of this paper) in the very nature of the document as a practical guide. In listing the possible applications of AI in libraries, its potential risks, and the questions that should guide professionals in considering whether to adopt AI technologies, the document implicitly conveys the premise that libraries have the opportunity to initiate projects to experiment with AI use and to stimulate critical thinking and, above all, a collaborative decision-making process about the role of AI in the library field and the sharing of experiences gained (IFLA 2025, 3). This is set in a context in which libraries are active participants in the collective creation of an artificial intelligence designed for the public good.[11]

In parallel with reflections on the opportunities and risks arising from the adoption of AI applications from ethical, social, economic, and organizational perspectives—and on strategies libraries can use to contribute in this scenario—the number of projects introducing the use of AI technologies in library activities is growing. The literature, in the form of essays, reports, surveys, and interviews, reflects the wealth of ongoing experiences in a variety of library institutions and

[9] The document constitutes Appendix I of the updated *IFLA Internet Manifesto* (IFLA 2024).

[10] Generative AIs are neural networks built with advanced machine learning techniques that can even generate content not pre-defined (unlike "traditional" neural networks), starting from an initial input (prompt). They are also characterized by being trained with unsupervised learning techniques, meaning that no human is present to indicate the correct solution to provide as output. ChatGPT is an example of a generative AI. "ChatGPT is a system for interacting with a so-called large language model, that is, a model of human language. In this specific case, with ChatGPT we are conversing with GPT. … GPT means Generative Pre-trained Transformer; it is generative, i.e. it can generate content (in this case text), it is trained (pre-trained) before being used, and it uses a machine learning technique based on the concept of a transformer, i.e. an innovative architecture based on neural networks." (Rossi 2024, 63).

[11] The concept of "AI for good" appears several times in the IFLA Statement (IFLA 2020, 2 and 9) to indicate various areas in which librarians can apply their skills to define what it means to use artificial intelligence oriented toward the public good. The notion of AI developed as a human-centric technology with the ultimate goal of improving human well-being is present in Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024, which establishes harmonized rules on artificial intelligence (AI Act). Finally, for the concept of *AI for social good* see (Floridi et al. 2018).

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-676

the diversity of implementations of AI technologies in library products and services.[12] The introduction of artificial intelligence (including generative AI) and machine learning in libraries has opened new prospects for automated creation, extraction, and enrichment of metadata and catalog records. Recent experiments documented in professional and academic literature show both enthusiasm for the technology's potential and a cautious attitude toward its current limitations, identified primarily in issues of accuracy and reliability[13] in record creation, concerns about conducting descriptive and representational practices in an ethical manner (Corrado 2021), and the instability of results obtained in record generation when using different input data (prompts) or making requests at different times, as seen in experiments with generative AIs like ChatGPT (Taniguchi 2024).

Finally, in April 2024 the *Final Report* produced by the Task Group on AI and Machine Learning for Cataloging and Metadata of the Program for Cooperative Cataloging (PCC)[14] was published (PCC Task Group on AI and Machine Learning for Cataloging and Metadata 2024). The report contains data from a survey conducted to map projects launched to use AI in cataloging and metadata creation activities within the PCC community, and to evaluate their impacts. Based on the survey results, the working group identified a set of themes of particular concern within the community—from the need to clearly dispel notions of an easy or immediate use of AI and machine learning technologies (which in fact require careful consideration of implementation choices and the collaboration of expert catalogers to be effectively integrated as part of technical services in libraries), to concerns about the general lack of resources specifically allocated to AI study and experimentation. The identified themes also formed the basis for developing recommendations and possible actions to coordinate and support the activities undertaken to experiment with AI technologies in the field of cataloging, thus promoting a strategic approach to integrating these tools into library processes and services. The document also provides a map—limited in geographic and chronological scope—of the places and areas of activity in this sector. Among these, a project particularly relevant to this paper is briefly mentioned, due to the substantial similarity of its objectives and technologies used. The reference is to the use of machine learning techniques by the OCLC[15] to improve the quality of the catalog and information retrieval by reducing duplicate

[12] By way of example of the extensive professional literature available on the subject, two contributions are cited here that bring together theoretical reflection on the opportunities and risks inherent in AI technologies with a broad and up-to-date overview of existing projects (Cox and Mazumdar 2022; Balnaves et al. 2025). Additionally, by way of illustration, some contributions can be mentioned regarding the many research efforts in the field of automated subject indexing and classification of documents, using tools ranging from NLP techniques to the use of ChatGPT for the creation of subject headings, highlighting results achieved and limitations found. Although not falling within the specific scope of this work, it was considered useful to cite at least a few of the many works on the topic given the relevance of this area of experimentation in the application of AI to the library sector. Volume 59 (8) (2021) of *Cataloging & Classification Quarterly* is a special issue of the journal entirely dedicated to the theme; see also (Chow, Kao, and Li 2024; Chou and Chu 2022; Saccucci and Salaba 2021; Suominen, Inkinen, and Lehtinen 2022; Manzoni 2023).

[13] See, for example, (Brzustowicz 2023; Chen and Li 2024).

[14] The Program for Cooperative Cataloging is an international program coordinated by the Library of Congress to promote cooperative cataloging.

[15] Online Computer Library Center (OCLC) is a cooperative library organization that offers products, services, and information systems to the libraries and other institutions in its network. It created and maintains WorldCat, the catalog that records the cataloging data of the libraries participating in the OCLC network.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-676

entries in bibliographic data records and ensuring the persistence of correct data. In the opening of Toves' article on OCLC's *Hanging Together* blog, we read

> "Any system aggregating data from thousands of sources needs sophisticated processes that mitigate duplication and ensure the correct data remain".[16]

Based on the same premises, the SBN Sommerso project was developed for the automatic deduplication—using machine learning—of similar and identical records, and for the recognition of new entities to be incorporated into the union catalog of the Italian library network.
In the Italian professional context as well, one can note a growing interest in issues related to artificial intelligence and its applications, as evidenced by contributions in the professional literature across various areas of reflection[17] and application of AI technologies. This latter aspect, in particular, was highlighted by a recent comparative survey carried out as part of the European project TLIT4U,[18] dedicated to the theme of artificial intelligence literacy and to the role of librarians in AI literacy pathways.[19] The comparative survey (conducted in four countries: Italy, Bulgaria, Mexico, and Bangladesh) started from the observation that data on the levels and modes of AI adoption in libraries were lacking, and aimed to investigate the perceived value and impact of AI, as well as the state of the art both in actual practices of AI use in libraries and in training—both the training received and that considered necessary for librarians to acquire new competencies. The analysis concludes that there is limited use of AI technologies by librarians and institutional support that is still insufficient to meet staff training and infrastructural requirements needed for the stable integration of AI in libraries, while at the same time underlining the urgency of adopting these technologies in order to maintain the social role of libraries (Tammaro, Zanichelli, and Del Carlo 2024, 13).

SBN Sommerso: Architectural Framework for Project Development

Operationally, the SBN Sommerso project is an application case of services developed in the context of I.PaC (Infrastructure and Digital Services for Cultural Heritage).[20] In particular, the project makes use of a specific subset of advanced processing services based on artificial intelligence applied to descriptive metadata, dedicated to deduplicating entities represented in catalog records. I.PaC is the national data space designed to archive, manage, and enrich the digital cultural heritage of the National Digital Ecosystem for Culture. Its development is coordinated by the Central Institute for the Digitization of Cultural Heritage (Digital Library) of the Ministry of Culture, as

[16] https://www.oclc.org/en/news/announcements/2023/leveraging-machine-learning-for-worldcat-de-duplication.html; https://hangingtogether.org/machine-learning-and-worldcat-improving-records-for-cataloging-and-discovery/.
[17] Among the many works published in recent years, relevant to the area of interest and context to which this work refers (with no claim to exhaustiveness), the following contributions can be mentioned: (Lana 2023; Ridi 2023; Morriello 2023; Roncaglia 2023b) in Biblioteche oggi Trends, 9 1 (2023), 1, L'intelligenza artificiale per le biblioteche, an issue entirely dedicated to the theme.
[18] TLIT4U – Improving Transliteracy Skills through Serious Games.
[19] The survey and the detailed results for Italy are described in (Tammaro, Zanichelli, and Del Carlo 2024).
[20] For an introduction to the I.PaC project, see the page available at: https://ipac.cultura.gov.it/#cose_ipac.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-676

part of the investments in culture provided by the PNRR (Italy's National Recovery and Resilience Plan).[21] I.PaC is one of the components of the ecosystem and is intended to enable cooperation among all the other environments that make up the ecosystem itself (Bartoli et al. 2024, 17–18).[22] Developed as an infrastructure for the integrated management of Italy's cultural heritage, I.PaC is configured not only to archive and preserve cultural resources but also as a dynamic environment for data enrichment aimed at sharing, using, and reusing of knowledge. As the central core of the national digital ecosystem, I.PaC offers a catalog of services organized into four functional macro-areas (24 ff). The first is dedicated to the ingestion of descriptive data and digital resources (assets); the second is designed to manage and process digital assets; the third and fourth macro-areas are dedicated respectively to Domain Knowledge Graphs, i.e.

> "semantic representations that describe the relationships between cultural heritage entities within specific domains of heritage, such as Archival, Bibliographic, ABAP and Multimedia" (25).

and to Cross-Domain Knowledge Graphs that

> "facilitate the representation of semantic relationships between digital entities belonging to different heritage domains" (26).

Domain Knowledge Graphs therefore aim to model highly specialized domains of knowledge and to capture the specific semantics of relevant entities and relationships, whereas the modeling of the Cross-Domain Graph[23] serves to overcome the descriptive specificities of disciplinary domains in order to foster the integration of information from different sources, to allow the reconstruction of contexts within which cultural heritage objects acquire meaning, and to connect museum objects, bibliographic resources, and archival documents (Bartoli et al. 2024, 32). To ensure cross-domain interconnection and navigation, data are also normalized and enriched through procedures that leverage AI techniques. Domain and Cross-Domain Graphs are fed by descriptive and cataloging metadata contributed by cooperating systems.

> "The informational objects sent to the infrastructure will undergo a series of processing steps in order to obtain enriched 'derivatives'. In other words, through processes of de-duplication, clustering, and

[21] For a detailed reconstruction of the historical and project context of the national digital ecosystem for culture, see pages 11–15 of (Bartoli et al. 2024).

[22] The ecosystem is composed of 4 environments: D.PaC (Digitization of cultural heritage), I.PaC (Infrastructure and digital services for cultural heritage), D.PaaS (Data product as a service), and the Access and Co-creation Platform.

[23] *A knowledge graph* is a structured representation of information that models data, concepts, entities, and relationships among them, representing the entities as nodes and the logical or semantic relationships between the nodes as edges. Below are the definitions of *Domain Knowledge Graph* and *Cross-Domain Knowledge Graph* in the I.PaC context, taken from the I.PaC Glossary. A *domain knowledge graph* is "obtained from a local knowledge graph enriched through rules. If no optimization and normalization is performed in I.PaC, it may coincide with the local knowledge graph. The domain knowledge graph conforms to the schema described in the target descriptive record format, of which it represents the extensional part." A *Cross-Domain Knowledge Graph* is "a graph of semantic relationships between entities (cultural heritage objects, authority records, controlled vocabularies) belonging to different domains of knowledge and descriptions…". The I.PaC Glossary is available online at: https://ipac.cultura.gov.it/wp-content/uploads/2024/08/Allegato-4-Glossario_I.PaC_202408.pdf.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-676

> mash-up based on rules and AI algorithms, I.PaC extends the dynamic and relational character of cultural information heritage" (Bartoli et al. 2024, 33).

At this point in the process—after having been subjected to the sophisticated "enhancement" elaborations described above—the data from participants in the national data infrastructure indeed contribute to the formation and growth of the I.PaC knowledge base.[24] At the center of the national digital ecosystem for culture, and as its technological engine, lies a complex information network fed by the local databases of participating systems. Within this network, the data are made interoperable through a model that standardizes their representational structures by identifying, via a process of semantic modeling, the key entities (concepts, objects) and defining meaningful connections between them. In other words, the original data formats or schemas,[25] which differ because they belong to different contexts or domains, are "converted" into a uniform graph structure. The representational paradigm used also makes the data available for interconnection with other entities present in the broader context of the semantic web.
Within this architectural framework, I.PaC applies AI techniques to the structured data in order to maximize interconnection, intra- and inter-domain navigation, and enrichment.[26] I.PaC applies AI in three areas: digital objects, navigation and user interaction,[27] and descriptive metadata (Bartoli et al. 2024, 34–37). In the area related to descriptive metadata, AI is used to perform clustering, reconciliation, data cleanup, and enrichment operations necessary to improve a data set that is quantitatively significant yet qualitatively heterogeneous, since it is fed by heterogeneous sources. For the purposes of this paper, attention will be focused only on the clustering algorithm because, as mentioned above, it is within the clustering process that algorithms designed to avoid data duplication are employed.[28]

SBN Sommerso: Machine Learning Applied to Bibliographic Data[29]

In the I.PaC context, machine learning algorithms support the production of clusters of duplicate entities, aimed at creating *Super Entities* useful for cross-domain data navigation. These algorithms are employed in the SBN Sommerso project to automate the process of identifying identical, sim-

[24] The I.PaC knowledge base is also fed through the contribution of SBN Index data, an information asset managed by ICCU.
[25] In the case of data exchanged within SBN, this refers to the UNIMARC record format.
[26] There is a strong temptation at this point to invoke the vivid and now inescapable image of *the architect and the oracle* used by Roncaglia to describe, respectively, the models known and practiced so far for implementing knowledge organization systems—linked to an encyclopedic structuring of knowledge—and the emerging statistical/probabilistic paradigms typical of how generative artificial intelligences operate, characterizing the scenario just described as one in which they might coexist. The reference here is to (Roncaglia 2023a).
[27] Through the development of chatbots deployed on portals for cultural heritage access and use.
[28] "To avoid such duplications, AI algorithms were developed that are capable of interpreting the context of entities by analyzing elements such as dates and places, qualifiers and other associated historical information, in order to group entities that are distinct only nominally but are semantically identical." (Bartoli et al. 2024, 36).
[29] SBN Sommerso is an ongoing project; in writing this section the author made use of technical project descriptions shared by the working group during periodic meetings, which have in part been included in the technical documentation. The author gratefully acknowledges those members of the working group who provided support and clarifications regarding the more technical project aspects utilized in this section.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-676

ilar, or new entities and generating the respective lists, including the percentage of semantic similarity detected in lists of similar entities.

From a technical perspective, the processing of UNIMARC records takes place in an architectural context structured into several components, each dedicated to managing different functionality corresponding to various phases of the process. The main phases can be summarized as: data acquisition of the records to be compared, metadata extraction from the UNIMARC records, application of deduplication and clustering algorithms, and finally presentation of results. These steps are outlined below. At the operational level, two distinct procedures have been set up for the acquisition of the bibliographic data to be compared. The system provides, on the one hand, for the upload and loading of SBN Index records via data acquisition modules made available by the I.PaC platform, and on the other hand a workflow dedicated to uploading packages of data in UNIMARC format external to the Index database—essentially those corresponding to SBN Sommerso. Once loaded, the UNIMARC data from the Sommerso are subjected to a formal check through a pre-validation process aimed at ensuring that the records meet integrity and completeness requirements, containing all data designated as mandatory. The system also performs data correctness checks on certain key fields and provides feedback on the results of the pre-validation process, which operators can use to monitor and analyze errors and to make informed choices about how to handle any records flagged as problematic. Once loaded into the system, the UNIMARC records have rules applied to them for extracting the metadata needed for subsequent clustering. For each entity[30] involved in the analysis and comparison process of the "sommerso" data, specific rules have been defined—for example, for identifying the key information to extract and for assigning a weight to a data element associated with the reference entity. The extracted metadata, structured appropriately to facilitate record comparison, are then used—depending on the case—either by the machine learning procedures or by the rule-based algorithms to form groups based on semantic or logical similarity. The system in fact applies both rule-based algorithms and a probabilistic approach involving AI semantic processing[31] to achieve clustering (i.e. grouping based on similarity criteria) as well as deduplication of records.[32] Finally, the system will provide users with feedback on the outcomes of the processing. The main result presented to users will consist of three lists: one for the records judged to be identical (i.e., those for which the algorithms found an exact match between the pairs examined); the list of similar records, which have partial matches and are accompanied by the percentage of similarity computed by the algorithms to assist the experts in the final decision;[33] and the list of new records, i.e. those with no correspondence in the Index database and therefore need to be integrated.

[30] The entities (or nodes, in the I.PaC context) considered in the SBN Sommerso project are Author (Person, Corporate Body, Conference), Series, Work, Manifestation, Printer's Mark, and Subject, defined in the I.PaC Graph relating to the bibliographic domain.

[31] Semantic similarity is defined as the distance between two points in a vector database. The closer the points are, the more likely it is (in the case of the SBN Sommerso project) that they represent the same thing.

[32] Clustering and deduplication are two distinct processes. In brief, one can say that clustering groups similar records based on measures of semantic/statistical distance, whereas deduplication is a specific action that identifies exact or nearly exact duplicates in order to eliminate or consolidate redundancies in the data.

[33] The percentage of semantic similarity returned to the operators helps to assist the experts in the final decision, as final control over the evaluations presented by the system is in any case entrusted to them.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-676

Conclusions

The SBN Sommerso project described here is conceived as a service both for the institutions that wish to integrate their collections—since it is intended to facilitate their entry into the Italian bibliographic infrastructure and to optimize data loading processes—and for the general public, understood as all those who rely on SBN catalog data as a tool for accessing recorded knowledge and for pursuing their various research objectives. It should be noted, however, that the project activities also serve as an experiment useful for the application of artificial intelligence techniques to bibliographic data, now implemented through machine learning. In this respect, the project represents a contribution in the context of an approach that heeds the calls of those in the library community who deem the participation of librarians in AI experiments to be fundamentally important, in order to actively take part in the process of constructing the emerging AI technologies under development in recent years. Being present and actively participating can be interpreted both in the specific context of SBN Sommerso—as acting as the "human in the loop"[34] by integrating one's own judgments and evaluations with those made by the AI system, to improve it in a highly specialized context—and as leveraging one's own expertise and values in the concrete shaping of technologies that serve the collective well-being of humanity and beyond.[35]

[34] "The concept of human in the loop—the human in the process—dates back to the dawn of computing and automation. It refers to the importance of incorporating human judgment and experience into the operations of complex systems (the 'loop' being automated). Today the term describes the training of AIs that incorporates human judgment. In the future, it may become more difficult to remain within the AI's decision-making process." (Mollick 2025, 52–53). On the theme of the human–machine relationship, I would refer to the stimulating reflections offered by Valentina Tanni in her work (Tanni 2025) where she provides, in the opinion of this author, a contribution with a novel vision and breadth in the current debate on the topic: "The holistic approach of cybernetics, which pushed artists and engineers to establish conversations with the machine, is particularly interesting today, at a historical moment in which the discourse on artificial intelligence is instead focused exclusively on the machine's learning and its supposed capacity to think, understand, create. On the one hand there is the computer, inert matter ready to be activated; on the other is the human creator, grappling with the task of making it become intelligent. Dialogue—a process through which it is possible to learn new, indispensable forms of coexistence—has been pushed into the background, supplanted by an ever more widespread rhetoric of substitution." (Tanni 2025, 11–12). And further: "To construct an alternative to the antiquated myth of the machine as an antagonistic subject, it is necessary to recover the perspective of relationship, to explore dialogue, to imagine and test possible forms of coexistence … It is only through the practice of conversation, accompanied by an in-depth study of the political and economic dynamics within which machines come into being, that we can regain some form of agency within the frenzy of technodeterminism." (Tanni 2025, 18).

[35] I would like to evoke here, in closing, the insights offered in (Bridle 2022) by quoting a short passage: "In this book I will also argue for the capacity for intervention and individuality of technology itself—or perhaps of technology yet to come: the moment, prophesied by many, when machines will become self-sufficient, self-aware, perhaps autonomous. Such a moment does not relieve us humans of responsibility or the ability to change our attitudes and behaviors. On the contrary, considering the agency of technology is an opportunity to have a serious and concrete reflection on how to ensure greater justice and equality to all the inhabitants of the planet: human, non-human, and machines. Technology remains, for now, chiefly in our hands, and we are able to repair, restore, and regenerate its engagement with the world and its effects on it." (Bridle 2022, 27).

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-676

References

Bartoli, Margherita, Luigi Cerullo, Lina Antonietta Coppola, Costantino Landino, Antonio Davide Madonna, Antonella Negri, Giovanni Pescarmona, Chiara Fauda Pichet, Margherita Porena, and Valentina Rossetti. 2024. "Verso la creazione di un Ecosistema digitale nazionale per la Cultura." *DigItalia* 19 (2): 11–48. https://doi.org/10.36181/digitalia-00101.

Balnaves, Edmund, Leda Bultrini, Andrew Cox, and Raymond Uzwyshyn. 2025. *New Horizons in Artificial Intelligence in Libraries*. Berlin–Boston: De Gruyter Saur. https://doi.org/10.1515/9783111336435.

Bridle, James. 2022. *Modi di essere. Animali, pianeta e computer: al di là dell'intelligenza umana*. Milano: Rizzoli.

Brzustowicz, Richard. 2023. "From ChatGPT to CatGPT: The Implications of Artificial Intelligence on Library Cataloging." *Information Technology and Libraries* 42 (3). https://doi.org/10.5860/ital.v42i3.16295.

Chen, Suzhen, and Mingyan Li. 2024. "AI for Cataloging and Metadata Creation: Perspectives and Future Opportunities from Cataloging and Metadata Professionals." *Technical Services Quarterly* 41 (4): 317–32. https://doi.org/10.1080/07317131.2024.2394919.

Chow, Eric H. C., T. J. Kao, and Xiaoli Li. 2024. "An Experiment with the Use of ChatGPT for LCSH Subject Assignment on Electronic Theses and Dissertations." *Cataloging & Classification Quarterly* 62 (5): 574–88. https://doi.org/10.1080/01639374.2024.2394516.

Corrado, Edward M. 2021. "Artificial Intelligence: The Possibilities for Metadata Creation." *Technical Services Quarterly* 38 (4): 395–405. https://doi.org/10.1080/07317131.2021.1973797.

Cox, Andrew M., and Suvodeep Mazumdar. 2022. "Defining Artificial Intelligence for Librarians." *Journal of Librarianship and Information Science* 56 (2). https://doi.org/10.1177/09610006221142029.

Cox, Andrew M., and Maria De Brasdefer. 2025. *IFLA Entry Point to Libraries and AI*. IFLA. https://repository.ifla.org/handle/20.500.14598/4034.

IFLA FAIFE (IFLA Committee on Freedom of Access to Information and Freedom of Expression). 2020. *IFLA Statement on Libraries and Artificial Intelligence*. IFLA. https://repository.ifla.org/handle/20.500.14598/1646.

Lana, Maurizio. 2023. "Leggere l'IFLA Statement on Libraries and Artificial Intelligence al tempo di ChatGPT." *Biblioteche oggi Trends* 9 (1): 4–12.

Manzoni, Laura. 2023. "L'indicizzazione semantica in automatico e semiautomatico nelle principali biblioteche europee." *Biblioteche oggi Trends* 9 (1): 57–64.

Mollick, Ethan. 2025. *L'intelligenza condivisa. Vivere e lavorare insieme all'AI*. Roma: Luiss University Press.

Morriello, Rossana. 2023. "Dati e metadati bibliotecari per l'intelligenza artificiale: per un'agency delle biblioteche." *Biblioteche oggi Trends* 9 (1): 38–45.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-676

PCC (Program for Cooperative Cataloging). 2024. "Final Report: PCC Task Group on Strategic Planning for AI and Machine Learning." PCC. https://www.loc.gov/aba/pcc/taskgroup/TG-Strategic-Planning-AI-final-report.pdf.

Ridi, Riccardo. 2023. "Intelligenza artificiale e web semantico: nessi reciproci, ambiguità e definizioni." *Biblioteche oggi Trends* 9 (1): 27–37.

Roncaglia, Gino. 2023a. *L'architetto e l'oracolo: forme digitali del sapere da Wikipedia a ChatGPT*. Bari–Roma: Laterza.

Roncaglia, Gino. 2023b. "Intelligenze artificiali generative e mediazione informativa: una introduzione." *Biblioteche oggi Trends* 9 (1): 13–26.

Rossi, Francesca. 2024. *Intelligenza artificiale: come funziona e dove ci porta la tecnologia che sta trasformando il mondo*. Bari–Roma: Laterza.

Saccucci, Caroline, and Athena Salaba. 2021. "Introduction to Artificial Intelligence (AI) and Automated Processes for Subject Access." *Cataloging & Classification Quarterly* 59 (8): 699–701. https://doi.org/10.1080/01639374.2021.2022058.

Suominen, Osma, Juho Inkinen, and Mona Lehtinen. 2022. "Annif and Finto AI: Developing and Implementing Automated Subject Indexing." *JLIS.it* 13 (1): 265–82. https://doi.org/10.4403/jlis.it-12740.

Tammaro, Anna Maria, Francesco Zanichelli, and Susanna Del Carlo. 2024. "L'IA in biblioteca: quale impatto? I risultati di un'indagine in Italia." *Biblioteche Oggi* 42 (8): 3–14. https://doi.org/10.3302/0392-8586-202408-003-1.

Tanni, Valentina. 2025. *Conversazioni con la macchina. Il dialogo dell'arte con le intelligenze artificiali*. Roma: Tlon.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-665

JLIS.it

SHARE Catalogue: Cooperation and Semantic Innovation in a Shared Library Ecosystem

Roberto Delle Donne[(a)]

a) University of Naples Federico II, https://orcid.org/0000-0001-8331-9436

Contact: Roberto Delle Donne, delledon@unina.it
Received: 20 June 2025; **Accepted:** 20 October 2025; **First Published:** 15 January 2026

ABSTRACT
This article presents the SHARE Catalogue project as an advanced model of library cooperation and semantic innovation. Developed through an agreement among Southern Italian universities, SHARE has built a Linked Open Data infrastructure based on open and interoperable ontologies (BIBFRAME and SHARE-VDE), enabling the representation of bibliographic resources according to an entity-relationship model inspired by FRBR and LRM, implemented in the BIBFRAME format within a Linked Open Data environment. The paper outlines the platform's architecture, collaborative practices, integration with Wikidata, and systemic implications for the evolution of the National Library Service. It positions SHARE as both a conceptual and technical laboratory for a sustainable and replicable model of semantic cataloguing.

KEYWORDS
Semantic Cataloguing; Linked Open Data; Library Cooperation; Interoperability.

SHARE Catalogue: cooperazione e innovazione semantica in un ecosistema bibliotecario condiviso

ABSTRACT
L'articolo presenta l'esperienza del progetto SHARE Catalogue come modello avanzato di cooperazione bibliotecaria e di innovazione semantica. Nato nell'ambito di una convenzione tra atenei del Mezzogiorno, SHARE ha sviluppato un'infrastruttura basata su Linked Open Data, fondata su ontologie aperte e interoperabili (BIBFRAME e SHARE-VDE), che consente di rappresentare le risorse bibliografiche secondo un modello entità-relazioni ispirato a FRBR e LRM, implementato nel formato BIBFRAME, in ambiente Linked Open Data. Il contributo illustra l'architettura della piattaforma, le pratiche collaborative, le modalità di integrazione con Wikidata e le implicazioni sistemiche per l'evoluzione del Servizio Bibliotecario Nazionale. L'articolo propone SHARE come laboratorio concettuale e tecnico per un modello sostenibile e replicabile di catalogazione semantica.

PAROLE CHIAVE
Catalogazione semantica; Linked Open Data; Cooperazione bibliotecaria; Interoperabilità.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-665

1. Introduzione. Dalla cooperazione all'interoperabilità

Nel 1984, con la firma del protocollo d'intesa tra il Ministero per i Beni Culturali e l'Università di Firenze (ICCU 1984), prendeva ufficialmente avvio il Servizio Bibliotecario Nazionale (SBN), un progetto ambizioso volto a costruire una rete cooperativa tra le principali biblioteche italiane attraverso una piattaforma condivisa di catalogazione partecipata. Il progetto era stato delineato inizialmente da Michel Boisset e Angela Vinay (Boisset e Vinay 1981), e successivamente presentato da quest'ultima al 30° Congresso nazionale dell'Associazione italiana biblioteche, tenutosi a Giardini Naxos nel novembre 1982 (AIB 1986). Coordinato fin dalle origini dall'Istituto Centrale per il Catalogo Unico (ICCU), SBN si proponeva di superare la frammentazione dei cataloghi e favorire l'accesso integrato all'informazione bibliografica e documentaria a livello nazionale. Già nella sua fase iniziale, il progetto puntava alla creazione di un'infrastruttura tecnologica comune, basata su standard aperti e interoperabili, capace di supportare servizi avanzati come il prestito interbibliotecario e la localizzazione dei documenti.

Nel novembre del 1992, a Rimini, l'Associazione italiana biblioteche (AIB 1993) tornava a interrogarsi sul significato e sulle prospettive della cooperazione interbibliotecaria, dieci anni dopo il Congresso di Giardini Naxos che aveva posto le premesse ideali e progettuali di SBN. Gli atti di quel XXXVIII Congresso dell'AIB, seppure segnati da una certa disomogeneità dei contributi, riflettono un clima di intensa partecipazione, in cui si affacciavano nuove pratiche e nuovi strumenti, ma anche il desiderio di superare le illusioni iniziali e di costruire su basi più solide un modello di cooperazione più maturo, più concreto, più visibile ed utile per gli utenti.

Il Servizio Bibliotecario Nazionale è oggi una rete di biblioteche di diversa tipologia (statali, universitarie, di enti locali, accademie e fondazioni) che condividono un'unica base di dati e un'infrastruttura comune per la catalogazione partecipata, la localizzazione dei documenti e l'accesso ai servizi. L'OPAC SBN consente la ricerca simultanea nei cataloghi delle biblioteche aderenti e offre servizi integrati come il prestito interbibliotecario (ILL SBN), basato su protocolli standardizzati che garantiscono l'interoperabilità tra software e piattaforme diverse. Tuttavia, come segnalato in recenti contributi sul rilancio di SBN, nonostante un iniziale impulso e l'integrazione di nuove risorse e authority file, persistono criticità legate alla frammentazione degli applicativi, alla lentezza nell'adozione dei Linked Data e alla difficoltà nel superamento di approcci tradizionali alla catalogazione, che rendono meno efficace la transizione verso modelli aperti, flessibili e cooperativi (Leombroni 2016).

In questo contesto, sono emerse nel tempo alcune iniziative sperimentali, avviate sotto il coordinamento dell'ICCU e in collaborazione con centri di ricerca come il VAST-LAB dell'Università di Firenze, volte alla trasformazione dei dati di SBN in Linked Open Data. Già a partire dal 2015 sono stati sviluppati modelli concettuali basati su FRBRoo (Functional Requirements for Bibliographic Records – object oriented) e CIDOC-CRM (Conceptual Reference Model), accompagnati dalla generazione di prototipi RDF (Resource Description Framework) e di interfacce SPARQL (ICCU 2015a), dalla partecipazione a progetti congiunti tra biblioteche universitarie e territoriali, come dimostrano anche le sperimentazioni del Polo digitale di Napoli (ICCU 2015b) e le attività documentate nei seminari promossi da AIB e ICCU (Aste, Mataloni, e Martinelli 2015; Borgi e D'Amato 2021). Sebbene ancora frammentarie, queste esperienze hanno indicato con chiarezza la necessità di una transizione verso una maggiore interoperabilità semantica, basata sull'adozione di standard aperti e vocabolari condivisi.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-665

A distanza di oltre quarant'anni dagli inizi di SBN, nel pieno di un processo di trasformazione tecnologica ben più radicale, è utile tornare a riflettere sulle possibili forme di cooperazione in ambito bibliotecario. In questo contesto, desidero presentare un'iniziativa alla quale ho contribuito in prima persona: il progetto SHARE (Scholarly Heritage and Access to Research). Sviluppatosi a partire dal 2013 nell'ambito di una convenzione tra gli atenei della Campania e della Basilicata, SHARE rappresenta un esempio rilevante di collaborazione bibliotecaria capace di coniugare visione strategica, rinnovamento tecnologico e attenzione alle "buone pratiche" (Delle Donne e Possemato 2017). Nato per iniziativa congiunta delle università di Napoli Federico II, Napoli L'Orientale, Napoli Parthenope, Salerno, Sannio, Basilicata, Campania "Luigi Vanvitelli", e successivamente esteso anche agli atenei di Napoli "Suor Orsola Benincasa", del Salento, di Cassino e alla Scuola Superiore Meridionale, SHARE ha puntato fin dall'inizio non solo alla razionalizzazione dei servizi, ma anche alla definizione di un vocabolario e di infrastrutture condivise, in grado di far dialogare patrimoni bibliografici diversi, contesti documentari eterogenei e molteplici esigenze informative.

Il progetto SHARE è stato successivamente ripreso e sviluppato nell'ambito di SHARE-VDE (Virtual Discovery Environment), un'iniziativa internazionale che coinvolge le maggiori biblioteche europee e nordamericane nella realizzazione di un ecosistema distribuito e integrato per la gestione condivisa di dati e servizi bibliotecari, promuovendo cluster tematici e soluzioni tecnologiche all'avanguardia (Hahn e Possemato 2023; Possemato 2019).[1] Questa evoluzione testimonia come SHARE rappresenti un ambito di sviluppo innovativo, capace di offrire spunti significativi per un possibile ripensamento dell'architettura di SBN e la sua eventuale transizione verso un'infrastruttura più compiutamente digitale.

Se nel 1992 la sfida raccolta dall'ICCU era quella di costruire una cooperazione funzionale centrata su uno standard catalografico condiviso, oggi tale sfida si ridefinisce su più piani: non soltanto come esigenza di compatibilità e integrazione semantica tra sistemi e dati, ma anche – e forse soprattutto – come necessità di rafforzare una cooperazione organizzativa tra soggetti istituzionali diversi, a livello territoriale e nazionale. Come ha evidenziato più volte Claudio Leombroni (2017), uno dei principali limiti storici di SBN è stato proprio quello di concentrarsi prioritariamente sugli aspetti tecnologici e applicativi, trascurando il consolidamento di una governance cooperativa stabile e realmente partecipata.

In questo contesto, SHARE può offrire un esempio concreto di evoluzione infrastrutturale e concettuale. Come osserva Nurmikko-Fuller (2024), le biblioteche digitali che adottano Linked Open Data non si limitano a essere meri depositi informativi, ma si configurano come nodi semantici attivi, capaci di facilitare connessioni dinamiche tra risorse, domini e contesti disciplinari eterogenei. In tale prospettiva, SHARE non è solo un catalogo collettivo in Linked Open Data, ma un ambiente cooperativo di sperimentazione metodologica, organizzativa e concettuale. La sua esperienza potrebbe fornire spunti di riflessione per un futuro ripensamento di SBN come infrastruttura pubblica della conoscenza, basata non solo su linguaggi comuni e interoperabilità semantica, ma anche su forme rinnovate di cooperazione istituzionale e civile.

[1] Si veda anche: https://www.share-vde.org/sharevde/clusters?l=en.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-665

2. Origini, obiettivi e governance del progetto SHARE

Il progetto SHARE nasce nell'ambito della programmazione triennale 2013-2015 del Ministero dell'Università e della Ricerca, con l'intento di rafforzare la cooperazione tra gli atenei del Mezzogiorno e di promuovere un sistema bibliotecario integrato, efficiente, sostenibile e aperto alla sperimentazione. A tale scopo, nel 2015 viene sottoscritta una Convenzione Interuniversitaria che definisce finalità, principi e forme organizzative del progetto, accompagnata da una Carta dei Servizi che stabilisce il riconoscimento reciproco degli utenti istituzionali, la condivisione e le forme di cooperazione tra le biblioteche aderenti.[2]

Fin dall'inizio, l'iniziativa SHARE si è articolata attorno a tre assi strategici: l'adozione di un modello cooperativo per gli acquisti e i servizi al pubblico; l'accesso integrato alle collezioni analogiche e digitali tramite piattaforme tecnologiche condivise; la promozione dell'editoria accademica in accesso aperto. A questi tre assi corrispondono rispettivamente: SHARE Services, il gruppo di coordinamento dei servizi di consultazione, prestito e *document delivery*; SHARE Catalogue e SHARE Discovery, che costituiscono le piattaforme integrate per la ricerca e il recupero delle risorse bibliografiche; SHARE Press, l'infrastruttura editoriale per la pubblicazione di riviste, libri e dati della ricerca.

Dal punto di vista istituzionale, la *governance* di SHARE si basa su un approccio paritario e partecipativo. Ogni ateneo aderente concorre con pari dignità alle decisioni strategiche e contribuisce alla definizione delle linee di sviluppo. Le attività progettuali sono coordinate da gruppi di lavoro tematici, che garantiscono il coinvolgimento attivo di bibliotecari, tecnici informatici e docenti, valorizzando competenze diffuse e promuovendo una cultura professionale condivisa. Questo orientamento ha favorito la costruzione di un'identità comune e la stabilizzazione di pratiche collaborative efficaci e replicabili, che rappresentano oggi uno dei maggiori punti di forza del progetto. Come hanno chiarito Guerrini e Possemato (2015), la transizione verso i Linked Open Data non è soltanto una questione tecnologica, ma coinvolge direttamente le finalità delle biblioteche, sia di quelle pubbliche sia di quelle universitarie, chiamate a rispondere anche alle istanze della "terza missione". In questo scenario, le biblioteche sono sempre più investite del compito di agire come mediatrici semantiche tra patrimoni informativi, soggetti descritti e comunità di utenti. Il progetto SHARE si è sviluppato proprio in questa direzione, assumendo la dimensione semantica come principio guida della cooperazione bibliotecaria.

3. Architettura di SHARE Catalogue e innovazione semantica

Il cuore tecnologico del progetto è rappresentato da SHARE Catalogue, un catalogo bibliografico integrato progettato in Linked Open Data e sviluppato insieme alla società di consulenza, progettazione e sviluppo di soluzioni tecnologiche @Cult. L'infrastruttura adotta lo standard BIBFRAME (Bibliographic Framework Initiative), promosso dalla Library of Congress come standard concepito per superare il formato MARC 21 (Library of Congress 2024; McCallum 2017). Questo impianto consente di superare i limiti dei tradizionali formati di catalogazione, centrati su record statici (*flat records*), e di adottare uno schema entità-relazioni per descrivere dinamicamente le risorse e i loro contesti d'uso.

[2] https://www.sharecampus.unina.it/.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-665

Nel framework SHARE, ogni risorsa è rappresentata come un grafo semantico interconnesso, in cui le entità bibliografiche (opere, espressioni, manifestazioni, item) sono identificate mediante URI persistenti e descritte tramite attributi e relazioni esplicite, secondo i principi del web semantico. Ciò permette una maggiore granularità descrittiva, un'integrazione più efficace con altri dataset culturali e scientifici e con fonti autorevoli (come VIAF, ISNI, Wikidata, BNF, LC authorities etc.), una piena interoperabilità con le ontologie e i vocabolari controllati già in uso nel dominio bibliografico, come IFLA-LRM (Library Reference Model) e RDA (Resource Description and Access).

SHARE Catalogue oggi adotta un'ontologia specifica, denominata SHARE-VDE Ontology, concepita come estensione del framework BIBFRAME per assicurare una migliore interoperabilità tra modelli bibliografici diversi, in particolare RDA e IFLA LRM. Come spiegano dettagliatamente Hahn e Possemato (2023), l'ontologia SHARE-VDE si basa su un approccio innovativo di modellazione attraverso insiemi di attributi (*attribute sets*), che descrivono entità bibliografiche come Opus, Work e OpusType. Questa estensione ontologica permette una gestione più raffinata delle relazioni tra opere, consentendo, ad esempio, la connessione e l'organizzazione automatica di diverse versioni e traduzioni di uno stesso contenuto bibliografico. La scelta di queste soluzioni semantiche avanzate rende SHARE non soltanto un esempio di applicazione pratica di BIBFRAME, ma anche un laboratorio metodologico capace di contribuire concretamente allo sviluppo e al perfezionamento degli standard internazionali.

Un ulteriore elemento distintivo è la scelta di strumenti per lo più open source, affiancata dall'adozione di ambienti di sviluppo condivisi basati su GitLab e su workflow distribuiti. La piattaforma consente non solo l'interrogazione semantica avanzata dei dati tramite endpoint RDF, ma anche la loro esportazione selettiva secondo diversi profili di metadati (Dublin Core, MODS, EDM), favorendone il riuso in contesti digitali eterogenei (Forziati et al. 2024).

Nel 2024, l'architettura della piattaforma LOD di SHARE è stata profondamente riorganizzata per migliorarne la flessibilità, la sostenibilità tecnica e la facilità di gestione. L'intervento ha riguardato la suddivisione modulare dei componenti, l'utilizzo di ambienti virtualizzati per semplificare l'installazione e l'aggiornamento dei servizi, e l'adozione di strumenti per il controllo delle versioni e la validazione automatica della coerenza semantica. La nuova infrastruttura consente interrogazioni complesse attraverso punti di accesso condivisi e favorisce una gestione trasparente e di qualità dei dati. La nuova piattaforma LOD rinnova anche l'interfaccia, progettata a partire da un'analisi dettagliata condotta su diverse tipologie di utenti, sia professionisti dell'informazione (come bibliotecari e sviluppatori), sia fruitori finali (come studenti, ricercatori e lettori), per facilitare l'accesso e l'esplorazione dei dati pubblicati.

L'infrastruttura semantica di SHARE-VDE è organizzata in componenti modulari e interoperabili che riflettono una precisa suddivisione funzionale: la Cluster Knowledge Base (CKB) funge da base semantica per la generazione e gestione dei cluster di entità; il Discovery Portal costituisce l'interfaccia di consultazione e navigazione del grafo bibliografico; lo User Empowerment System (UES) consente agli utenti, in particolare bibliotecari e catalogatori, di contribuire attivamente al miglioramento e alla validazione dei dati; l'Entity Management System (EMS) supporta la gestione condivisa delle entità bibliografiche secondo un approccio distribuito e collaborativo, attraverso una avanzata interfaccia utente (JCricket) che consente alla comunità di catalogatori di migliorare la qualità delle entità prodotte automaticamente attraverso i processi di "entity resolution" e

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-665

conversione, in una modalità collaborativa. Questa architettura consente una gestione flessibile e scalabile dell'intero ciclo di vita dei dati, valorizzando la partecipazione degli attori istituzionali e favorendo la co-evoluzione tra l'infrastruttura e la "comunità di pratica" che la sostiene.[3]
Attraverso questa infrastruttura, SHARE Catalogue non è semplicemente una vetrina delle collezioni bibliografiche degli enti partecipanti, ma uno strumento attivo di connessione, arricchimento e disseminazione dei saperi, in grado di dialogare con la rete semantica globale e di contribuire alla costruzione di un ecosistema documentario realmente condiviso.

4. Riuso, interoperabilità e sostenibilità: ricadute sistemiche del modello SHARE

La strutturazione dei dati in BIBFRAME, l'adozione di ontologie condivise e l'esposizione semantica secondo i principi FAIR (Findable, Accessible, Interoperable, Reusable) permettono a SHARE Catalogue di essere non solo un punto di accesso integrato, ma anche un nodo attivo di scambio semantico tra ambienti diversi, facilitando la connessione tra differenti basi di dati bibliografiche e promuovendo il riuso avanzato delle informazioni. I metadati bibliografici prodotti e pubblicati in RDF possono infatti essere facilmente riutilizzati da altri sistemi informativi culturali e accademici, integrati in piattaforme di *digital curation*, oppure interrogati da strumenti automatici per la generazione di *knowledge graph* e servizi di *discovery* avanzati.
La scelta di standard internazionali e di software in larga parte open source favorisce la sostenibilità economica e tecnica del sistema. Le trasformazioni dei dati da MARC 21 a BIBFRAME avvengono tramite pipeline personalizzabili e documentate, mentre l'interfaccia SPARQL consente la costruzione di servizi dinamici basati su query intelligenti. Questo impianto rende possibile lo sviluppo di applicazioni per l'arricchimento semantico automatico, visualizzazioni interattive, aggregazioni tematiche e collegamenti a file di autorità esterni. Inoltre, la piattaforma LOD mette a disposizione un layer di API che consente l'interazione e lo scambio di dati e servizi anche con soggetti terzi, contribuendo alla costruzione di un ecosistema bibliografico fortemente collaborativo e interoperabile.
L'evoluzione della mappatura semantica ha avuto un impatto decisivo sui meccanismi di raggruppamento automatico delle entità bibliografiche. Come sottolineano Forziati et al. (2024), la possibilità di convertire direttamente i dati in formato UNIMARC – adottato dalla maggior parte degli atenei partecipanti a SHARE – nel modello BIBFRAME, senza passaggi intermedi in MARC 21, consente una rappresentazione più coerente e dettagliata delle entità. Ciò migliora l'affidabilità dei cluster generati e potenzia l'interoperabilità semantica tra risorse eterogenee. Questo approccio si è rivelato essenziale per facilitare la riconciliazione automatica e il collegamento tra entità, promuovendo un'aggregazione intelligente delle informazioni.
Inoltre, il sistema è progettato per supportare la gestione delle diverse versioni dei dati e la tracciabilità puntuale di tutte le modifiche apportate, elementi essenziali per garantire trasparenza e affidabilità nei processi di cooperazione. Grazie alla capacità di dialogo tra sistemi, le entità bibliografiche possono essere "riconciliate", ossia identificate e collegate a entità corrispondenti

[3] https://wiki.share-vde.org/wiki/Main_Page.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-665

presenti in archivi esterni (ad esempio, con gli autori su VIAF, Wikidata o Idref), favorendo una progressiva convergenza verso un ecosistema informativo distribuito, aperto e dialogico.
Queste caratteristiche rendono il progetto SHARE una risorsa strategica per tutti quei contesti bibliotecari che intendano adottare un'infrastruttura semantica di nuova generazione, capace di valorizzare il patrimonio documentario esistente e di integrarlo nei circuiti internazionali della ricerca e della comunicazione scientifica.
Come sottolineato sia da McCallum (2017) sia da Hahn e Possemato (2023), una delle maggiori potenzialità offerte dalla transizione semantica verso BIBFRAME e SHARE-VDE risiede proprio nella capacità di creare un ecosistema informativo aperto e flessibile, sostenibile tecnicamente ed economicamente nel lungo periodo. L'adozione di strumenti open source e standard internazionali come RDF, insieme alla possibilità di integrare e riutilizzare ontologie esterne e autorità semantiche globali (come Wikidata), garantisce una maggiore durabilità dei dati, nonché la possibilità di aggiornarli dinamicamente. SHARE, pertanto, non si limita a rispondere a bisogni contingenti: ambisce a essere un modello replicabile di interoperabilità avanzata, orientato al confronto e all'interazione con l'evoluzione continua degli standard semantici internazionali.
SHARE può costituire, per molti aspetti, un'alternativa avanzata all'OPAC per la navigazione e la rappresentazione semantica delle risorse, ma non è progettato per sostituire un *discovery tool* generalista. La piattaforma privilegia l'esposizione in Linked Open Data di risorse bibliografiche interoperabili, mentre l'accesso integrato a fonti elettroniche commerciali e a contenuti full-text resta affidato a strumenti specifici, nel rispetto delle limitazioni imposte dai diritti d'uso.

5. Interconnessioni semantiche e dati collaborativi: il ruolo di Wikidata

Uno degli sviluppi più rilevanti del progetto SHARE è rappresentato dall'integrazione con Wikidata, la *knowledge base* collaborativa mantenuta dalla Wikimedia Foundation, che funge da snodo per il web della conoscenza aperta (Forziati e Lo Castro 2018).
In SHARE-VDE, Wikidata non è soltanto una fonte di dati esterni da cui importare identificativi o proprietà: essa partecipa attivamente alla costruzione delle identità bibliografiche, concorrendo alla definizione e all'arricchimento dei cluster semantici. Attraverso meccanismi di "entity resolution" e collegamenti bidirezionali, le entità descritte nel grafo RDF di SHARE possono essere ricondotte a corrispondenti entità di Wikidata, e viceversa, contribuendo alla generazione di identificatori condivisi e all'alimentazione dinamica della Cluster Knowledge Base di SHARE. Questo modello ibrido – fondato sulla coesistenza tra dati curati istituzionalmente e fonti collaborative – permette di superare la dicotomia tra authority file chiusi e vocabolari aperti, favorendo un'integrazione semantica realmente distribuita.[4]
Il progetto di allineamento semantico tra SHARE Catalogue e Wikidata, che ha riguardato nella prima fase operativa oltre duecentomila entità, ha l'obiettivo di stabilire corrispondenze puntuali tra le entità bibliografiche descritte in BIBFRAME e le entità di Wikidata, a partire da differenti tipologie di agenti (persone e organizzazioni) e il proposito di allargare la stessa prassi a soggetti e opere. L'utilizzo in Wikidata della proprietà P3987 (SHARE Catalogue author ID), che mette in

[4] https://wiki.share-vde.org/wiki/Main_Page.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-665

relazione gli item di Wikidata con gli identificatori dei cluster degli autori di SHARE Catalogue, consente di rendere espliciti tali collegamenti e di favorire una navigazione fluida tra ambienti informativi diversi e interoperabili.
Questo approccio ha permesso l'attivazione di endpoint SPARQL federati, in grado di interrogare simultaneamente sia il grafo RDF di SHARE sia quello di Wikidata, favorendo ricerche integrate, funzionalità avanzate e iniziative collaborative. La comunità dei bibliotecari coinvolti ha contribuito alla creazione di nuovi item su Wikidata, al miglioramento di quelli esistenti e alla documentazione dei processi, secondo una logica di condivisione e mutuo supporto tra archivi di dati istituzionali e infrastrutture collaborative a livello globale. In tal modo, SHARE ha svolto anche un ruolo formativo e di potenziamento delle competenze, contribuendo a sviluppare una comunità di pratica capace di operare efficacemente nel web semantico.

6. L'evoluzione semantica della catalogazione: da MARC all'Entity Modelling

L'adozione del framework BIBFRAME da parte di SHARE non è quindi un mero aggiornamento formale, ma rappresenta un passo importante verso l'adozione dei paradigmi del web semantico. Tale evoluzione si fonda sui modelli concettuali introdotti da FRBR e successivamente riformulati in LRM, che ridefiniscono l'identità bibliografica non più come una descrizione monolitica racchiusa in un record, ma come un insieme di entità correlate (autori, opere, espressioni, manifestazioni, contesti) dinamicamente rappresentabili e collegabili in un grafo semantico. In questo scenario, formati come BIBFRAME risultano molto più compatibili con tali modelli rispetto al tradizionale MARC, offrendo una struttura più flessibile per l'interoperabilità e l'arricchimento dei dati (Possemato 2024).
Questa visione si collega a quanto affermato da Guerrini e Possemato (2015) già dieci anni fa quando osservavano che il superamento del record MARC a favore di una modellazione fondata su entità e relazioni rappresenta una svolta culturale, prima ancora che tecnica, nella rappresentazione della conoscenza. Va tuttavia ricordato che tale modellazione, nata originariamente nell'ambito dei database relazionali e adottata nei modelli concettuali bibliografici come FRBR, non implica necessariamente l'adozione del web semantico, ma può essere potenziata da esso. La struttura reticolare dei Linked Open Data favorisce infatti l'esplicitazione e l'interconnessione dei nessi semantici, valorizzando le traiettorie interpretative che legano tra loro le risorse bibliografiche.
Questo cambiamento, come abbiamo visto nei paragrafi precedenti, comporta una trasformazione epistemologica: si abbandona la logica documentale, basata su record statici, per adottare un modello fondato su entità distinte – come persone, opere, soggetti – e sulle relazioni che le connettono nel tempo e nei contesti.
Come evidenziato da Forziati et al. (2024), i più recenti sviluppi del progetto SHARE Catalogue hanno reso possibile la conversione delle risorse fisiche direttamente dal formato UNIMARC al modello BIBFRAME, senza necessità di una trasformazione intermedia in MARC 21. Questa modalità semplifica l'intero processo di transizione semantica, riducendo la complessità, migliorando la coerenza dei dati e incrementando l'affidabilità dei cluster generati in fase di normalizzazione automatica.
Il processo di clusterizzazione nei progetti della SHARE Family si basa su un sistema di identificazione e raggruppamento automatico delle entità bibliografiche e di authority che mira a superare

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-665

la frammentarietà dei record provenienti da fonti eterogenee. Attraverso l'utilizzo di algoritmi di normalizzazione e "matching" semantico, le entità simili vengono riunite in cluster logici, ciascuno dotato di un identificatore persistente, con la possibilità di gestire le varianti, le ambiguità e i conflitti di attribuzione. Questi cluster alimentano una Knowledge Base centralizzata, che costituisce la base per le funzionalità di *discovery* offerte dal sistema, finalizzate alla navigazione strutturata e alla visualizzazione delle entità bibliografiche e di authority. Il *clustering* non è un processo unidirezionale e chiuso, ma può essere sottoposto a verifiche, revisioni e correzioni da parte dei catalogatori, grazie agli strumenti messi a disposizione dal sistema UES (User Empowerment System) e, nello specifico, dal linked data editor JCricket.[5]
L'esperienza di SHARE ha contribuito ad arricchire questo dibattito, anche grazie alla proposta di creazione di estensioni di BIBFRAME specificamente concepite per il contesto MAB (Musei, Archivi, Biblioteche), in cui l'integrazione tra diverse tipologie documentarie richiede modelli flessibili, adattabili e semanticamente coerenti.

7. Un modello da replicare? Criticità e potenzialità per SBN

L'esperienza di SHARE rappresenta un caso di studio significativo all'interno di una riflessione più ampia sul futuro di SBN. Se da un lato SBN costituisce un'infrastruttura consolidata e capillarmente diffusa, dall'altro presenta limiti strutturali che ne frenano l'evoluzione verso modelli più aperti, modulari e replicabili a livello internazionale, con una limitata circolazione delle soluzioni adottate al di fuori del contesto nazionale. In questo scenario, SHARE dimostra come sia possibile realizzare un'infrastruttura distribuita, cooperativa e orientata allo scambio strutturato di dati, in grado di dialogare efficacemente con piattaforme europee (come Europeana e OpenAIRE) e globali (come Wikidata), pur operando in un contesto decentralizzato e con software locali differenziati.
Va nondimeno ricordato che negli ultimi anni, anche all'interno di SBN, sono emersi importanti segnali di rinnovamento. Un esempio è il progetto di riconciliazione tra le voci di autorità dell'OPAC SBN e le entità di Wikidata, avviato grazie al contributo volontario del Gruppo Wikidata per Musei, Archivi e Biblioteche (GWMAB). La proprietà P396 di Wikidata, relativa agli autori persona di SBN, è stata creata nel 2013,[6] ma soltanto di recente, con l'ampliamento della visibilità dell'authority file di SBN e con lo sviluppo di strumenti di riconciliazione automatica, si è raggiunto un traguardo rilevante: oltre 200.000 voci collegate, nonché la creazione di plug-in per browser e strumenti personalizzati in Javascript a supporto dell'attività editoriale su Wikidata.[7]
Un ulteriore passo in avanti è rappresentato da Alphabetica, la nuova piattaforma del Servizio Bibliotecario Nazionale, che non si limita a offrire un'interfaccia di accesso modernizzata, ma integra in un unico ambiente di ricerca e navigazione i contenuti dell'OPAC SBN e di altri database gestiti dall'ICCU, come Internet Culturale. Attraverso un approccio visuale e user-friendly, Alphabetica migliora significativamente l'esperienza d'uso e valorizza il patrimonio informativo che SBN ha costruito nel tempo (Atturo e Ravelli 2024; Castro 2022).

[5] https://wiki.share-vde.org/wiki/Main_Page.
[6] https://www.wikidata.org/wiki/Property:P396.
[7] https://www.wikidata.org/wiki/Wikidata:Gruppo_Wikidata_per_Musei,_Archivi_e_Biblioteche/SBN.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-665

Alla luce di queste evoluzioni, non si tratta di sostituire SBN, ma di integrarne l'architettura attraverso modelli interoperabili e complementari, capaci di dialogare tra loro e contribuire congiuntamente all'evoluzione dell'ecosistema bibliografico nazionale. La coesistenza di soluzioni differenti, purché capaci di dialogare tra loro, può infatti rafforzare l'intero ecosistema informativo nazionale. In questo contesto, le pratiche sviluppate da SHARE offrono percorsi concreti di transizione verso la modellazione entità-relazioni, l'esposizione in Linked Open Data e l'adozione di workflow collaborativi, senza rinunciare all'eredità catalografica esistente.
Questa traiettoria si inserisce in un più ampio processo di evoluzione semantica che investe anche il contesto italiano, sempre più consapevole della necessità di superare la logica monolitica del record MARC in favore di modelli più flessibili e interoperabili. In questo quadro, SHARE non è un'eccezione, ma un laboratorio avanzato di sperimentazione in sintonia con i percorsi di aggiornamento intrapresi da molte istituzioni bibliografiche e culturali, che vedono nel web semantico un'opportunità strategica per rafforzare le connessioni tra dati, patrimoni e saperi.

8. Intelligenza artificiale e modellazione semantica nei sistemi bibliografici

Negli ultimi anni l'integrazione tra Linked Open Data e intelligenza artificiale ha aperto prospettive inedite per l'arricchimento automatico dei metadati, il riconoscimento di entità, l'analisi delle relazioni semantiche e la disambiguazione dei contesti informativi (Mixter 2024). Sistemi di *natural language processing*, reti neurali e modelli di "apprendimento profondo" possono contribuire a migliorare la qualità e la coerenza dei dati bibliografici, specialmente nei contesti eterogenei e multilinguistici. Alla luce di queste potenzialità, SHARE Catalogue potrebbe evolvere verso una piattaforma intelligente capace non solo di aggregare dati, ma anche di suggerire inferenze, allineamenti e connessioni tra risorse, valorizzando appieno la dimensione reticolare della conoscenza.
L'adozione di ontologie trasparenti, tracciabili e condivise – come quelle sviluppate in SHARE-VDE – costituisce una condizione fondamentale per uno sviluppo eticamente sostenibile dell'intelligenza artificiale applicata alla bibliografia. La presenza di dati strutturati secondo standard aperti e interoperabili consente infatti di addestrare modelli predittivi su basi verificabili e riproducibili, riducendo i rischi di opacità algoritmica. In questo senso, SHARE non è solo un'infrastruttura tecnica, ma un laboratorio epistemologico in cui sperimentare forme di intelligenza computazionale orientate alla qualità, alla verifica e all'inclusività dei saperi.
Nel contesto dei Linked Open Data, la funzione della biblioteca si espande oltre la mera intermediazione documentaria. Essa si configura come nodo di intelligenza collettiva, in cui competenze umane e capacità computazionali si intrecciano nella costruzione, nella rilettura e nella connessione delle conoscenze. In questa prospettiva, l'intelligenza artificiale non sostituisce l'intermediazione bibliotecaria, ma la potenzia e la rende scalabile, aprendo alla possibilità di esplorazioni personalizzate, contestualizzate e culturalmente sensibili.

9. Conclusioni. Dalla cooperazione alla rete: una sfida condivisa

Le sinergie tra biblioteche, oggi, faticano ancora a tradursi in una cooperazione strutturata sul piano della condivisione delle risorse e della razionalizzazione dei servizi. Ma proprio questa fragilità

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-665

dimostra quanto la sfida della cooperazione sia oggi, più che mai, anche intellettuale e politica: essa richiede la capacità di ideare modelli inclusivi, sostenibili e orientati alla diffusione del sapere. L'esperienza di SHARE dimostra che è possibile progettare sistemi bibliografici in grado di dialogare con l'esterno, mantenendo al tempo stesso coerenza interna, cura nella qualità dei dati e disponibilità al cambiamento.

Le prospettive di sviluppo delle biblioteche internazionali che aderiscono a SHARE comprendono l'ampliamento a nuovi ambiti disciplinari (con i progetti SHARE-Art e SHARE-Music), la nascita di comunità di pratica transdisciplinari e la partecipazione attiva alle principali reti della scienza aperta. In tale direzione, SHARE si configura non solo come iniziativa tecnologica, ma come infrastruttura pubblica cooperativa, promotrice di una rinnovata cultura della collaborazione bibliotecaria, ispirata a principi di interoperabilità, trasparenza e apertura.

Questo approccio realizza pienamente una visione volta a trasformare le biblioteche in *snodi attivi del web semantico*, capaci di contribuire alla co-costruzione di una conoscenza pubblica, condivisa e interrogabile attraverso modelli aperti. In questa prospettiva, la cooperazione bibliotecaria rafforza e rinnova la propria azione civile e culturale, estendendola anche all'ambiente digitale, grazie a linguaggi comuni, ontologie aperte e responsabilità condivise.

Questa infrastruttura prende forma attraverso una "comunità di pratica" internazionale che include bibliotecari, informatici e specialisti di metadati, attivamente coinvolti nello sviluppo collaborativo dell'ontologia SHARE-VDE e nei processi decisionali della cosiddetta Share Family. Come documentato nel portale tecnico dell'iniziativa Share Family, la piattaforma non è un ambiente neutro, ma un ecosistema distribuito modellato da soggetti reali che negoziano, revisionano e validano congiuntamente le scelte semantiche, gli attributi descrittivi e le pratiche di *clustering*. Ne deriva un modello in cui l'interoperabilità non è solo un'esigenza tecnica, ma una responsabilità condivisa: ogni nodo bibliotecario contribuisce alla costruzione di un vocabolario comune e alla curatela partecipativa dell'identità bibliografica.

L'adozione di modelli entità-relazioni e di ontologie comuni – come quelle sviluppate in SHARE-VDE – implica una ridefinizione delle relazioni tra oggetti, soggetti e contesti. Come rileva Nurmikko-Fuller (2024), le ontologie non sono mai neutrali: riflettono visioni del mondo, priorità interpretative e modalità di rappresentazione del sapere. In questa prospettiva, l'intelligenza artificiale potrà costituire uno strumento formidabile per potenziare l'intermediazione bibliotecaria, purché si fondi su basi semantiche trasparenti, verificabili e culturalmente situate.

L'adozione di standard semantici nel contesto del web aperto non costituisce solo un'evoluzione tecnica, ma un cambiamento epistemologico e partecipativo, in cui le biblioteche diventano nodi attivi di una conoscenza condivisa, dinamica e radicata nei contesti culturali. SHARE interpreta pienamente questa visione, proponendo un modello replicabile di alleanza tra competenze umane e tecnologie semantiche, al servizio di una circolazione della conoscenza più equa e condivisa.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-665

Riferimenti bibliografici*

AIB (Associazione italiana biblioteche). 1986. *La cooperazione: il Servizio bibliotecario nazionale, Giardini-Naxos, 21-24 novembre 1982.* Messina: AIB.

AIB (Associazione italiana biblioteche). 1993. *Biblioteche insieme: gli spazi della cooperazione. Atti del XXXVIII Congresso nazionale dell'Associazione italiana biblioteche, Rimini, 18-20 novembre 1992.* A cura di Paolo Malpezzi. Roma: AIB.

Aste, Margherita, Maria Cristina Mataloni, e Luca Martinelli. 2015. "Linked Data: il mondo di Internet e il ruolo delle biblioteche, degli archivi e dei musei." *Digitalia* 10 (1/2): 65–73. https://digitalia.cultura.gov.it/article/view/1474.

Atturo, Valentina e Elena Ravelli. 2024. "L'evoluzione dell'authority work in SBN. Dalle origini ad Alphabetica e prospettive future." *JLIS.it* 15 (1):45-58. https://doi.org/10.36253/jlis.it-572.

Boisset, Michel, e Angela Vinay. 1981. "Università Europea e Servizio bibliotecario Nazionale." *Il Ponte* 37 (5): 394–97.

Borgi, Elena, e Lianna D'Amato. 2021. *Il progetto Linked Open Data del CoBiS. Interoperabilità e arricchimento semantico dei dati bibliografici.* https://cobis.to.it/wp-content/uploads/2021/05/Cobis-Lod-Unimi-2021.pdf.

Castro, Elisabetta. 2022. "Alphabetica: uno strumento poliedrico di accesso ai servizi bibliografici nazionali." *DigItalia* 17 (1): 39–47.

Delle Donne, Roberto, e Tiziana Possemato. 2017. "SHARE Catalogue: un'esperienza di cooperazione." *Biblioteche Oggi* 35: 21–9.

Forziati, Claudio, Annalisa Di Sabato, Rossella Molisso, e Chiara Mugnano. 2024. "La nuova LOD Platform di SHARE Catalogue: un'evoluzione nel segno delle pratiche collaborative della Share Family." In *Parsifal. Un modello di collaborazione bibliotecaria per condividere la conoscenza registrata*, a cura di Silvano Danieli, 105–25. Firenze: Firenze University Press.

Forziati, Claudio, e Valeria Lo Castro. 2018. "La connessione tra i dati delle biblioteche e il coinvolgimento della comunità: il progetto SHARE Catalogue-Wikidata." *JLIS.it* 9 (3): 109–20. https://doi.org/10.4403/jlis.it-12488.

Guerrini, Mauro, e Tiziana Possemato. 2015. *Linked Data per biblioteche, archivi e musei. Perché l'informazione sia del web e non solo nel web.* Milano: Editrice Bibliografica.

Hahn, Jim, e Tiziana Possemato. 2023. *The Share-VDE ontology: a BIBFRAME extension for linked data discovery.* CERN European Organization for Nuclear Research. https://zenodo.org/records/8414627.

ICCU (Istituto Centrale per il Catalogo Unico). 1984. *Protocollo di intesa fra Ministero per i beni e le attività culturali e le Regioni per il progetto speciale di Servizio Bibliotecario Nazionale.* https://www.iccu.sbn.it/it/accordi-convenzioni/protocolli-dintesa/protocollo-di-intesa-fra-ministero-per-i-beni-e-le-attivita-culturali-e-le-regioni-per-il-progetto-speciale-di-servizio-bibliotecario-nazionale/index.html.

* Data di ultima consultazione dei siti web: 11 giugno 2025

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-665

ICCU. 2015a. *Gruppo di lavoro LOD SBN*. https://www.iccu.sbn.it/it/attivita-servizi/gruppi-di-lavoro-e-commissioni/pagina_0005.html.

ICCU. 2015b. *Progetto ICCU – Polo digitale Napoli*. https://www.iccu.sbn.it/it/attivita-servizi/attivita-nazionali/pagina_0007.html.

Leombroni, Claudio. 2017. "Tra *krisis* e *kairòs*: sul futuro di SBN." *Digitalia* 1/2 (2017): 127–35. https://www.bv.ipzs.it/bv-pdf/007/MOD-BP-17-107-017_2294_1.pdf.

Library of Congress. 2024. *BIBFRAME Initiative.* https://www.loc.gov/bibframe/.

McCallum, Sally. 2017. "BIBFRAME Development." *JLIS.it* 8 (3): 71–85. https://doi.org/10.4403/jlis.it-12415.

Mixter, Jeff. 2024. "An Introduction to Library Linked Data." *OCLC Next Blog*, 26 marzo. https://blog.oclc.org/next/intro-to-library-linked-data/?utm_source=chatgpt.com.

Nurmikko-Fuller, Terhi. 2024. *Linked Data for Digital Humanities.* London-New York: Routledge.

Possemato, Tiziana. 2019. *Share Virtual Discovery Environment in Linked Data (SHARE-VDE). State-of-the-art.* Fiesole. https://www.casalini.it/retreat/web_content/2019/presentations/possemato.pdf.

Possemato, Tiziana. 2024. *Entity Modeling: la terza generazione della catalogazione.* Firenze: Firenze University Press.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-666

JLIS.it

NaMo, utility for populating missing date fields in authority records*

Silvia Cagnizi[a], Stefano Bargioni[b]

a) Pontifical University Santa Croce in Rome, https://orcid.org/0009-0005-7312-6928
b) Pontifical University Santa Croce in Rome, https://orcid.org/0000-0001-7679-2874

Contact: Silvia Cagnizi, s.cagnizi@pusc.it; Stefano Bargioni, bargioni@pusc.it
Received: 23 June 2025; **Accepted:** 23 October 2025; **First Published:** 15 January 2026

ABSTRACT
The article introduces NaMo, a tool developed to achieve effective and integrated bibliographic control and to enrich person entities in library catalogs, with particular reference to the work carried out by the Library of the Pontifical University Santa Croce. In a context of increasing use of entities and URIs within the semantic web, NaMo leverages Wikidata to integrate and update biographical information — such as birth and death dates — through catalogers' contributions, using SPARQL queries and a dynamic interface. The tool operates on two data sets: one consisting of recently updated records in Wikidata, and another enabling retrospective work, both aimed at enriching authority entities. NaMo facilitates the identification of records with incomplete data, supporting critical review and validation processes. It proves effective not only in data enrichment but also in error correction and disambiguation, highlighting the importance of standardized query tools and collaboration across bibliographic systems. The study also explores the potential extension of NaMo to other systems such as SBN, GND, and BNF.

KEYWORDS
Identifiers; Authority Control; RDA; Koha; Wikidata; SBN; URBE.

NaMo, strumento per l'aggiunta di date mancanti a record di autorità

ABSTRACT
L'articolo presenta NaMo, uno strumento sviluppato per ottenere un controllo bibliografico efficace ed integrato e per arricchire le entità (persona) nei cataloghi bibliotecari, in particolare in riferimento al lavoro svolto dalla Biblioteca della Pontificia Università della Santa Croce. In un contesto di crescente utilizzo di entità e URI nel web semantico, NaMo sfrutta i dati di Wikidata per integrare e aggiornare mediante il lavoro dei catalogatori le informazioni biografiche, come le date di nascita e morte, attraverso un sistema di query SPARQL e un'interfaccia dinamica. Lo strumento interviene lavorando su due insiemi di dati: quelli con aggiornamenti recenti in Wikidata e quelli che permettono un lavoro retrospettivo, con l'obiettivo di arricchire le entità. NaMo permette quindi di individuare comodamente record con dati incompleti, facilitando interventi critici di revisione e validazione e rivelandosi in grado di affrontare non solo l'arricchimento di dati, ma anche la correzione di errori e la gestione di omonimie, evidenziando l'importanza di strumenti di interrogazione standardizzati e di collaborazione tra sistemi bibliografici. Lo studio affronta anche la possibilità di estendere NaMo ad altri sistemi come SBN, GND e BNF.

PAROLE CHIAVE
Identificatori; Controllo di autorità; RDA; Koha; Wikidata; SBN; URBE.

* Gli autori hanno cooperato nella stesura e nella revisione dell'articolo; tuttavia, ogni autore ha redatto parti specifiche dell'articolo. Silvia Cagnizi: Introduzione, Casi d'uso, Conclusioni; Stefano Bargioni: Metodo, Statistiche, Applicazioni della logica di NaMo in altri ambiti, Conclusioni, Appendice tecnica.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-666

Introduzione

All'interno dei cataloghi delle biblioteche contemporanee sempre più frequentemente si fa ricorso all'uso di entità, a seguito dell'incremento costante del bisogno di progredire nel controllo bibliografico universale (Bianchini e Sardo 2022). Le operazioni di catalogazione si occupano anche e soprattutto della gestione dell'authority file:[1] la produzione di voci di autorità ben strutturate permette l'identificazione delle entità (Possemato 2024) e contribuisce ad accrescere il controllo sui dati.

Come ben spiegato in (Manzoni 2022), il processo di identificazione è "l'attività caratterizzante il web semantico, all'interno del quale ogni oggetto è rappresentato mediante un URI" (Manzoni 2022, 10) che ne garantisca sia la stabilità che l'unicità: per questo motivo anche nel catalogo della Biblioteca della Pontificia Università della Santa Croce (d'ora in poi PUSC) tale questione è stata messa al centro delle operazioni di authority control. Inoltre, da circa un paio d'anni il nostro catalogo contribuisce all'Authority File Centralizzato del catalogo Parsifal,[2] che esprime attraverso uno strumento tangibile la collaborazione tra le biblioteche della Rete URBE, circa venti, i cui cataloghi locali sono redatti utilizzando il formato MARC21. La necessità di identificare in maniera inequivocabile le entità condivise dai cataloghi di tali biblioteche ha messo ancora di più al centro del lavoro gli identificatori: Parsifal è diventato un hub, composto di cluster che raggruppano tutte le diverse rappresentazioni della stessa entità; lo stesso Parsifal possiede inoltre un proprio identificatore (denominato "urbe")[3] che consente di collegare le singole voci di autorità dei cataloghi locali all'entità, cioè al cluster identificato. I linked data prodotti da Parsifal alimentano il web semantico e contribuiscono al recupero delle entità (Bianchini 2024).

Il catalogo della Biblioteca della PUSC,[4] gestito attraverso il software open source Koha, è dotato di un ricco authority file,[5] all'interno del quale gli identificatori principalmente utilizzati sono Wikidata,[6] VIAF, ISNI, URBE.[7]

A seguito dei sempre più numerosi casi rilevati dall'esperienza di catalogazione di molteplici cluster relativi alla stessa entità in VIAF, il catalogo della PUSC sta sviluppando in maniera sistematica i collegamenti con gli altri identificatori e in particolare con le entità di Wikidata;[8] esse, infatti, nonostante siano numericamente più contenute rispetto ai database internazionali VIAF e ISNI, presentano dati qualitativamente molto più affidabili (Possemato, Di Sabato, e Moi 2024, 98), specialmente per gli autori antichi.[9] Contemporaneamente ci si è impegnati nell'arricchimento collaterale dello stesso Wikidata, che permette, tra le altre cose e al contrario di VIAF, a chiunque di effettuare la fusione tra entità duplicate, garantendo quindi un migliore controllo e una maggiore attendibilità, nonché limitata dispersione dell'informazione.

---

[1] A proposito di authority control, si legga (Wiederhold e Reeve 2021).

[2] Il catalogo Parsifal è stato inaugurato l'11 maggio 2023 con l'organizzazione di un convegno, i cui atti sono pubblicati in (Danieli 2024), consultabile in open access al link https://media.fupress.com/files/pdf/24/14537/41461.

[3] https://www.loc.gov/standards/sourcelist/standard-identifier.html.

[4] https://catalogo.pusc.it/.

[5] Le registrazioni di autorità sono più di 120mila a fronte di circa 220.000 risorse.

[6] All'interno del catalogo sono circa il 70% i record ad avere l'identificatore Wikidata.

[7] https://parsifal.urbe.it/parsifal/homeSearch.

[8] Per una storia della collaborazione tra Wikidata e la Rete URBE, si veda (Pellizzari di San Girolamo 2024).

[9] Sul confronto tra VIAF e Wikidata, si veda (Bianchini, Bargioni, e Pellizzari di San Girolamo 2021).

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-666

Ogni entità presente nel nostro catalogo può essere collegata alla corrispondente entità in Wikidata attraverso una proprietà[10] che riporta l'identificatore PUSC di un agente: il collegamento tra i due database avviene in maniera reciproca,[11] segue cioè la tecnica denominata *data round-tripping*.[12]
Le voci di autorità del catalogo della Biblioteca della PUSC si presentano quindi per una gran parte dotate di identificatore Wikidata. Nel nostro catalogo, però, numerose sono anche le voci "povere" di dati rilevanti per l'identificazione: si tratta di record per i quali al momento della creazione non si avevano a disposizione ulteriori informazioni; e quelle in cui, nonostante la completezza dei dati, le informazioni necessitano di un aggiornamento.
Ma non è tutto: l'importanza di questo e degli altri identificatori si estrinseca anche nel modello di autorità "Persona" definito dalla Commissione per la revisione delle Varianti locali[13] di URBE. In questo modello, la cui esigenza nasceva già a partire dall'introduzione di RDA[14] nella Rete URBE, sono state definite alcune indicazioni utili per la compilazione delle voci di autorità (Leoni 2024, 131), con diversi livelli di completezza.
Il modello di voce di autorità minimale è stato definito come composto da una serie di campi minimi obbligatori, al fine di avere una identificazione più possibile chiara e di snellire il lavoro di quelle biblioteche con personale numericamente ridotto all'interno di URBE; tra questi campi ne emergono due fondamentali per il nostro discorso: 024 (identificatori) e 046 (date).
Dell'importanza degli identificatori abbiamo già parlato; il tag 046, che la Rete URBE ha deciso di compilare in formato EDTF (Extended date/time format),[15] riveste un'importanza cruciale non soltanto per l'identificazione ma anche per la formalizzazione delle date che spesso nel tag 100$d vengono espresse con qualificatori derivati dal linguaggio corrente: si tratta soprattutto di quegli autori, vissuti in epoche remote, a cui non possono essere associate date precise e per i quali si fornisce spesso, qualora possibile, una datazione approssimativa. Inoltre il tag 046 consente di avere date precise al giorno, rendendo l'identificazione, per la quale si considerano sufficienti nome, cognome e data di nascita/morte, più accurata.[16]
Va anche osservato che la nostra utenza ha sottolineato in diverse occasioni l'utilità del posizionamento temporale – e quindi anche culturale – degli autori in catalogo. Le informazioni spazio-temporali orientano molto nelle funzioni utente di "trovare", "identificare" ed "esplorare" incluse nel modello LRM (Guerrini e Sardo 2018, 70-72).
Lo sviluppo del tool NaMo prende le mosse da questa situazione di partenza. Il meccanismo di funzionamento è quello dell'accoppiamento delle due entità: la voce di autorità del nostro catalogo, e l'elemento corrispondente in Wikidata, collegati attraverso i rispettivi identificatori.

[10] Si tratta della proprietà P5739, usata da circa 70.000 elementi.
[11] Questa proprietà può essere alimentata attraverso l'utilizzo di Mix'n'match: https://mix-n-match.toolforge.org/#/catalog/5702. Sull'uso di Mix'n'match, si veda (Pellizzari di San Girolamo 2024, 150).
[12] Tale funzione viene utilizzata anche da altre importanti biblioteche nel mondo, come ad esempio Parsifal, la Library of Congress, la National Library of Israel, la Biblioteca Nazionale di Catalogna. https://www.wikidata.org/wiki/Wikidata:Data_round-tripping.
[13] La *Lista delle varianti locali alle regole di catalogazione AACR2/RDA ammesse nei cataloghi della Rete URBE* è stata approvata nel 2009 dall'Assemblea dei Bibliotecari di URBE ed è consultabile all'indirizzo https://www.urbe.it/formazione/userfiles/file/manoni/varianti_locali_URBE_testo_finale.pdf.
[14] https://www.rdatoolkit.org/about.
[15] https://www.loc.gov/standards/datetime/.
[16] Per questi motivi è stato scelto di utilizzare per NaMo il tag 046 invece del sottocampo d del tag 100.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-666

L'obiettivo che ci siamo prefissati al momento della creazione di questo strumento è quello della migliore identificazione e dell'arricchimento informativo,[17] attraverso l'aggiunta di date mancanti. Dimostreremo come gli automatismi di NaMo siano molto utili al metadatatore che intende arricchire il proprio authority file, ma evidenzieremo anche le limitazioni intrinseche che impediscono l'aggiornamento automatico del catalogo con le informazioni raccolte. Risultano infatti imprescindibili l'esperienza e la professionalità del catalogatore, che attraverso la capacità di analisi critica potrà risolvere le frequenti problematiche di omonimia, assegnazioni erronee, sovrapposizioni improprie e persino identificatori inaccurati.
La risoluzione di tali incongruenze richiede l'intervento del giudizio umano, essenziale per la rettifica e la validazione dei dati catalografici.

Metodo

Durante la fase di progettazione dello strumento, sono stati evidenziati due insiemi di dati ai quali applicarlo.
Il primo, che peraltro diede origine all'idea di NaMo, è relativo al lavoro corrente ed è rappresentato dagli elementi di Wikidata contenenti P5739 che vengono arricchiti costantemente nel tempo grazie al lavoro ordinario su Wikidata, portato avanti dagli utenti più svariati.[18] Questo primo insieme è in costante aggiornamento, perché non considera soltanto le modifiche alle date di nascita o morte (rispettivamente registrate nelle proprietà P569 e P570), ma viene alimentato anche dalle modifiche effettuate su qualsiasi altra proprietà dell'elemento Wikidata.[19] Richiede pertanto di essere seguito con regolarità nel tempo, sia per riportare in catalogo date di morte molto recenti, sia per registrare date di nascita o morte aggiunte recentemente da un qualunque utente di Wikidata al corrispondente elemento.
Il secondo insieme di dati, ben più vasto, si riferisce al lavoro pregresso ed è costituito dalle voci di autorità per le quali manca una data di nascita o di morte nel nostro catalogo; tali voci vengono selezionate da NaMo in Wikidata in base a una decade scelta dall'operatore, all'interno della quale cade l'anno di morte.[20] Visto il numero elevato, di cui si è discusso in precedenza, di riconciliazioni con Wikidata tramite P5739, si tratta quasi esclusivamente di autori minori o di autori di cui la nostra biblioteca ha scarso posseduto. Infatti, queste voci, negli anni in cui la metadatazione non era ancora sviluppata, non furono corredate da elementi identificativi sufficienti.
Nella realizzazione di NaMo sono state quindi introdotte due sezioni, ovvero due tipi di ricerca: recenti e nel passato. I due casi differiscono per la query di individuazione degli elementi in Wikidata, ed hanno in comune tutte le altre componenti di NaMo, vale a dire l'intersezione con le voci di autorità di Koha, il filtro in base alle date mancanti in Koha, e la visualizzazione in HTML.

[17] Sull'argomento si legga (Possemato 2023).
[18] Il Gruppo di lavoro sull'Authority Control costituisce la stragrande maggioranza di questi utenti: si vedano le statistiche all'indirizzo https://bambots.brucemyers.com/NavelGazer.php?property=P5739.
[19] Tramite https://wdrc.toolforge.org/ si potrebbe concentrare la ricerca solo sugli elementi a cui sono state modificate le date P569 o P570, ma si basa su un dataset derivato, che ha un ritardo misurabile in mesi.
[20] Avremmo potuto indifferentemente utilizzare la decade in cui cada la data di nascita, ma si è scelto di procedere così per uniformità con il primo insieme di dati.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-666

La pagina web di NaMo, riprodotta nell'immagine seguente, è dinamica:[21] esegue la ricerca su Wikidata, chiede a Koha l'elaborazione della risposta e presenta i risultati a schermo sotto forma di una tabella.

Figura 1 - L'homepage di NaMo, con le istruzioni e le sezioni per i due tipi di ricerca.

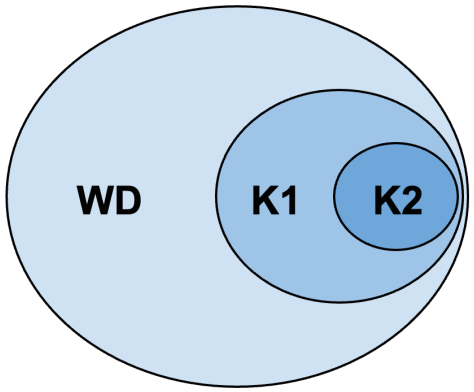
La ricerca di elementi con P5739 modificati di recente in Wikidata viene limitata ad un certo numero di giorni, nella misura massima di 365. Il default è 30, in considerazione che una persona che si occupi di metadatazione ci lavori con frequenza mensile. L'operatore potrà comunque cambiare questo periodo, in base all'impegno da dedicare a questa attività.
Nel caso delle ricerche nel passato, invece, si tratta di fissare un periodo di tempo nel quale potrebbe ricadere la data di morte di un autore. È stata scelta la durata di un decennio, e quindi l'operatore potrà indicare un anno qualunque e la ricerca sarà impostata automaticamente sul periodo intercorso tra il primo anno di quel decennio e l'anno scelto.[22] Sarà sua cura ricordare di avere già lavorato e completato un certo periodo, in modo da non dover effettuare la stessa ricerca più volte.
In entrambi i casi, le query non dovranno durare eccessivamente, per evitare attese inutili e soprattutto per non incorrere nel timeout di 60 secondi impostato dal server SPARQL di Wikidata, che implicherebbe un errore e quindi una risposta vuota.

[21] Le interrogazioni descritte più avanti e la composizione dei risultati da presentare a schermo sono svolte da un codice scritto in JavaScript, che fa uso della libreria jQuery https://jquery.com/ ed è incluso nella pagina HTML stessa.
[22] Ad esempio, se si sceglie il 1874 la ricerca verrà effettuata nell'intervallo 1871-1874.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-666

Il risultato di ogni query è un insieme di valori della proprietà P5739, vale a dire un elenco di identificatori di voci di autorità del catalogo Koha.[23] A ogni identificatore viene associato il valore delle date di nascita (anche se assente) e di morte, contenuti nelle rispettive proprietà dell'elemento di Wikidata.
Grazie alla completa apertura del suo codice sorgente, è stata aggiunto a Koha uno script in linguaggio Perl che riceve in ingresso l'elenco suddetto di identificatori; a seguito di una interrogazione del database, lo script restituisce solo quegli identificatori corrispondenti a record di autorità aventi il tag specifico[24] delle date non popolato in almeno una delle due parti. Insieme ad ogni identificatore, viene anche restituito il valore delle due date. Lo script di NaMo che ha effettuato le due interrogazioni è ora in grado di confrontare ogni coppia di date e mantenere solo i casi in cui si presenti una differenza.
Si tratta quindi dell'intersezione di due insiemi di dati, e di un filtro sul risultato dell'intersezione, come rappresentato dalla figura seguente.



WD Elementi di Wikidata con P5739, modificati di recente o con data di morte compresa in una determinata decade

K1 Voci di autorità del catalogo Koha appartenenti all'insieme WD ma con almeno una data mancante

K2 Voci di autorità dell'insieme K1 in cui almeno una data differisce da quelle presenti nell'insieme K1

Figura 2 - Rappresentazione insiemistica dei risultati delle ricerche e di quello finale.

La tabella costruita con i dati così ottenuti è composta come nella figura seguente:

[23] L'elenco consta ordinariamente di alcune centinaia di identificatori. Per ragioni di performance e di facilitazione del lavoro, conviene ottimizzare il sistema su questo ordine di grandezza, o inferiore.
[24] Il tag specifico e tecnicamente formattato, nel nostro caso del MARC21, è il tag 046. I sottocampi presi in considerazione in questo lavoro sono 'f' e 'g', rispettivamente per la data di nascita e la data di morte. Per una descrizione completa del tag 046 si veda https://www.loc.gov/marc/authority/ad046.html.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-666

	68777	Rovida, Cesare, morto 1862		1862
	Q56708351	Cesare Rovida	1785-09-19	1862
	65942	Rayneri, Giovanni Antonio		
	Q50328946	Giovanni Antonio Rayneri	1810-03-02	1867-06-04
	59782	Palmieri, Michele, 1779-	17791103	
	Q6012969	Michel Palmieri de Micciché	1779-11-03	1863-02-09
	54509	Montanelli, Giuseppe		
	Q3108102	Giuseppe Montanelli	1813-01-01	1862-06-17
	51162	Martínez de la Rosa, Francisco, 1787-1862		
	Q27357	Francisco Martínez de la Rosa	1787-03-10	1862-02-07

Figura 3 - Esempio di risultati per il periodo di data di morte 1861-1870.

Come si può notare, ogni riga della tabella è composta da celle a loro volta suddivise in due parti da un tratto di colore arancione: la parte superiore contiene dati provenienti dal catalogo Koha, la parte inferiore riporta dati provenienti dal corrispondente elemento di Wikidata.
La prima colonna contiene le rispettive icone, la seconda colonna gli identificatori, la terza colonna il tag 100[25] della forma scelta nel catalogo Koha e l'etichetta in italiano in Wikidata.[26] La quarta e la quinta colonna contengono le date di nascita e di morte e ne permettono il confronto.
I due identificatori della seconda colonna sono link attivi: l'identificatore della nostra voce di autorità conduce all'interfaccia Koha di modifica della voce stessa, per poter applicare la data o le date mancanti, mentre l'identificatore di Wikidata permette la visualizzazione dell'elemento. Entrambi aprono una nuova finestra del browser.
Per facilitare l'operatività e rendere più intuitivo lo strumento, sono stati aggiunti un bottone per la copia delle date (in formato EDTF) che possono essere inserite nel catalogo Koha e una evidenziazione in giallo della riga in lavorazione.
Le varie combinazioni possibili tra i dati delle colonne del nome e delle date costituiscono i casi d'uso descritti di seguito. E tra questi si trovano anche alcune positive conseguenze, non previste inizialmente nel progetto di questa utility.

Casi d'uso

In questa sezione raccoglieremo casi di utilizzo di NaMo al fine di mostrare alcune delle circostanze più rilevanti nelle quali lo strumento è stato utilizzato non solo per arricchire le entità (del nostro catalogo e in alcuni casi anche di Wikidata), ma anche per correggere alcuni errori di vario genere.

[25] Include i sottocampi abcd. Per una descrizione completa del tag 100 si veda https://www.loc.gov/marc/authority/ad100.html.
[26] Si noti la forma, sempre diretta, propria delle etichette in Wikidata.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-666

Arricchimento

	67852	Rodríguez Vicente, María Encarnación		
	Q131744085	María Encarnación Rodríguez Vicente	1925	1988
	67205	Rinaldi, Mario, morto 1985		1985
	Q21543404	Mario Rinaldi	1903-11-01	1985
	66892	Richard, Pierre, 1894-	1894	
	Q104524909	Pierre Richard	1894	1989-08-25

Figura 4 - Esempi di entità da arricchire.

Nel primo caso, quello ordinario, si può incrementare l'informazione all'interno della nostra voce di autorità: come si vede nella figura 4, i dati presentati risultano incompleti all'interno del nostro catalogo.
L'operazione da svolgere consiste nell'accedere al record di autorità PUSC, cliccando sul suo ID nella prima colonna, e inserire la datazione mancante sia nel tag 046 in formato EDTF, sia nel sottocampo d del tag 100. A questo punto la funzione di NaMo può considerarsi conclusa, mentre il catalogatore può proseguire verificando se la voce abbia bisogno di ulteriori aggiustamenti, modifiche, o se possa essere completata con informazioni provenienti da Wikidata o anche da altre fonti.

Errore derivato da omonimia

	127364	Wood, James, 1920-	19200722	
	Q6145777	James Wood	1760-12-14	1839-04-23

Figura 5 - Esempio di abbinamento errato.

In questo caso siamo in presenza di un evidente errore: due omonimi vissuti a circa 2 secoli di distanza sono stati erroneamente abbinati sia nel nostro catalogo che in Wikidata. Questa situazione potrebbe derivare da una procedura di riversamento automatico di dati da VIAF[27] all'interno di PUSC, oppure da un errore in sede di catalogazione da parte di un utente Wikidata o del catalogatore in Koha. Per risolvere il problema l'operatore dovrà scollegare le due entità, eliminando l'eventuale identificatore 024 Wikidata errato nel catalogo e l'ID PUSC dall'elemento Wikidata e verificando la corretta attribuzione dei record bibliografici. Successivamente, si potrà effettuare una ricerca per verificare se esistano degli identificatori corretti per ciascuna entità: a questo punto, nel caso in cui esista un elemento corrispondente in Wikidata, è necessario ristabilire il

[27] Tale procedura è stata effettuata due volte: nel marzo del 2019 e nel giugno del 2020. L'argomento è stato trattato più approfonditamente in (Bargioni 2020).

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-666

reciproco collegamento in Koha e Wikidata che garantisca il corretto match, proseguendo infine con l'inserimento di tutti i dati necessari all'identificazione.

Errore derivato da inversione di date

	1203	Aldricus Cenomanensis, morto 856	0856	
	Q592129	Aldrico di Le Mans	0800	0856-01-11

Figura 6 - Esempio di data di morte registrata erroneamente come data di nascita.

In rari casi abbiamo rilevato lo scambio tra la data di nascita e quella di morte, derivante da un errore di compilazione del tag 046. In questo caso specifico il sottocampo per la data di nascita ($f) è stato utilizzato al posto di quello per la data di morte ($g). Dal punto di vista operativo, trattandosi di un autore del IX secolo, occorrerà prima consultare il repertorio (Volpi 1994), che è stato individuato all'interno della Rete URBE come fonte primaria[28] per la datazione degli autori antichi e per l'individuazione dei titoli uniformi.
La datazione all'interno di Koha quindi deve essere uniformata a quella presente nella voce del repertorio Volpi corrispondente all'autore in questione e va corretta inserendola nel giusto sottocampo. La conformità alla fonte primaria del repertorio Volpi impone di non inserire le date all'interno della voce di autorità anche nel caso in cui esse si trovino in Wikidata.[29] Altri tipi di arricchimento sono invece possibili: ad esempio luoghi di nascita e morte, lingue associate, ecc.

Errore di datazione nel Catalogo PUSC

	48193	Luigi Maria da Carpi, O.F.M., morto 1885		1885
	Q113257483	Luigi Maria da Carpi	1815-12-18	1877-03-17

Figura 7 - Abbinamento corretto che evidenzia un errore di datazione.

In questo esempio si può notare che le date di morte indicate, seppur vicine nel tempo, sono differenti: il catalogatore dovrà per prima cosa verificare se si tratti di omonimia oppure se le due entità si riferiscano alla stessa persona.[30] Appurato che si tratta della stessa persona e che l'errore è nel catalogo PUSC, il catalogatore a questo punto dovrà correggere la data di morte e inserire quella di nascita.

[28] Questo si è reso necessario anche per conservare ed aumentare l'uniformità dei punti di accesso, che favorisce la corretta clusterizzazione all'interno del catalogo comune Parsifal.
[29] Non è stato ancora individuato un meccanismo per evitare che questi casi continuino a presentarsi in NaMo a causa della difformità tra la fonte prescelta nelle Varianti locali di URBE e i dati presenti in Wikidata.
[30] Le verifiche in questi casi possono essere effettuate attraverso un'analisi di altre fonti catalografiche, repertori e opere attribuite.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-666

Errore derivato da altre agenzie di catalogazione

	136966	Rojas Gálvez, Ignacio, 1971-	19711221	
	Q109526815	Ignacio Rojas Gálvez	1971-12-21	1971-12-21

Figura 8 - Errore generato da Library of Congress.

Questo caso evidenzia l'importanza del controllo bibliografico e della collaborazione tra le biblioteche. Come si vede, in Wikidata è stata inserita la stessa data per la nascita e la morte dell'autore; andando a verificare all'interno del cluster VIAF,[31] si noterà che l'errore è derivante dalla voce di autorità[32] della Library of Congress. In questo caso, la data corrisponde a quella della nascita: basterà correggere in Wikidata e, volendo, inviare una comunicazione all'agenzia statunitense per contribuire al mantenimento dei cataloghi in maniera globale.
In questo esempio risulta maggiormente rilevante l'aspetto di cooperazione e di controllo bibliografico universale, rispetto all'arricchimento del catalogo PUSC.

Statistiche

Si riportano di seguito due tabelle dei conteggi dei risultati ottenuti. Le due tabelle vengono presentate a scopo esemplificativo per mostrare le funzionalità dello strumento: i dati mostrati, infatti, dipendono sia dall'attività in Wikidata sugli elementi con P5739, che può variare fortemente nel tempo, sia dagli autori inseriti in catalogo per periodo di morte. Le tabelle forniscono una panoramica del lavoro di ricerca in Wikidata e Koha, delle proporzioni delle intersezioni e del filtro finale, e soprattutto della quantità di lavoro richiesto per l'aggiornamento del catalogo e quindi dei vantaggi nella qualità dell'identificazione.

giorni	trovati (WD)	ricerca e filtro in PUSC (K1)	da lavorare (K2)
1	467	40	39
10	3869	248	242
20	6156	471	458
30	9856	668	644
40	12114	771	744
50	13505	867	838
60	18308	1096	1064

Tabella 1 – Esempi di ricerche effettuate nella sezione "Recenti".

[31] https://viaf.org/en/viaf/187553661.
[32] https://web.archive.org/web/20250930221452/https://id.loc.gov/authorities/names/no2019082094.html.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-666

periodo	trovati (WD)	ricerca e filtro in PUSC (K1)	da lavorare (K2)
2021-2024	1871	316	312
2001-2010	3942	205	199
1991-2000	3066	141	136
1981-1990	2395	89	84
1971-1980	2026	70	68
1961-1970	1770	68	67
1951-1960	1371	65	65
1941-1950	1335	69	65
1931-1940	1107	37	36
1921-1930	746	35	35

Tabella 2 – Esempi di ricerche effettuate nella sezione “Nel passato”.

Applicazioni di NaMo in altri ambiti

Visti gli esiti dell’utilizzo di NaMo nell’ambito del catalogo Koha, è sorta l’ipotesi di estenderne l’uso, anche solo teorico, ad altre realtà, soprattutto a SBN per cui è in corso un’attività di arricchimento delle voci di autorità (persona).[33] Tratteremo nell’ordine BNF, GND e SBN.

Nel caso della Biblioteca Nazionale Francese (BNF),[34] è possibile per esempio usare la query seguente:

```
PREFIX rdf:  <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX bio:  <http://vocab.org/bio/0.1/>
SELECT ?surname ?givenName ?birth ?death
WHERE {
  ?id rdf:type foaf:Person .
  ?id foaf:givenName ?givenName .
  ?id foaf:familyName ?surname .
  OPTIONAL { ?id bio:birth ?birth . }
  OPTIONAL { ?id bio:death ?death . }
  VALUES ?id {
    <http://data.bnf.fr/ark:/12148/cb15562698n#about>
```

[33] Dal 2023 SBN utilizza Wikidata per arricchire le proprie voci di autorità; si veda: https://www.iccu.sbn.it/it/SBN/catalogazione-e-manutenzione-del-catalogo-sbn/progetti-di-implementazione-dei-dati-dellindice-sbn/arricchimento-dei-dati-di-authority-di-sbn-attraverso-il-confronto-e-lo-scambio-con-wikidata-/. Tale collaborazione e i risultati del primo anno di lavoro sono stati presentati nel novembre 2024 a Bologna; per la presentazione a cura di Elena Ravelli si vedano i seguenti documenti: https://www.wikidata.org/wiki/File:Authority_file_di_SBN_e_Wikidata_-_Wikidata_Days_Bologna_2024.pdf, https://www.wikidata.org/wiki/File:Authority_file_di_SBN_e_Wikidata_-_Wikidata_Days_Bologna_2024.webm.

[34] In questa biblioteca viene utilizzato il formato Unimarc, dove le date di nascita e morte sono registrate rispettivamente in 103 sottocampo a e in 103 sottocampo b.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-666


```
    <http://data.bnf.fr/ark:/12148/cb16176918m#about>
    <http://data.bnf.fr/ark:/12148/cb10546293p#about>
    <http://data.bnf.fr/ark:/12148/cb10693624f#about>
  }
}
```

da eseguire tramite l'endpoint SPARQL https://data.bnf.fr/sparql/.
Nel caso del GND,[35] si potrebbe usare la query

```
PREFIX :     <https://d-nb.info/gnd/>
PREFIX gndo: <https://d-nb.info/standards/elementset/gnd#>
SELECT ?id ?label ?birthdate ?deathdate WHERE {
    ?id gndo:preferredNameForThePerson ?label .
    OPTIONAL { ?id gndo:dateOfBirth ?birthdate . }
    OPTIONAL { ?id gndo:dateOfDeath ?deathdate . }
    VALUES ?id {
      :117458031
      :121501833
      :118766570
      :129554278
    }
}
```

da eseguire tramite l'endpoint SPARQL: https://zbw.eu/beta/sparql-lab/?endpoint=https://zbw.eu/beta/sparql/gnd/query.

Per applicare questa utility anche a SBN, la ricerca equivalente a K1 andrebbe realizzata basandosi su un server SPARQL, o tramite API che permettano di recuperare in maniera performante e poco dispendiosa per il server interrogato le voci di autorità recuperate con la ricerca Wikidata, ovviamente adattata da P5739 a P396.[36] Non pare vi siano strumenti di questo tipo disponibili per SBN.[37]
È stato invece fatto ricorso, almeno teorico, alla presenza in Wikidata del qualificatore "soggetto indicato come" (P1810), che frequentemente accompagna la suddetta P396. Quando presente, P1810 riporta in copia la forma scelta di SBN, che a sua volta può contenere le date di nascita o morte, o entrambe. La figura seguente mostra un esempio per RMSV010751,[38] il cui elemento Wikidata corrispondente è Q72406.[39]

[35] Il GND, o Gemeinsame Normdatei, è gestito dalla Biblioteca Nazionale Tedesca e dalle altre biblioteche che fanno parte del network dei paesi di lingua tedesca.
[36] P396 è la proprietà di Wikidata per registrare i VID: https://www.wikidata.org/wiki/Property:P396.
[37] Esiste un triplestore SBN con l'endpoint SPARQL https://triplestore.iccu.sbn.it/sparql ma non è dedicato alle voci di autorità. Le API documentate in https://api.iccu.sbn.it/devportal/apis e in particolare le API dei Nomi non contengono metodi di ricerca adatti allo scopo.
[38] http://id.sbn.it/bid/RMSV010751.
[39] http://www.wikidata.org/entity/Q72406#P396.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-666

Figura 9 - La voce RMSV010751 come appare nell'OPAC SBN al 15 marzo 2025.

Come si può notare, la forma "Harries, Carl Dietrich <1866-1923>" è presente in SBN e anche in Wikidata.[40]

Figura 10 - La dichiarazione P396 della voce RMSV010751 in Wikidata (Q72406) al 15 marzo 2025.

È pertanto possibile ricorrere ad un escamotage, che risulta certo limitato – come esaminato poco sotto – ma ha il seppur minimo vantaggio di unire le due interrogazioni WD e K1 in una sola. La query SPARQL è riportata in "Appendice tecnica".
I limiti di questa soluzione sono dovuti al fatto che in realtà, per la normativa corrente,[41] in SBN la qualificazione si usa soltanto per disambiguazione interna; in assenza di omonimi, le date vengono inserite soltanto nell'apposito campo Datazione; tali date pertanto non comparirebbero nel qualificatore P1810 di Wikidata e non sarebbero sfruttabili per questo lavoro. Inoltre, va tenuto presente che tra le parentesi uncinate a volte non vengono riportate solo le date, ma anche altre informazioni disambiguanti, rendendo complicata la stesura del filtro sulle due date. Con le sole date verrebbero comunque individuati casi come nella tabella seguente.

[40] Va notato che è possibile che P1810 non contenga un valore aggiornato, dato che non è previsto un automatismo che lo assicuri.
[41] Regole Italiane di Catalogazione (REICAT), disponibili in pdf al link https://www.iccu.sbn.it/export/sites/iccu/documenti/2015/REICAT-giugno2009.pdf. I costanti aggiornamenti sono riportati su https://norme.iccu.sbn.it/index.php/Normative_catalografiche. Nello specifico, per le qualificazioni consultare https://norme.iccu.sbn.it/index.php?title=Norme_comuni/Authority_file/Nomi_di_persona/Registrazione_di_authority/Qualificazioni.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-666

	CFIV007224 Q183074	Juana Inés : de la Cruz Juana Inés de la Cruz	– 1651	– 1695
	CFIV080315 Q313001	Lodge, David <1935- > David Lodge	1935 1935	– 2025
	SBLV250082 Q1083104	Parigi, Alfonso <m. 1656> Alfonso Parigi il Giovane	– 1606	1656 1656

Tabella 3 - Esempi di casi teorici per l'arricchimento in SBN.

Rispetto al caso basato su Koha e ai due casi basati su SPARQL, si nota la complessità e il rischio di errori insiti in questa soluzione per SBN, e soprattutto il ridotto numero di situazioni individuate dalla query.

Conclusioni

Lo strumento NaMo è un esempio di riuso di dati specifici di Wikidata per migliorare le voci di autorità. A differenza di alcuni contesti come ad esempio Alphabetica,[42] Parsifal[43] e CoBiS[44] che sfruttano il riuso nell'OPAC, NaMo opera un riuso dei dati di Wikidata al fine di registrarli all'interno del catalogo.
Wikidata, oltre a essere utile per le riconciliazioni, è anche fonte di dati costantemente e prontamente aggiornati che possono essere utilizzati per migliorare l'identificazione; la sollecitudine nell'aggiornamento del proprio catalogo influisce anche sull'immagine della biblioteca, che mostra di essere proattiva soprattutto nel caso di cambiamenti biografici di personaggi importanti (morte, cambio di incarico pubblico o di stato di canonizzazione, ecc.).
Il lavoro ha inoltre evidenziato l'importanza della granularità dei dati, in particolare delle date, e della necessità di disporre di strumenti di interrogazione per raggiungere le entità da arricchire; inoltre è stata messa in luce l'importanza di avere a disposizione dati pubblicati in formati standard come XML o RDF, oppure interrogabili con un server SPARQL. Questa esigenza è particolarmente evidente per i grandi sistemi, come è il caso di SBN.
Le operazioni di arricchimento hanno avuto come rilevante effetto collaterale l'individuazione dei tipici errori di gestione delle entità, quali conflazioni e attributi errati. Attraverso la visualizzazione tabellare è stato possibile rilevare le incongruenze con un semplice confronto visivo; questo dimostra la necessità di un controllo e di un intervento umano sui dati.
Tra i possibili sviluppi futuri si può considerare il trattamento delle date incerte, che in MARC21

[42] Alphabetica fa riuso di dati di Wikidata e Wikimedia in tempo reale: si confronti per esempio la voce di autorità del matematico Robert L. Devaney https://opac.sbn.it/bid/UFIV023576 con la corrispondente pagina "protagonisti" https://alphabetica.it/web/alphabetica/protagonisti-risultati/-/s/results?input=UFIV023576!%3B!Devaney!%3B!Robert!%3B!L.!%3B!%3C1948-!%3B!%3E. Per un ulteriore approfondimento sul tema si veda (Bianchini, Bargioni, e Pellizzari di San Girolamo 2022, 281-286).
[43] Un esempio di riuso di Wikipedia in Parsifal: https://parsifal.urbe.it/parsifal/searchNames?n_cluster_id=14957.
[44] Un esempio di entità in CoBiS: https://dati.cobis.to.it/agent/95a5yja38danygu695b5yc1j70u30cr.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-666

coinvolgono i sottocampi s e t del tag 046. Questi casi sicuramente possono apportare maggiore qualità e precisione alla descrizione delle entità legate a autori antichi, ma richiedono la massima attenzione e soprattutto l'adesione ad eventuali norme proprie della biblioteca, che per esempio possono includere il ricorso a determinati repertori.
Inoltre si può prevedere di estendere il metodo di NaMo agli arricchimenti di luoghi di nascita e morte, occupazioni, gruppi di appartenenza, ecc.[45]
Infine, come mostrato dalle query teoricamente applicabili a BNF e GND, disporre di un server SPARQL con cui interrogare i dati del catalogo da arricchire renderebbe decisamente più semplici le operazioni, grazie alle query federate previste dal protocollo SPARQL stesso,[46] con cui si potrebbe interrogare contemporaneamente anche Wikidata.

Appendice tecnica

Di seguito sono elencate le query utilizzate per NaMo e per le eventuali applicazioni in altri ambienti. La notazione ${…}, mutuata dal linguaggio JavaScript, all'interno delle query indica un parametro che va sostituito con valori impostati dall'utente.

La query SPARQL utilizzata per le ricerche recenti:

```
# elementi con PUSC ID P5739 modificati negli ultimi N giorni
SELECT ?person ?PUSC ?personLabel ?dod ?dod_prec ?dob ?dob_prec ?mod_time WHERE {
    ?person wdt:P5739 ?PUSC;
      wdt:P31 wd:Q5;
      schema:dateModified ?mod_time.
    FILTER(?mod_time > ((NOW()) - "P${giorni_recenti}D"^^xsd:duration))
    OPTIONAL {
      ?person p:P569 ?dob_st.
      ?dob_st rdf:type wikibase:BestRank;
        psv:P569 ?dob_v.
      ?dob_v wikibase:timePrecision ?dob_prec;
        wikibase:timeValue ?dob.
    }
    ?person p:P570 ?dod_st.
    ?dod_st rdf:type wikibase:BestRank;
        psv:P570 ?dod_v.
    ?dod_v wikibase:timePrecision ?dod_prec;
        wikibase:timeValue ?dod.
    SERVICE wikibase:label { bd:serviceParam wikibase:language "it,la,mul,en". }
}
```

[45] In MARC21, i luoghi di nascita e morte sono descrivibili nel tag 370; i gruppi di appartenenza nel tag 373, le occupazioni nel tag 374, il genere nel tag 375, ecc. In tutti questi casi va prevista la registrazione dell'URI del concetto. In Wikidata, i dati suddetti si trovano rispettivamente nelle proprietà P19, P20, P108, P106, P21.
[46] Per una trattazione esaustiva del protocollo SPARQL, si veda (DuCharme 2013) e la pagina https://www.w3.org/TR/sparql11-query/.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-666

La query SPARQL utilizzata per le ricerche nel passato:

```
SELECT ?person ?PUSC ?personLabel ?dod ?dod_prec ?dob ?dob_prec WHERE {
    ?person wdt:P5739 ?PUSC;
      p:P570 ?dod_st.
    ?dod_st rdf:type wikibase:BestRank;
      psv:P570 ?dod_v.
    ?dod_v wikibase:timePrecision ?dod_prec;
      wikibase:timeValue ?dod.
    FILTER(
      ?dod < "${fine_periodo_nel_passato}-12-31T23:59:59Z"^^xsd:dateTime
    )
    FILTER(
      ?dod > "${inizio_periodo_nel_passato}-01-01T00:00:00Z"^^xsd:dateTime
    )
    OPTIONAL {
      ?person p:P569 ?dob_st.
      ?dob_st rdf:type wikibase:BestRank;
        psv:P569 ?dob_v.
      ?dob_v wikibase:timePrecision ?dob_prec;
        wikibase:timeValue ?dob.
    }
    SERVICE wikibase:label { bd:serviceParam wikibase:language "it,la,mul,en". }
}
```

La query SQL utilizzata in Koha in base agli identificatori reperiti in una delle due query SPARQL precedenti:[47]

```
SELECT authid, ExtractValue(`marcxml`,'//datafield[@tag="100"]/*') t100,
    ExtractValue(`marcxml`,'//datafield[@tag="046"]/subfield[@code="f"]') t046f,
    ExtractValue(`marcxml`,'//datafield[@tag="046"]/subfield[@code="g"]') t046g
FROM auth_header
WHERE authid IN ( ${lista_di_ID} )
    AND (ExtractValue(`marcxml`,'//datafield[@tag="046"]/subfield[@code="f"]') = ''
    OR ExtractValue(`marcxml`,'//datafield[@tag="046"]/subfield[@code="g"]') = '')
ORDER BY authid DESC
```

[47] Si tenga presente che in Koha le voci di autorità sono conservate nella tabella “auth_header”, campo ”marcxml”, nel formato omonimo. Con i dati in formato XML, nel database MariaDB di Koha si ricorre alla funzione “ExtractValue” e alla sintassi XPATH per accedere ai valori di “datafield” e “subfield”.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-666

La query per SBN:[48]

```
# elementi con SBN ID P396 modificati negli ultimi 2 giorni contenenti "soggetto
# indicato come" P1810, a sua volta contenente "<"
SELECT ?person ?sbn ?personLabel ?forma_in_sbn ?dob ?dob_prec ?dod ?dod_prec # ?mod_time
WHERE {
    ?person wdt:P396 ?sbn;
      wdt:P31 wd:Q5;
      schema:dateModified ?mod_time.
    FILTER(?mod_time > ((NOW()) - "P2D"^^xsd:duration))
    OPTIONAL {
      ?person p:P569 ?dob_st.
      ?dob_st rdf:type wikibase:BestRank;
      psv:P569 ?dob_v.
      ?dob_v wikibase:timePrecision ?dob_prec;
      wikibase:timeValue ?dob.
    }
    ?person p:P570 ?dod_st.
    ?dod_st rdf:type wikibase:BestRank;
      psv:P570 ?dod_v.
    ?dod_v wikibase:timePrecision ?dod_prec;
      wikibase:timeValue ?dod.

    ?person p:P396 ?sbn_st .
    ?sbn_st pq:P1810 ?forma_in_sbn .
    FILTER(CONTAINS(?forma_in_sbn, "<"))

    SERVICE wikibase:label { bd:serviceParam wikibase:language "it,la,mul,en". }
}
```

Conteggi PUSC (17.6.2025)

110745 personal name:
`https://catalogo.pusc.it/cgi-bin/koha/opac-authorities-home.pl?op=do_search&type=opac&authtypecode=PERSO_NAME`

024 $2 wikidata sono 64255 (58,0%):

```
SELECT count(*) FROM auth_header
WHERE authtypecode = 'PERSO_NAME'
AND ExtractValue(`marcxml`,'//datafield[@tag="024"]/subfield[@code="2"]') LIKE
'%wikidata%';
```

In Wikidata ci sono 69905 elementi con P5739 (63,1%):

```
SELECT (COUNT( DISTINCT ?q) as ?c) {?q wdt:P5739 ?i.}
```

Conteggi SBN (17.6.2025)

[48] Per eseguire la query e visualizzarne i risultati si può utilizzare il seguente link: https://w.wiki/EVun.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-666

Le voci di autorità di tipo persona visibili in OPAC SBN sono 3771027:[49]
`https://opac.sbn.it/risultati-autori?core=autori&filter_nocheck%3A6021%3ATipo_nome=Persona%3AA`

In Wikidata ci sono 190029 elementi con P396 (5,03%):
```
SELECT (COUNT( DISTINCT ?q) as ?c) {?q wdt:P396 ?vid.}
```

[49] Questo conteggio non include le voci di livello 05, come spiegato in (Atturo e Ravelli 2024, 51).

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-666

Riferimenti bibliografici*

Atturo, Valentina, e Elena Ravelli. 2024. "The Evolution of Authority Work in SBN. From Origins to Alphabetica and Future Prospects." *JLIS.it* 15 (1): 45-58. https://doi.org/10.36253/jlis.it-572.

Bargioni, Stefano. 2020. "From Authority Enrichment to AuthorityBox. Applying RDA in a Koha environment." *JLIS.it* 11 (1): 175–89. https://doi.org/10.4403/jlis.it-12595.

Bianchini, Carlo. 2024. "Nessun catalogo è un'isola." In *Parsifal: un modello di collaborazione bibliotecaria per condividere la conoscenza registrata*, a cura di Silvano Danieli, 57-72. Firenze: Firenze University Press. https://doi.org/10.36253/979-12-215-0356-2.10.

Bianchini, Carlo, Stefano Bargioni, and Camillo Carlo Pellizzari di San Girolamo. 2021. "Beyond VIAF: Wikidata As a Complementary Tool for Authority Control in Libraries". In: *Information Technology and Libraries* 40 (2). https://doi.org/10.6017/ital.v40i2.12959.

Bianchini, Carlo, e Lucia Sardo. 2022. "Wikidata: A New Perspective towards Universal Bibliographic Control." *JLIS.it* 13 (1): 291-311. https://doi.org/10.4403/jlis.it-12725.

Bianchini, Carlo, Stefano Bargioni, e Camillo Carlo Pellizzari di San Girolamo. 2022. "Le voci di autorità dei nomi di persona in SBN e Alphabetica: problemi e prospettive." *Bibliothecae.it* 11 (1): 247-314. https://doi.org/10.6092/issn.2283-9364/15078.

Danieli, Silvano, a c. di. 2024. *Parsifal: un modello di collaborazione bibliotecaria per condividere la conoscenza registrata.* Firenze: Firenze University Press. https://doi.org/10.36253/979-12-215-0356-2.

DuCharme, Bob. 2013. *Learning SPARQL, 2nd Edition*. Cambridge: O'Reilly Media, Inc.

Guerrini, Mauro, e Lucia Sardo. 2018. *IFLA Library Reference Model (LRM): un modello concettuale per le biblioteche del XXI secolo*. Milano: Editrice Bibliografica.

Leoni, Cristiana. 2024. "Quindici anni di catalogazione in URBE: dalle Varianti locali (2009) alla Commissione sulle Varianti locali e la catalogazione in URBE (2024)." In *Parsifal: un modello di collaborazione bibliotecaria per condividere la conoscenza registrata*, a cura di Silvano Danieli, 147-62. Firenze: Firenze University Press. https://doi.org/10.36253/979-12-215-0356-2.15.

Manzoni, Laura. 2022. *Identificatori*. Roma: Associazione italiana biblioteche.

Pellizzari di San Girolamo, Camillo Carlo. 2024. "Storia della collaborazione tra Wikidata e le biblioteche della Rete URBE nel controllo di autorità." In *Parsifal: un modello di collaborazione bibliotecaria per condividere la conoscenza registrata*, a cura di Silvano Danieli, 147-62. Firenze: Firenze University Press. https://doi.org/10.36253/979-12-215-0356-2.17.

Possemato, Tiziana. 2023. "Linked data: un'opportunità per il riuso." *DigItalia* 2: 134-46. https://doi.org/10.36181/digitalia-00081.

Possemato, Tiziana. 2024. *Entity modeling: la terza generazione della catalogazione*. Firenze: Firenze University Press. https://doi.org/10.36253/979-12-215-0393-7.

* Ultima data di consultazione dei siti web: 28 giugno 2025.

JLIS.it vol. 17, no. 1 (January 2026)
ISSN: 2038-1026 online

DOI: 10.36253/jlis.it-666

Possemato, Tiziana, Annalisa Di Sabato, e Alessandra Moi. 2024. "Parsifal: armonizzare la tradizione con la modernità. L'Authority file condiviso di URBE come nuovo terreno di collaborazione." In *Parsifal: un modello di collaborazione bibliotecaria per condividere la conoscenza registrata*, a cura di Silvano Danieli, 81-101. Firenze: Firenze University Press. https://doi.org/10.36253/979-12-215-0356-2.12.

Volpi, Vittorio. 1994. *DOC: dizionario delle opere classiche: intestazioni uniformi degli autori, elenco delle opere e delle parti componenti, indici degli autori, dei titoli e delle parole chiave della letteratura classica, medievale e bizantina*. Milano: Editrice Bibliografica.

Wiederhold, Rebecca A., e Gregory F. Reeve. 2021. "Authority Control Today: Principles, Practices, and Trends." *Cataloging & Classification Quarterly* 59 (2-3): 129-58. https://doi.org/10.1080/01639374.2021.1881009.