

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.

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-

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 Fortegueriana <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).

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.

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 2003¹³

- 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.

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 legale¹⁵ 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 2017¹⁶ 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".

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 inglese¹⁸ 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.

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 depositarie 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 Library²³ 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.

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/>.

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 archive³⁰ 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 longitudinale³¹ (ad esempio, studiando l'evoluzione di un fenomeno nel tempo attraverso i documenti del corpus) o di tipo trasversale³² (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.

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.

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 automatizzata³⁷ 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_\(Enciclopedia-dell'Italiano\)/](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_\(deep_learning_architecture\)](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).

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 2006⁴⁸ 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 incrementale⁴⁹. 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-Project⁵² 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 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”.

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>.

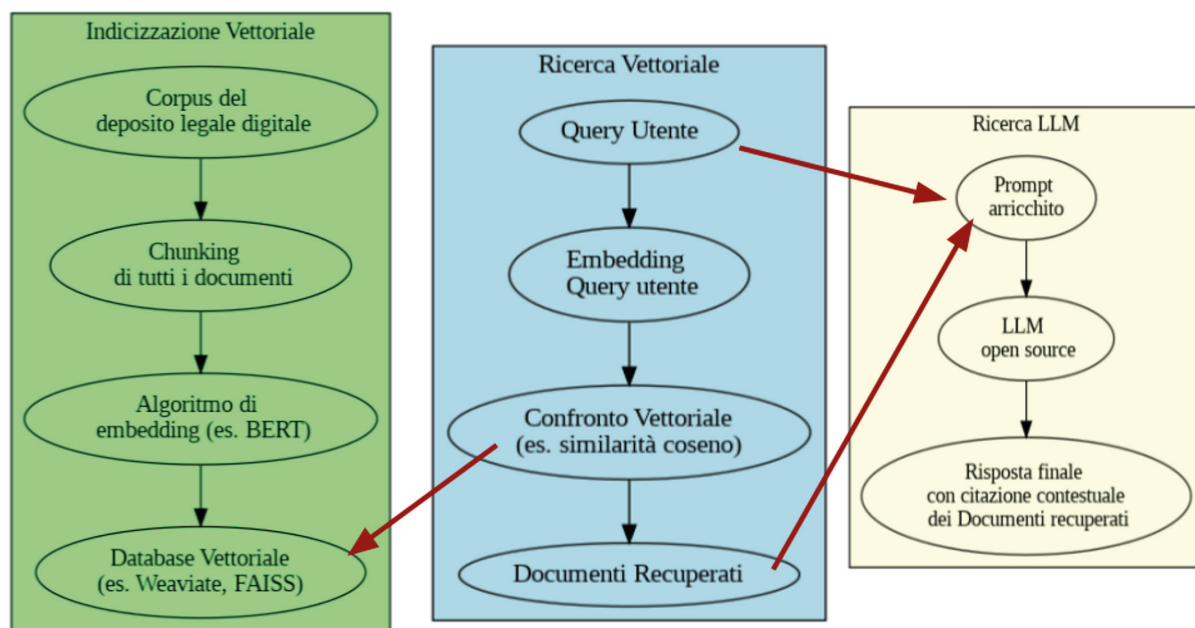


Fig. 1 Modello di computational access con un sistema RAG per il deposito legale digitale

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. BERT⁵⁷);
- 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 coseno⁵⁸) 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.

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 Mistral⁶¹) 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 controllati⁶⁴ 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_\(modello_linguistico\)](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_\(Neologismi\)/](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).

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 culturali⁶⁶, 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).

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.

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.

Riferimenti bibliografici

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

Fan, Wenqi, Yajuan 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, Osmo 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 Sammlungsgedanke 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.

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, Osmo, 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, Osmo, 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 cybersecurity 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>.

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>.