digitcult @Scientific Journal on Digital Cultures

DigitCult
ISSN 2531-5994
VOL. 11, NO. 1
(2026)

# Modelli di classificazione e di gestione dinamica delle collezioni digitali tra smart digital libraries e Intelligenza Artificiale. Sfide e prospettive

## Classification and Dynamic Management Models for Digital Collections between Smart Digital Libraries and Artificial Intelligence. Challenges and Perspectives

Nicola Barbuti
0000-0003-0817-4235
Università degli Studi di Bari "Aldo Moro"
ROR 027ynra39
nicola.barbuti@uniba.it

10.54103/2531-5994/31789

## Abstract

**[It.]** La crescita delle raccolte digitizzate e born-digital, l'accelerazione della trasformazione digitale e il recente dilagare dell'Intelligenza Artificiale richiedono un ripensamento dei modelli e delle pratiche di gestione delle collezioni. Le digital libraries non possono più essere trattate come depositi di surrogati, immagini, file o record statici e chiusi: sono ambienti nei quali le risorse vengono prodotte, descritte, pubblicate, collegate, verificate, versionate, corrette e riusate. Questo contributo propone una rielaborazione dei concetti di digitizzazione, digital heritage, metadati, provenance, lifecycle, Principi FAIR, knowledge graph e AI-assisted cataloguing, identificandoli come componenti di un unico modello teorico-metodologico. La classificazione delle collezioni digitali è interpretata non come attribuzione puntuale e definitiva di una classe, ma come processo documentato, verificabile e versionabile, nel quale l'AI contribuisce a produrre ipotesi, riconoscimenti e connessioni da ricondurre a responsabilità che hanno matrice nelle istanze e nelle riflessioni biblioteconomiche, archivistiche e umanistico-digitali.

**Parole chiave:** Gestione dinamica; Digital heritage; Smart digital libraries; Knowledge graphs; Classificazione automatica..

**[En.]** The growth of digitised and born-digital collections, the acceleration of digital transformation and the recent spread of Artificial Intelligence require a reconsideration of collection management models and practices. Digital libraries can no longer be treated as repositories of surrogates, images, files or static and closed records: they are environments in which resources are produced, described, published, linked, verified, versioned, corrected and reused. This paper proposes a theoretical and methodological dynamic model to manage digital collection, which integrates the concepts of digitisation, digital heritage, dynamic metadata, provenance, lifecycle, FAIR Principles, knowledge graph and AI-assisted cataloguing. The automatic classification of digital collections is interpreted not as the punctual and definitive attribution of a class, but as a documented, verifiable and versioning process in which AI contributes to producing



Received: 29/05/2026
Accepted: 16/07/2026
Published: 06/07/26



Milano University Press

DigitCult
ISSN 2531-5994
VOL. 11, NO. 1
(2026)

hypotheses, recognitions and connections that must be traced back to responsibilities rooted in library, archival and digital-humanities principles and reflections.

**Keywords:** Dynamic management; Digital heritage; Smart digital libraries; Knowledge graphs; Automatic classification.

# 1. Introduzione

La trasformazione digitale delle istituzioni culturali ha modificato in profondità la produzione, l'organizzazione e l'utilizzo delle collezioni documentarie. Biblioteche, archivi, musei e gallerie hanno riversato in rete quantità crescenti di risorse digitali in immagini, descrizioni e ambienti di accesso. L'accelerazione quantitativa va letta nel contesto conseguente alla pandemia del 2020, alla temporanea sospensione della fruizione fisica e ai successivi programmi pubblici di investimento, intrapresi in seguito all'incremento della domanda di soluzioni digitali connesse con il patrimonio culturale da fruire in remoto. La crescita quantitativa, tuttavia, non ha presupposto modelli capaci di interpretare, classificare, organizzare, gestire, preservare e comunicare le collezioni digitali. La presenza di file in rete non coincide con la loro intellegibilità documentaria: una risorsa può essere accessibile e agevolmente fruibile ma povera di contesto, tecnicamente conservata ma incapace di testimoniare il processo che l'ha prodotta e restituire le informazioni sul suo ciclo di vita.

Per questa ragione, il riferimento al 2020 non va assunto come cesura meramente cronologica, ma come indicatore di un mutamento di scala: la chiusura temporanea degli spazi fisici di consultazione, l'urgenza della fruizione remota e la concentrazione di investimenti pubblici hanno reso evidente che l'accesso digitale non può più essere trattato come servizio accessorio. La quantità di risorse prodotte ha mostrato, in modo più netto di prima, che la digitizzazione non è solo attività tecnica di riproduzione, ma è processo di costruzione di oggetti documentari destinati a essere descritti, classificati, interrogati e preservati nel tempo. Ne consegue che la gestione delle collezioni digitali necessita di un ripensamento profondo sul piano epistemologico e metodologico, oltre che tecnico.

Infatti, sebbene le *digital libraries* a tema patrimonio culturale abbiano ampliato l'accesso e reso consultabili risorse altrimenti disperse e difficilmente raggiungibili fisicamente, molte di esse hanno travasato nella dimensione digitale impianti concettuali e approcci gestionali peculiari al patrimonio analogico, rimanendo legate al modello del *repository*: contenitori di oggetti digitali variamente organizzati e corredati da metadati descrittivi, immagini e PDF, interrogabili soprattutto attraverso ricerche testuali che puntano a descrizioni catalografiche. Questo modello ha aumentato la disponibilità delle risorse digitali, ma non sempre ne ha accresciuto la comprensione, trascurando di documentarne la provenienza, la struttura, il rapporto con l'artefatto analogico, le

* Il presente contributo rappresenta una sintesi di un più ampio lavoro monografico in lavorazione e di prossima pubblicazione.

DigitCult
ISSN 2531-5994
VOL. 11, NO. 1
(2026)

trasformazioni subite e le responsabilità descrittive che ne determinano la forma e lo stato correnti.

Il problema diventa più evidente quando le collezioni comprendono risorse digitali eterogenee sia per genesi, formato e tipologia, che per categorie di artefatti culturali rappresentati. L'impianto concettuale delle descrizioni bibliografiche, archivistiche, museali e storico-artistiche resta indispensabile, ma deve confrontarsi con oggetti differenti per generazione composti da bit, formati, strutture bidimensionali, tridimensionali e born-digital, metadati, relazioni, licenze, log, versioni e derivazioni.

L'ingresso dell'Intelligenza Artificiale (IA) nella produzione e gestione del digitale a tema culturale rende la revisione ancora più urgente. La letteratura scientifica recente mostra che metodi e funzionalità di Natural Language Processing (NLP), modelli matematici complessi basati su Reti Neurali Convoluzionali (CNN) sempre più evolute, tecniche avanzate di Optical Character Recognition (OCR), Handwritten Text Recognition (HTR), Intelligent Character Recognition (ICR) sono sperimentate a supporto di asset biblioteconomici quali discovery, reference, raccomandazione, arricchimento di metadati, catalogazione, soggettazione, classificazione, analisi del layout, supporto alla preservazione (Cox, Pinfield, & Rutter, 2019; Cox & Mazumdar, 2024, pp. 333-334; Huang, 2024; Engel et al., 2025, p. 261). Quanto emerge è che quando l'IA entra nei processi descrittivi non produce soltanto efficienza, ma modifica nella sostanza il rapporto tra competenza professionale, interpretazione documentaria, responsabilità istituzionale e produzione di conoscenza.

La domanda urgente, quindi, può essere formulata in questi termini: *come può la biblioteconomia elaborare modelli di classificazione automatica e gestione dinamica delle collezioni digitali che integrino metadati, provenance, lifecycle, knowledge graph e IA, senza indebolire responsabilità scientifica, verificabilità dei processi e controllo esperto?*

Una prima risposta potrebbe essere definire le linee per evolvere l'attuale modello di gestione biblioteconomica descrittivo-statico in un modello relazionale e computazionalmente trattabile, nel quale ogni classificazione, arricchimento, correzione o inferenza sia registrato e reso accessibile nel tempo come *evento* del ciclo di vita della risorsa digitale (Barbuti, Ferilli, & Caldarola, 2022; Barbuti, 2024). Il suo valore dipende tanto dalla qualità della rappresentazione, quanto dalla possibilità di evolvere in *record* registrando nei metadati informazioni accreditate, verificabili e affidabili su storia, funzioni, trasformazioni e condizioni di uso e riuso che ne documentino il ciclo di vita.

E qui si rende indispensabile un breve chiarimento del concetto di *record digitale*. Un *dato digitale* è una porzione di informazione formalizzata; esso diventa *risorsa* quando lo si identifica e descrive, ed evolve in *record* quando è collocato in un sistema che ne esplicita responsabilità, identificazione, contesti e relazioni. La sequenza di bit di un'immagine non dice da quale originale derivi, quale processo l'abbia generata, quali diritti ne regolino l'uso o quale funzione svolga in una collezione. Perché il *dato* diventi *record* occorre una trama di metadati che ne definisca la sostanza documentaria. Classificare una risorsa digitale, quindi, significa inserirla in una rete di intellegibilità, nella quale contenuto rappresentato, unità strutturale, derivato di consultazione, workflow, grafo semantico, validazione o input per un modello di riconoscimento non sono attri-

DigitCult
ISSN 2531-5994
VOL. 11, NO. 1
(2026)

buti isolati, ma relazioni formalizzate. Ne deriva che il *record* non è soltanto una scheda descrittiva, né un contenitore di campi compilati. È la forma documentaria attraverso cui un oggetto digitale diventa identificabile, attribuibile, verificabile e discutibile. In esso devono poter coesistere informazioni stabilizzate e informazioni provvisorie, dati prodotti da operatori umani e dati generati da sistemi automatici, elementi descrittivi e tracce di processo. La sua qualità non dipende dalla fissità, ma dalla capacità di rendere intellegibili le trasformazioni che registra (Barbuti, 2022, pp. 181-182).

In questa transizione, l'IA non va assunta *tout court* come sostituto dell'organizzazione e gestione delle collezioni, ma va utilizzata in tutto il suo potenziale come componente di workflow documentati, interpretabili, revisionabili e orientati alla costruzione di memoria digitale preservabile e, nel contempo, trasferibile.

La riflessione, dunque, va centrata soprattutto sullo statuto epistemico delle classificazioni automatiche. Una proposta prodotta da un sistema computazionale non implica una validazione esperta, motivo per cui questa deve essere distinta, attribuita, documentata e, se accolta, incorporata nella risorsa digitale, evidenziandone la responsabilità documentaria.

Il presente contributo si focalizza su questo punto cruciale: non l'introduzione generica, casuale o circostanziale di automazione pseudo-intelligente nei processi biblioteconomici, ma la definizione di condizioni e linee biblioteconomiche che consentano di gestire dinamicamente i *record* computazionali come atti documentari dotati di agente, tempo, versione, affidabilità e stato di validazione. In assenza di tale distinzione, una classificazione prodotta da un modello rischia di essere assorbita nel record come se avesse lo stesso statuto di una decisione professionale, cancellando proprio la differenza che dovrebbe renderla valutabile.

Pur conservando un impianto teorico-metodologico, si riconduce l'orizzonte programmatico della riflessione a una proposta verificabile: definire condizioni, livelli e responsabilità attraverso cui l'automazione possa entrare nella gestione dinamica delle collezioni digitali senza trasformarsi in autorità descrittiva non controllata, e la classificazione sia esito della ridefinizione del rapporto tra dato, metadato, record, relazione, inferenza e responsabilità.

## 2. Digitalizzazione, digitizzazione e digital heritage: fondamenti concettuali per una biblioteconomia del digitale

La disambiguazione lessicale e concettuale è il primo passaggio per definire un modello gestionale dinamico delle collezioni digitali. In Italia, il termine *digitalizzazione* è usato per indicare indistintamente generazione di risorse da analogico, innovazione dei processi, produzione di immagini, automazione di servizi e creazione di risorse native o derivate. Questa polisemia ha prodotto un'ambiguità metodologica: nella digitalizzazione identificata con la produzione di "copie", l'oggetto digitale resta un derivato funzionale all'originale; quando, invece, si riconosce la specificità della digitizzazione, la risorsa

DigitCult
ISSN 2531-5994
VOL. 11, NO. 1
(2026)

può essere descritta come entità documentaria dotata di struttura, storia, contesto e relazioni (Brennen & Kreiss, 2014).

Nel lessico internazionale, *digitalization* e *digitization* distinguono l'innovazione dei processi dai flussi di generazione di entità digitali (Brennen & Kreiss, 2014; Bloomberg, 2018). Nel contesto culturale, questa distinzione è decisiva: la *digitalizzazione* include pratiche, organizzazioni, servizi e modelli di relazione tra istituzioni; la *digitizzazione* riguarda «la creazione e materializzazione di entità digitali» sia derivate da scansione di artefatti analogici che born-digital, integrate da metadati capaci di registrarne «*provenance*, ciclo di vita e relazioni tra contenuti e contesti» (Barbuti, 2024, p. 79). Confondere i livelli porta a misurare un progetto sulla quantità di file prodotti o sulla visibilità dell'interfaccia, mentre la qualità scientifica e culturale dipende dalla coerenza dell'intero ciclo: selezione, acquisizione, generazione, controllo, metadatazione, esposizione, uso, preservazione e riuso. Del resto, la digitizzazione non è mai *neutra*: risoluzione, formato, denominazione delle risorse, profilo colore, struttura dei pacchetti e modalità di esposizione determinano ciò che sarà leggibile, interrogabile, preservabile, verificabile, affidabile e trasferibile. Gli interventi automatici devono perciò essere documentati con lo stesso rigore delle scelte descrittive e di classificazione.

Il che rende indispensabile un chiarimento sull'imperversante istanza secondo cui le risorse digitali di artefatti culturali riproducono *digital twins* degli originali. La letteratura più recente sembra avere finalmente ridefinito il concetto di *gemello digitale* emancipandolo da quello di *duplicato virtuale di un originale*. Nelle applicazioni ad architetture, siti, manufatti monitorabili e ambienti nei quali dati sensoriali e modelli computazionali sostengono decisioni conservative, esso è inteso come rappresentazione *data-driven* e aggiornata, utile a monitoraggio, conservazione, gestione del rischio, simulazione e fruizione immersiva (Dagnaw, Capuano, & Muccini, 2026). Nondimeno, l'estensione della nozione alle risorse e alle collezioni digitali a tema patrimonio è del tutto inappropriata, poiché non vi è alcuna identità operativa tra originale e rappresentazione, ma generazione di una nuova entità documentaria.

Ne consegue che ridurre indiscriminatamente la digitizzazione a produzione di *digital twins* nei quali *si travasano* "copie" o gemelli virtuali di artefatti analogici ne oscura componenti decisive. Una pagina scansionata può apparire corretta sul piano visivo, ma essere priva di metadati tecnici affidabili e, quindi, non verificabile; può essere dettagliatamente descritta, ma scollegata dalla sequenza fisica del volume; può essere consultabile, ma non associata a diritti chiari o a una strategia di preservazione. In ciascun caso, l'oggetto esiste come file, ma non raggiunge lo statuto di risorsa culturalmente governata, affidabile e tracciabile. Anche l'IA cambia funzione a seconda di questa interpretazione: se il digitale è mero *travaso* generativo di *gemelli digitali*, l'automazione è mero acceleratore di attività tecniche; se, invece, esso è ecosistema documentario, l'IA diventa agente che produce e integra metadati nel *lifecycle* delle risorse, propone classificazioni, sostanzia i record con rigore descrittivo documentato e ne definisce la trama narrativa, incidendo sull'esplorazione delle collezioni.

La distinzione, perciò, deve focalizzarsi sulla delimitazione del suo campo di pertinenza. Nei casi in cui il *gemello digitale* è costruito per simulare comportamento, mo-

DigitCult
ISSN 2531-5994
VOL. 11, NO. 1
(2026)

nitorare condizioni fisiche, integrare flussi sensoriali o sostenere decisioni conservative, esso appartiene a un paradigma modellistico coerente. Quando, invece, il termine viene applicato a una fotografia digitale, a un file derivato o a una creazione digitale riproducente un artefatto culturale, esso rischia di mascherare la natura documentaria dell'oggetto prodotto. Il *record digitale* non riproduce un originale: lo rappresenta, lo descrive, lo reinterpreta attraverso parametri strutturali e tecnici e lo inserisce in un sistema di relazioni che non ha alcuna coincidenza con l'artefatto fisico (Barbuti, 2024, pp. 77-79).

La risorsa digitale di contenuto culturale, quindi, va interpretata come entità autonoma, connessa al referente analogico quando questo esiste, ma dotata di proprietà specifiche: formato, struttura dati, storia, tecnica, regime di accesso, metadati, vita operativa e riusi. La sua qualità non dipende dalla sola somiglianza visuale con l'originale, ma dalla tracciabilità del processo che l'ha prodotta: acquisizione, colore, illuminazione, post-produzione, denominazione, struttura delle cartelle, metadatazione e modalità di accesso. Ogni scelta incide sulla possibilità di studiare, verificare, preservare e riusare la risorsa. Per le risorse born-digital, in assenza di originali analogici, diventano cruciali strutture, funzionamento, dipendenze software, relazioni tra file, versioni e modalità di fruizione. L'autonomia non va intesa come separazione dall'oggetto rappresentato, ma come riconoscimento del fatto che l'entità digitale possiede una propria storia esistenziale e operativa. Un'immagine master, un derivato di accesso, un file OCR, un pacchetto METS, una *manifest* IIIF o una descrizione RDF non sono copie equivalenti dello stesso bene, ma componenti diverse di un ecosistema documentario, ciascuna con funzioni, vincoli e responsabilità specifiche.

Nella prospettiva GLAM – Galleries, Libraries, Archives and Museums – la responsabilità descrittiva diventa centrale. Ogni istituzione produce un punto di vista sull'oggetto digitale: bibliografico, archivistico, museale, tecnico, giuridico, curatoriale o comunicativo. Un modello dinamico deve consentire la coesistenza di questi livelli, evitando che una scheda unica appiattisca la complessità. Le collezioni digitali possono integrare domini diversi in piattaforme, aggregatori, ontologie, *knowledge graphs* e soluzioni *phygital*, ma questa convergenza non elimina le differenze; piuttosto, obbliga a modellarle e organizzarle come componenti attive di un ecosistema comune (Barbuti & De Bari, 2024a). Le risorse digitali relative a un libro, un fascicolo archivistico, una fotografia, un reperto museale, un oggetto 3D o un archivio born-digital non possono essere trattate come file equivalenti, pur essendo esposte in un unico ambiente virtuale; nondimeno, i contenuti rappresentati possono essere legati tra loro da relazioni non rilevabili negli originali analogici, ma chiaramente evidenziabili nell'ecosistema digitale in cui sono attivi.

In questa prospettiva riteniamo possano essere tracciate le linee per definire il *digital heritage*: esso non coincide con aggregazioni di artefatti analogici riprodotti in record digitali, ma identifica l'ecosistema di relazioni tra risorse digitizzate e born-digital, archivi nativi digitali, pubblicazioni elettroniche, collezioni web, oggetti multimediali, ambienti interattivi, dati di ricerca, narrazioni digitali, contenuti generati dagli utenti e tracce d'interazione in sistemi interoperabili (Bettella, Carrer, & Turetta, 2022). Questa nuova dimensione di esposizione e trasmissione del messaggio culturale impone un ripensamento del concetto stesso di *memoria culturale*: essa non racconta più solo ar-

DigitCult
ISSN 2531-5994
VOL. 11, NO. 1
(2026)

tefatti, ma preserva processi, relazioni, stati, versioni e condizioni di interpretabilità (Beaudoin, 2012; Duranti & Shaffer, 2012).

## 3. Metadati dinamici, provenance, lifecycle e FAIRification

Sebbene la formula del metadato come *dato su un dato* resti utile, essa è insufficiente per collezioni digitali complesse. Il metadato non deve limitarsi a descrivere una risorsa, ma deve renderla trattabile, accessibile, trovabile, intellegibile, interoperabile, preservabile, verificabile e riusabile. La qualità dell'oggetto digitale dipende dalla rete di informazioni che ne rende comprensibili origine, struttura e trasformazioni; per questo il passaggio dal *metadato descrittivo* al *metadato dinamico* è uno dei fondamenti della gestione digitale.

Il *metadato descrittivo* tradizionale risponde a domande di identificazione e accesso che fissano l'informazione in una sequenza predeterminata e poco estendibile: titolo, autore, data, soggetto, collocazione, descrizione fisica. Nelle collezioni digitali, questi elementi devono integrarsi con metadati tecnici, amministrativi, strutturali, di preservazione, di diritti, di processo, di uso e riuso. La risorsa, perciò, diventa comprensibile solo se il record tiene insieme il piano dell'artefatto rappresentato, del file, del processo di produzione, dell'interfaccia, delle relazioni con altre risorse e delle interazioni degli utenti. In questo modo, i metadati diventano *dinamici*, arricchendosi e trasformandosi di pari passo con la vita della risorsa e, contemporaneamente, raccontandola nel tempo in quanto parte costitutiva del *record digitale culturale*. La loro assenza può produrre perdita di significato anche quando l'integrità dei bit è mantenuta, soprattutto in digital libraries caratterizzate da pluralità di tecnologie e di funzioni (Xie & Matusiak, 2016; Biagetti, 2019).

Le informazioni sulla *provenance* costituiscono il nucleo fondante di tale architettura, perché consentono di ricostruire origine, agenti, processi, passaggi, trasformazioni e responsabilità. Nei metadati delle collezioni relative al patrimonio culturale, la *provenance information*, integrando «contextual metadata detailing the origin and history of data», si ricollega alla *data trustworthiness* peculiare ai metadati del patrimonio culturale (Massari et al., 2026, p. i196) e diventa la condizione di autenticità su cui si fonda la sostanza del digitale. Se un sistema conserva solo il suo stato finale, cancella la storia della conoscenza prodotta; se registra gli stati intermedi, abilita audit, revisione e apprendimento.

Il dettaglio della *provenance* diventa, in questo quadro, un criterio di affidabilità. Non basta sapere che una risorsa appartiene a un progetto o a un istituto: occorre distinguere acquisizione, elaborazione, controllo, classificazione, pubblicazione, validazione e riuso. Ciascuna di queste azioni può essere compiuta da agenti diversi e con strumenti differenti; se non sono rappresentate, la risorsa perde la possibilità di essere valutata nella sua genealogia documentaria.

Il *lifecycle* estende questa prospettiva all'intera vita della risorsa. Selezione, acquisizione, preparazione, ripresa o generazione nativa, controllo qualità, master e derivati, metadatazione, classificazione, esposizione, correzione, arricchimento, uso, riuso, mi-

DigitCult
ISSN 2531-5994
VOL. 11, NO. 1
(2026)

grazione e preservazione producono informazioni necessarie alla comprensione futura. Un modello dinamico deve registrare tali eventi come parte strutturale del record. Da qui discende la *responsabilità documentaria*: ogni metadato, classificazione, normalizzazione, inferenza o interazione deve essere riferibile a un agente, umano o computazionale, e a un contesto operativo. Una classe proposta da un modello non ha lo stesso statuto epistemico di una classe validata da un esperto; la distinzione tra proposta automatica, correzione umana, validazione istituzionale, versione superata e stato corrente è quindi un requisito scientifico, non un dettaglio tecnico. La *responsabilità documentaria* è perciò il punto di raccordo tra *lifecycle* e *classificazione*. Essa impone di attribuire ogni passaggio a un soggetto, a una procedura o a un sistema, ma anche di specificarne lo statuto epistemico. Una proposta algoritmica non è falsa perché automatica, né vera perché statisticamente plausibile: è un'informazione generata in condizioni determinate, che deve essere sottoposta a controllo e registrata come tale.

I *Principi FAIR – Findable*, *Accessible*, *Interoperable*, *Reusable* – e la *FAIRification* dei dati sono oggi riferimento progettuale indispensabile per questa trasformazione (Wilkinson et al., 2016). *Findability* richiede identificatori persistenti, metadatazione ricca e connessioni semantiche; *Accessibility* necessita di protocolli stabili e diritti chiari; *Interoperability* richiede standard, vocabolari controllati e scambio tra sistemi; *Reusability* richiede licenze, *provenance*, qualità e documentazione dei metodi. Nei contesti GLAM, la FAIRification non consiste nel correggere a posteriori descrizioni povere, ma nel progettare metadati capaci di sostenere usi e riusi nel tempo, in dialogo con standard e profili quali MAG, METS, Dublin Core, IIIF, TEI, RDF e CIDOC CRM (He, Ma, & Zhang, 2017). In tale prospettiva, i Principi FAIR operano come requisiti di architettura informativa, in quanto non descrivono soltanto la destinazione ideale dei dati, ma orientano la scelta degli identificatori, la stabilità dei protocolli, la qualità dei vocabolari, la leggibilità delle licenze e la possibilità di esportare o interrogare i record in ambienti diversi. Il loro valore cresce quando sono applicati all'intera collezione e non soltanto a singoli oggetti.

La dimensione del *riuso* è decisiva. Un dataset di immagini utilizzato per addestrare un modello di classificazione deve indicare come le immagini sono state prodotte, quali classi sono state attribuite, da chi, con quali criteri e con quali possibili *bias*. Senza questa documentazione, il riuso può moltiplicare errori e generare classificazioni opache e, perciò, inaffidabili. Dunque, i requisiti FAIR devono operare come condizione di sostenibilità conoscitiva, non come etichetta conclusiva giustapposta *a posteriori*. Uno dei principali ostacoli alla FAIRification, infatti, è la separazione tra *metadatazione*, *classificazione* e *preservazione*. Nella gestione digitale, queste funzioni devono essere interdipendenti: la metadatazione definisce i record, la classificazione li rende conoscibili e interrogabili, la preservazione conserva file e condizioni di intellegibilità. Se una risorsa è classificata automaticamente, il processo produce un evento rappresentabile nei metadati; se viene corretta, produce una nuova versione; se viene collegata a un grafo, modifica dinamicamente la rete delle relazioni.

Ne discende che le informazioni sulla classificazione automatica devono anch'esse essere registrate nei metadati come *eventi*: occorre sapere se una classe deriva da algoritmo, regola, catalogatore, vocabolario controllato, validazione o inferenza di grafo,

DigitCult
ISSN 2531-5994
VOL. 11, NO. 1
(2026)

quando è stata attribuita, con quale livello di confidenza e rispetto a quale versione della risorsa. Questa unità di processo rende più concreta la nozione di "gestione dinamica": essa non modifica continuamente record senza controllo, ma permette di sapere perché, quando e da chi un record è stato ridefinito. Il risultato automatico, quindi, deve essere trattato come proposta classificatoria da validare e documentare: l'automazione, infatti, può arricchire il record solo se il suo output è attribuito, verificato e inscritto nel *lifecycle* della risorsa. L'aggiornamento, perciò, diventa parte integrante di attente policy di *preservation by design*, che vanno integrate nel ciclo produttivo delle risorse fin dalla loro generazione, mantenendo la storia delle modifiche disponibile come componente del documento.

## 4. Smart digital libraries, knowledge graph, classificazione automatica e IA per un modello di gestione dinamica

La transizione dalla *biblioteca intelligente* alla *smart digital library* può essere compresa solo dopo aver chiarito il ruolo di metadati dinamici, provenance, lifecycle e Principi FAIR. L'IA non può produrre da sola biblioteche intelligenti, in quanto occorre stabilire quali funzioni possano essere automatizzate, quali debbano restare sotto controllo esperto, quali risultati possano entrare nei record e come vadano documentati gli output generati dai sistemi computazionali. La letteratura sull'*intelligent library* ha mostrato quanto l'adozione non pienamente consapevole di funzionalità di IA nella gestione biblioteconomica impatti su competenze, servizi, infrastrutture e modelli organizzativi (Cox, Pinfield, & Rutter, 2019; Huang, 2024, pp. 887-889). La transizione, pertanto, non è tecnologica in senso stretto: una biblioteca digitale non diventa *smart* per la sola adozione di tecnologie avanzate. Essa riguarda il modo in cui la biblioteca digitale organizza la propria autorità descrittiva quando la parte prevalente di attività e servizi è mediata da sistemi computazionali.

Una *smart digital library* deve rendere espliciti i propri criteri di automazione, distinguere tra raccomandazione, inferenza e validazione, e impedire che la velocità di produzione dei dati ne indebolisca la controllabilità; deve organizzare risorse, processi e responsabilità in modo da sostenere relazioni, inferenze, aggiornamenti controllati, servizi personalizzati e interazioni documentabili. Recenti modelli di *smart digital libraries* sono stati descritti come ambienti tecnologici innovativi che integrano *smart technology*, *smart services*, *smart people*, *smart governance* e *smart places* (Gul & Bano, 2019, p. 764; Yunus, Ismail, & Osman, 2025, pp. 176-177). Nella gestione delle collezioni digitali, tali ambienti acquistano valore se hanno metadati FAIRificati, relazioni esplicite e criteri di valutazione verificabili, altrimenti rischiano di rimanere autoreferenziali.

Alcune esperienze recenti hanno interpretato la *smart digital library* come ambiente in cui integrare e mettere in dialogo linguaggi del Semantic Web, ontologie concettuali e sistemi di riconoscimento intelligente applicati a collezioni digitali, per rendere le risorse interoperabili e interrogabili anche attraverso i contenuti delle immagini e dei

DigitCult
ISSN 2531-5994
VOL. 11, NO. 1
(2026)

metadati. L'interesse di questo orientamento non consiste nella somma di componenti tecnologiche, ma nello spostamento dell'attenzione dai metadati centrati sull'artefatto analogico a una trama narrativa in cui relazioni semantiche, provenance, interazione utente e riconoscimento intelligente diventano elementi di costruzione di *basi di conoscenza*. La *smartness* nasce quando interfaccia, motore semantico, metadati, servizi e pratiche professionali sono progettati insieme (Barbuti, Ferilli, & Caldarola, 2022, p. 1).

In questa linea si collocano anche alcune recenti sperimentazioni *phygital*. I modelli sviluppati non sostituiscono l'oggetto fisico con un doppio digitale, ma generano nuove entità in continuità relazionale tra artefatto analogico, sua descrizione, oggetto digitale, espansioni associate all'originale, dispositivi di accesso e percorsi di interazione elaborati dai fruitori. Nei domini culturali, l'accesso *phygital* può trasformare un manoscritto, una fotografia o un ambiente in elementi di attivazione e accesso a relazioni e contesti, mantenendo riconoscibile il rapporto con la fonte e con il processo che ha generato l'espansione (Barbuti & De Bari, 2024b, pp. 84-88). Ne consegue che la gestione dinamica include non più solo collezioni e ambienti digitali, ma anche oggetti e ambienti fisici quando diventano accessi a percorsi interattivi nei quali le risorse digitali non sono copie, ma estensioni interpretative. L'*unità di cura* diventa allora composita, in quanto artefatto analogico, risorsa, narrazione, metadati, dispositivi, azioni dell'utente e condizioni di preservazione vanno governati come elementi dello stesso processo.

A parere di chi scrive, i *knowledge graphs* sono oggi la forma più coerente per rappresentare questa complessità. Una *base dati* registra *attributi*, un *grafo di conoscenza* rappresenta *relazioni*. Nella digitizzazione del patrimonio culturale, un oggetto digitale rappresenta un originale, appartiene a una collezione, è descritto da un record, classificato secondo uno schema, connesso a un autore, a un luogo, a un evento, a un agente, a una licenza o a un'interpretazione. Un grafo di conoscenza rende interrogabile questa rete senza frammentarla in campi isolati, in quanto le tecnologie del Semantic Web offrono strutture *machine-readable* e facilitano l'integrazione di informazioni granulari provenienti da fonti diverse (Davis & Heravi, 2021, pp. 21-22). Metadati nativamente in *Open Data* o in *Linked Open Data* spostano questo principio a monte del processo, consentendo di progettare record che nascono come entità relazionali e non come schede statiche da trasformare successivamente (Daquino, 2021, pp. 91-94).

Recenti sperimentazioni sull'utilizzo di *graph database* e schemi ontologici in una smart digital library sono orientate su questa logica. L'ontologia non normalizza soltanto nomi o classi: definisce relazioni ammissibili, inferenze possibili, livelli descrittivi da distinguere e regole di coerenza. In un modello di gestione dinamica, il grafo non è il risultato finale di un processo di pubblicazione, ma l'ambiente in cui i record vivono, entrano in relazione, evolvono, si trasformano (Barbuti, Ferilli, & Caldarola, 2022, pp. 3-4). Il crescente rapporto tra *knowledge graphs* e IA consente di vincolare e controllare le inferenze automatiche: una classe proposta da un modello di classificazione può essere verificata rispetto a relazioni note, vincoli cronologici, tipologie documentarie o appartenenze di collezione. Allo stesso tempo, un sistema automatico può suggerire relazioni non ancora registrate, che, se necessario, il curatore valuta, verifica e valida. La dinamicità nasce da questa cooperazione tra inferenza computazionale e validazione semantica.

DigitCult
ISSN 2531-5994
VOL. 11, NO. 1
(2026)

L'integrazione tra *knowledge graphs* e IA richiede, però, una distinzione epistemologica. Da un lato, il grafo corrisponde a un modello architettonico-sistematico dell'organizzazione del sapere: categorie, relazioni e vincoli sono espliciti. Dall'altro, l'IA generativa opera secondo un paradigma statistico-probabilistico, capace di produrre contenuti sulla base di modelli appresi, ma non necessariamente fondato su relazioni dichiarate e verificabili (Roncaglia, 2023). L'integrazione è possibile solo se non si confondono i due statuti: il grafo può vincolare, controllare e rendere esplicite le inferenze; l'IA può suggerire relazioni, classi o aggregazioni da sottoporre a verifica. La cooperazione tra inferenza computazionale e validazione semantica consente di trasformare una classificazione in *nodo-evento*, una validazione in *relazione tra agente*, *classe* e *risorsa*, una correzione in nuova versione del record. Tale precisazione risponde a una difficoltà teorica decisiva: l'organizzazione tramite *knowledge graphs* mira a rendere esplicita una struttura di conoscenza; l'IA generativa, invece, tende a produrre sintesi plausibili a partire da regolarità statistiche. Il loro dialogo può essere fecondo solo se il secondo paradigma non dissolve il primo e la generazione resta vincolata alla verificabilità delle relazioni e alla possibilità di distinguere ciò che è stato descritto, ciò che è stato inferito e ciò che è stato validato.

In questo quadro, l'*AI-assisted cataloguing* rientra come pratica di supporto a descrizione, soggettazione e classificazione automatica, non come sostituzione della responsabilità gestionale biblioteconomica. La necessità di benchmark comparabili tra catalogazione umana e AI-assisted conferma che il problema non è soltanto produrre record, ma misurarne qualità, coerenza e affidabilità (Engel et al., 2025, pp. 261-262). L'uso di BERT per l'attribuzione di *subject headings* mostra che i modelli linguistici possono sostenere la soggettazione, purché siano ricondotti a vocabolari, regole e scelte istituzionali (Chou & Chu, 2022, p. 808). Analogamente, è dimostrato che una classificazione automatica basata sulla Classificazione Decimale funziona meglio quando si appoggia a dati già qualificati da bibliotecari (Kragelj & Kljajić Borštnar, 2021, p. 755).

Lungi dal sostituire la classificazione che conosciamo, l'automazione produce ipotesi, pattern, similarità, etichette e aggregazioni da sottoporre a controllo. La sua funzione più rilevante è far emergere regolarità, anomalie e connessioni difficilmente gestibili su larga scala. Il limite resta chiaro: un modello non conosce da solo il valore documentario delle classi che attribuisce. Esso deve operare in un ciclo *human-in-the-loop*, nel quale contributo computazionale e giudizio professionale restano distinti, registrati e reciprocamente produttivi. La proposta automatica non va cancellata dalla validazione, ma registrata come stato precedente. L'errore non è scarto da eliminare, ma diventa informazione sul comportamento del modello, sulla qualità del dataset e sulla complessità della collezione. In questo modo, il sistema può apprendere dalle correzioni, il curatore può ricostruire il percorso decisionale e l'istituzione può dimostrare la qualità dei propri processi, trasformando in risorsa anche la memoria dell'errore.

È utile distinguere tre forme di classificazione automatica AI-assisted. La prima è semantico-testuale, fondata su NLP, modelli linguistici, *embeddings*, soggettazione automatica e classificazioni bibliografiche. La seconda è visuale e strutturale, basata su *computer vision*, *CNN*, *document image analysis*, *layout analysis* e riconoscimento di

DigitCult
ISSN 2531-5994
VOL. 11, NO. 1
(2026)

parti del libro o del documento. La terza è relazionale, costruita su *knowledge graphs* e inferenze che derivano non solo dal contenuto della risorsa, ma dalla rete in cui essa è inserita. Un modello gestionale dinamico deve integrare queste tre forme senza confonderle, perché hanno statuti diversi e richiedono livelli differenti di validazione (Barbuti & Caldarola, 2026). La distinzione consente anche di evitare una sovrapposizione frequente: non tutte le classificazioni automatiche hanno la stessa funzione. Una classe semantico-testuale può sostenere la soggettazione; una classe visuale può segnalare una struttura materiale o grafica; una classe relazionale può emergere dalla posizione della risorsa nel grafo. Solo la registrazione del metodo che l'ha prodotta consente di valutarne l'uso nel record.

A esemplificazione, una classificazione AI-assisted basata su riconoscimento visuale di immagini può rivelare informazioni implicite nei metadati tradizionali, quali layout, elementi decorativi, supporti, parti del volume, tecniche di stampa, grafie, immagini, pattern ricorrenti. Infatti, il riconoscimento non è mera estrazione di segmenti da traslitterare. Gli attuali sistemi di OCR, HTR e ICR rendono interrogabile ciò che prima era solo visibile, alimentano indici, abilitano ricerche semantiche, collegano parole e immagini e producono nuove possibilità di classificazione. Anche l'estrazione automatica diventa un *evento* del *lifecycle* della risorsa. Nei modelli *phygital*, testo riconosciuto, immagine digitale, record catalografico, riferimento fisico e interfaccia di accesso possono essere collegati nello stesso percorso, generando un sistema di relazioni che rimanda dall'oggetto alla sua espansione computazionale e da questa alla responsabilità descrittiva che la sostiene (Barbuti & De Bari, 2024a). Alcuni studi sul *deep learning* basato su riconoscimento applicato alla classificazione AI-assisted di contenuti storici distinguono compiti di *document classification*, *layout structure* e *content analysis*, insistendo sulla necessità di *ground truth* e metriche comparabili (Lombardi & Marinai, 2020; Aouinti et al., 2022; Nikolaidou et al., 2022, p. 305).

Dato che le classificazioni incidono su accesso, visibilità, ricerca, preservazione e memoria istituzionale, può essere utile anche avvalersi di tecniche di *explainable AI*. Una classificazione errata, infatti, può rendere invisibile una risorsa o collocarla in un contesto improprio. Difficoltà di *accountability* e interpretabilità sono centrali nei processi automatizzati e richiedono sistemi ibridi fondati su conoscenza di dominio, nei quali metriche, visualizzazioni, livelli di confidenza e tracciamento delle decisioni sono componenti della qualità del record (Cox, Pinfield, & Rutter, 2019; Barbuti & Caldarola, 2026, pp. 121-122).

## 5. Conclusioni: sfide e prospettive

Provando a tirare le fila, un *modello di gestione dinamica* delle collezioni digitali può essere delineato come processo teorico-operativo, nel quale risorse eterogenee per genesi, formato e struttura sono descritte, classificate, verificate, validate, preservate e rese accessibili attraverso metadatazione, classificazione automatica, validazione esperta, arricchimento semantico, registrazione della *provenance* e aggiornamento continuo

DigitCult
ISSN 2531-5994
VOL. 11, NO. 1
(2026)

delle informazioni sul *lifecycle*. La sua *unità di cura* granulare è il *record*, inteso come insieme composto da risorsa digitale, metadati, eventi, relazioni e stati di validazione. I livelli in cui il modello può essere articolato sono sette: documentario, digitale, semantico, computazionale, provenienziale, interattivo e valutativo. Questi non vanno intesi come compartimenti separati: il livello documentario stabilisce l'identità culturale della risorsa; quello digitale ne descrive forma, formato e condizioni tecniche; il semantico ne organizza classi e relazioni; il computazionale registra strumenti e modelli impiegati; quello provenienziale attribuisce agenti ed eventi; quello interattivo documenta usi e riusi; il valutativo misura qualità, affidabilità e responsabilità. Il workflow può essere applicato anche in modo modulare, ma deve sempre conservare le informazioni necessarie a ricostruire il percorso che conduce dall'oggetto digitale al record e dal record alla conoscenza.

Il processo generativo include digitizzazione o elaborazione born-digital, creazione dell'oggetto, metadatazione, classificazione automatica preliminare, arricchimento semantico, validazione, esposizione in *smart digital library*, interazione e riuso, aggiornamento e trasferibilità del record e dei *knowledge graphs*.

Il modello si riconduce a tre istanze di riferimento. La prima è che la gestione delle collezioni digitali non può fondarsi su una concezione statica del record. Esso è l'*unità di cura granulare* sia *semplice* che *complessa* quando composta da file, metadati, relazioni, versioni, eventi e contesti. La classificazione automatica è distribuita lungo il ciclo di vita, mentre la preservazione coincide con la capacità di mantenere intellegibili origine, struttura, verifiche, usi, trasformazioni e riusi. L'accessibilità dipende dalla qualità delle relazioni generate dalla FAIRification.

La seconda istanza riguarda i metadati. La distinzione tra digitalizzazione e digitizzazione evidenzia che produrre risorse digitali significa generare entità da governare, preservare e valorizzare nel tempo. I metadati collegano oggetto e processo, dato e record, record e grafo, grafo e conoscenza. Senza informazioni su *provenance* e *lifecycle* nella narrazione dei metadati, le collezioni rischiano di restare aggregati statici di immagini e file; con esse, possono diventare basi di conoscenza evolutive e dinamiche.

La terza istanza riguarda l'IA e i sistemi di riconoscimento intelligente. IA, OCR, HTR, ICR, CNN e NLP sono utili se trattati come contributi al processo descrittivo, non come verità autonome. La loro efficacia dipende da dataset, modelli, contesti, metriche, esplicabilità e validazione. L'IA aiuta senza dubbio ad affrontare scala, complessità e relazioni, ma diventa critica quando tende a sostituire il giudizio esperto o a produrre metadati non tracciabili né verificabili.

La qualità di questo modello risiede nell'integrare in un processo organizzato tradizioni dei domini culturali, computazione e accessibilità multidimensionale, includendo anche soluzioni *phygital* nelle quali fisico e digitale sono rappresentati, interrogati, preservati e trasferiti come componenti di un medesimo ecosistema.

Nondimeno, le sfide da superare sono molteplici. Una prima riguarda i dataset. Manoscritti, libri a stampa, fotografie, mappe, carteggi, oggetti museali e archivi born-digital presentano variabilità di contenuti visuali che rendono difficile l'addestramento di modelli generici. I dataset annotati sono spesso piccoli, sbilanciati, scarsamente do-

DigitCult
ISSN 2531-5994
VOL. 11, NO. 1
(2026)

cumentati o non riusabili. Per il patrimonio GLAM sono necessari dataset coerenti con i Principi FAIR, annotati con vocabolari condivisi, corredati da *provenance* e accompagnati da *ground truth* e *benchmark* comparabili (Nikolaidou et al., 2022, p. 305).

Altra sfida è l'interoperabilità senza appiattimento. Gli standard MAG, METS, Dublin Core, IIIF, TEI, RDF e CIDOC CRM offrono strumenti diversi, da integrare in profili coerenti. Il problema non è scegliere uno standard unico, ma rappresentare i diversi livelli della risorsa: descrittivo, tecnico, strutturale, semantico, provenienziale, conservativo e trasformativo. L'interoperabilità perde valore se cancella la specificità delle tradizioni documentarie, diventa produttiva quando permette a differenze e relazioni di coesistere.

Non meno sfidante delle precedenti è la governance dell'IA. Ogni classificazione automatica deve essere accompagnata da informazioni su modello, dataset, versione, data, parametri, livello di confidenza e stato di validazione. Le istituzioni devono evitare sia il rifiuto dell'IA per timore dell'errore, sia l'adozione acritica per efficienza presunta. La via più solida è una governance trasparente, nella quale l'IA sia usata con funzione di supporto, esplorazione e arricchimento, mantenendo esplicita la responsabilità umana.

Resta, infine, la questione delle competenze. La gestione dinamica richiede figure professionali capaci di dialogare con standard, algoritmi, modelli semantici, policy di preservazione e pluralità di utenti. La formazione umanistica-digitale deve includere metadatazione computazionale, qualità dei dati, *explainability*, diritti, FAIRification e progettazione di workflow. Senza professionalità ibride, le infrastrutture rischiano di essere tecnicamente avanzate ma culturalmente deboli.

La prospettiva più promettente riguarda la costruzione di dataset relativi a GLAM annotati e riusabili, la definizione di profili semantici capaci di collegare MAG, METS, IIIF, TEI, RDF e *knowledge graphs*, lo sviluppo di modelli di IA verificabili e l'adozione di workflow *human-in-the-loop*. La classificazione automatica va progettata come arricchimento controllato e dovrebbe produrre ipotesi e relazioni, ma il record deve conservare provenienza, verifica, validazione e stato della conoscenza. Anche le soluzioni *phygital* vanno valutate secondo questo criterio: non come spettacolarizzazione dell'accesso, ma come estensione documentata dell'artefatto culturale originale, capace di collegare presenza fisica, risorsa digitale, metadati, interazione e responsabilità descrittiva.

In questa prospettiva, la *smart digital library* del prossimo futuro sarebbe non più solo un'interfaccia efficiente di consultazione, ma un ecosistema capace di aggiornare descrizioni, relazioni e servizi in modo documentato. Ogni tecnologia sarebbe valutata rispetto alla sua capacità di migliorare intellegibilità, accesso, preservazione, riuso e trasferibilità. La domanda cui rispondere, dunque, non è più quanto un'istituzione possa automatizzare, ma quali decisioni possa rendere più tracciabili, quali relazioni possa esplicitare e quale memoria culturale del digitale possa consegnare al futuro. La sfida non è digitizzare di più o automatizzare più rapidamente, ma trasformare il dato in record, il record in relazione, la relazione in messaggio, il messaggio in conoscenza e la conoscenza in memoria digitale affidabile, preservabile, intellegibile, riusabile e trasferibile.

DigitCult
ISSN 2531-5994
VOL. 11, NO. 1
(2026)

# Bibliografia

Aouinti, F., Eyharabide, V., Fresquet, X., & Billiet, F. (2022). Illumination detection in IIIF medieval manuscripts using deep learning. *Digital Medievalist*, 15(1), 1-18.

Barbuti, N. (2022). *La digitalizzazione dei beni documentali. Metodi, tecniche, buone prassi*. Editrice Bibliografica.

Barbuti, N. (2024). Non solo immagini. Collezioni digitali in cerca d'autore nella biblioteca di Babele. In M. Gatto, A. Squeo, & S. Silvestri (a cura di), *Informatica umanistica, Digital Humanities: verso quale modernità?* (pp. 77-94). Cacucci.

Barbuti, N., & Caldarola, T. (2026). A dynamic system for the automatic classification of complex resources in digital libraries: Design and preliminary evaluation. *Digital Library Perspectives*, 42(1), 119-140.

Barbuti, N., & De Bari, M. (2024a). Travelling to new knowledge perspectives. Digital expansions for the interactive fruition of books and libraries. In A. Carpallo Bautista & M. Salamanca López (a cura di), *El manuscrito medieval: del pergamino al metadato* (pp. 269-290). Editorial Dykinson.

Barbuti, N., & De Bari, M. (2024b). Libri e biblioteche tra museabilità e musealizzazione digitale: sogno o realtà? In A. Di Silvestro & D. Spampinato (a cura di), *Proceedings del XIII Convegno Annuale AIUCD. Me.Te. Digitali. Mediterraneo in rete tra testi e contesti* (pp. 84-88). Università di Catania.

Barbuti, N., Ferilli, S., & Caldarola, T. (2022). A pilot of smart digital library user-centered: The Project SMARTER. In G. M. Di Nunzio, B. Portelli, D. Redavid, & G. Silvello (Eds.), *Proceedings of the 18th Italian Research Conference on Digital Libraries* (CEUR Workshop Proceedings, Vol. 3160).

Beaudoin, J. E. (2012). Context and its role in the digital preservation of cultural objects. *D-Lib Magazine*, 18(11/12).

Bettella, C., Carrer, Y., & Turetta, G. (2022). La valutazione FAIRness di un archivio digitale certificato: tra principi teorici e azioni pratiche. *DigItalia*, 17(1), 113-133.

Biagetti, M. T. (2019). *Le biblioteche digitali. Tecnologie, funzionalità e modelli di sviluppo*. FrancoAngeli.

Bloomberg, J. (2018, 29 aprile). Digitization, digitalization, and digital transformation: Confuse them at your peril. *Forbes*. https://www.forbes.com/sites/jasonbloomberg/2018/04/29/digitization-digitalization-and-digital-transformation-confuse-them-at-your-peril/

Brennen, S., & Kreiss, D. (2014, 8 settembre). Digitalization and digitization. *Culture Digitally*. https://culturedigitally.org/2014/09/digitalization-and-digitization/

Chou, C., & Chu, T. (2022). An analysis of BERT (NLP) for assisted subject indexing for Project Gutenberg. *Cataloging & Classification Quarterly*, 60(8), 807-835.

CIDOC CRM. (s.d.). *Conceptual Reference Model*. https://www.cidoc-crm.org/

Cox, A. M., & Mazumdar, S. (2024). Defining artificial intelligence for librarians. *Journal of Librarianship and Information Science*, 56(2), 330-340.

Cox, A. M., Pinfield, S., & Rutter, S. (2019). The intelligent library: Thought leaders' views on the likely impact of artificial intelligence on academic libraries. *Library Hi Tech*, 37(3), 418-435.

DigitCult
ISSN 2531-5994
VOL. 11, NO. 1
(2026)

Dagnaw, G. A., Capuano, R., & Muccini, H. (2026). Digital twins for cultural heritage: A systematic analysis of the state of the art. *ACM Computing Surveys*, 58(9), Article 237, 1-35.

Daquino, M. (2021). Linked Open Data native cataloguing and archival description. *JLIS.it*, 12(3), 91-104.

Davis, E., & Heravi, B. (2021). Linked data and cultural heritage: A systematic review of participation, collaboration, and motivation. *ACM Journal on Computing and Cultural Heritage*, 14(2), Article 21.

Duranti, L., & Shaffer, E. (Eds.). (2012). *The memory of the world in the digital age: Digitization and preservation*. UNESCO.

Engel, J. Y., Do, D. T., Salem, B., & Cunningham, J. (2025). Artificial intelligence in library cataloging: A review of literature. *Journal of Library Metadata*, 25(4), 261-276.

Gul, S., & Bano, S. (2019). Smart libraries: An emerging and innovative technological habitat of 21st century. *The Electronic Library*, 37(5), 764-783.

He, Y., Ma, Y. H., & Zhang, X. R. (2017). "Digital heritage": Theory and innovative practice. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W5, 335-342.

Huang, Y. H. (2024). Exploring the implementation of artificial intelligence applications among academic libraries in Taiwan. *Library Hi Tech*, 42(3), 885-905.

International Image Interoperability Framework. (s.d.). *IIIF*. https://iiif.io/

Kragelj, M., & Kljajić Borštnar, M. (2021). Automatic classification of older electronic texts into the Universal Decimal Classification – UDC. *Journal of Documentation*, 77(3), 755-776.

Lombardi, F., & Marinai, S. (2020). Deep learning for historical document analysis and recognition: A survey. *Journal of Imaging*, 6(10), Article 110.

Massari, A., Peroni, S., Tomasi, F., & Heibi, I. (2026). Representing provenance and track changes of cultural heritage metadata in RDF: A survey of existing approaches. *Digital Scholarship in the Humanities*, 41(Supplement 1), i196-i213.

Nikolaidou, K., Seuret, M., Mokayed, H., & Liwicki, M. (2022). A survey of historical document image datasets. *International Journal on Document Analysis and Recognition*, 25, 305-338.

Roncaglia, G. (2023). *L'architetto e l'oracolo. Forme digitali del sapere da Wikipedia a ChatGPT*. Laterza.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, Article 160018.

Xie, I., & Matusiak, K. K. (2016). *Discover digital libraries: Theory and practice*. Elsevier.

Yunus, N., Ismail, M. N., & Osman, G. (2025). Smart library themes and elements: A systematic literature review. *Journal of Librarianship and Information Science*, 57(1), 175-190.