

Dig *Italia*

Anno XV, Numero 2 - **2020**

Rivista del digitale nei beni culturali

ICCU-ROMA



ICCU

Istituto centrale per il catalogo unico
delle biblioteche italiane e per le informazioni bibliografiche
www.iccu.sbn.it

Copyright © ICCU - Roma

La riproduzione totale o parziale del contenuto della rivista
è ammessa con obbligo di citazione

«*Digitalia*», rivista del digitale nei beni culturali, è una rivista *peer reviewed*
e segue il codice etico delle pubblicazioni

I contributi possono essere proposti alla rivista dagli autori o su iniziativa del Comitato Scientifico. Gli articoli vengono sottoposti al Comitato di Redazione per un primo accertamento sulla corrispondenza con i campi di ricerca della rivista. I contributi delle sezioni Saggi e Progetti vengono indirizzati in forma anonima ad almeno uno studioso di comprovata competenza sui temi affrontati. I revisori fanno pervenire i loro giudizi alla redazione (favorevole alla pubblicazione, favorevole con modifiche/miglioramenti, non favorevole). Se il giudizio finale è positivo, viene comunicata agli autori l'accettazione del contributo, insieme ad eventuali indicazioni suggerite dai valutatori, di cui si garantisce comunque l'anonimato.

Digitalia

Rivista del digitale nei beni culturali
ISSN 1972-6201
Anno XV, Numero 2 - Dicembre 2020

In copertina:

L'immagine è una libera elaborazione grafica della testa della statua di Apollo del I sec. d.c. (Civitavecchia, Museo Nazionale), copia da un originale greco avvicicabile all'Apollo di Leochares (IV sec. a.c.)

Direttore Fondatore

Marco Paoli

Direttore Responsabile

Simonetta Buttò

Comitato di Redazione

Capo Redattore:

Elisabetta Caldelli

Amalia Maria Amendola
Valentina Atturo
Lucia Basile
Laura Borsi
Flavia Bruni
Elisabetta Castro
Massimina Cattari
Silvana de Capua
Carla Di Loreto
Maria Cristina Di Martino
Vilma Gidaro
Egidio Incelli
Maria Cristina Mataloni
Massimo Menna
Lucia Negrini
Paola Puglisi
Alice Semboloni
Vittoria Tola
Maria Lucia Violo

Grafica & Impaginazione

MLA&Partner - Roberta Micchi

Produzione e Stampa

Istituto Poligrafico e Zecca dello Stato S.p.A.
Roma

Editore

ICCU

Istituto centrale per il catalogo unico
delle biblioteche italiane
e per le informazioni bibliografiche
Viale Castro Pretorio, 105
00185 Roma
T +39 06 49.210.425
F +39 06 49.59.302
email: ic-cu.digitalia@beniculturali.it
<http://digitalia.sbn.it>

In attesa di registrazione al Tribunale di Roma



Comitato Scientifico

Osvaldo Avallone
Giovanni Bergamin
Dimitri Brunetti
Simonetta Buttò
Rossella Caffo
Rosaria Campioni
Maria Carla Cavagnis Sotgiu
Laura Ciancio
Flavia Cristiano
Gianfranco Crupi
Andrea De Pasquale
Maria Cristina Di Martino
Pierluigi Feliciati
Marina Giannetto
Maria Guercio
Mauro Guerrini
Klaus Kempf
Maurizio Messina

Maria Cristina Misiti
Maria Teresa Natale
Marco Paoli
Don Valerio Pennasso
Alberto Petrucciani
Massimo Pistacchi
Marco Pizzo
Paola Puglisi
Roberto Raieli
Gino Roncaglia
Maria Letizia Sebastiani
Giovanni Solimine
Laura Tallandini
Anna Maria Tammaro
Costantino Thanos
Antonella Trombone
Paul Gabriele Weston

SOMMARIO

dicembre 2020

SAGGI

- Alphabetic**, il nuovo portale per la ricerca integrata:
un salto di qualità per le biblioteche italiane 9
di Simonetta Buttò
- Sistema di ricerca integrato: un nuovo catalogo
di servizi per le biblioteche** 16
di Luigi Cerullo, Maria Cristina Mataloni
- Non solo ebook:
la biblioteca digitale per l'e-learning** 26
di Anna Maria Tammaro
- Dal Catalogo generale dei beni culturali
al knowledge graph del patrimonio
culturale italiano: il progetto ArCo** 43
di Chiara Veninata

PROGETTI

- Il progetto inDICES: misurare
l'impatto della cultura digitale** 59
di Sara Di Giorgio, Claudio Prandoni
- New Approaches to OCR
for Early Printed Books** 74
di Nikolaus Weichselbaumer, Mathias Seuret, Saskia Limbach,
Rui Dong, Manuel Burghardt, Vincent Christlein
- ARIADNEplus: l'avventura continua** 88
di Franco Niccolucci
- L'adozione del IIF nell'ecosistema digitale
della Biblioteca Apostolica Vaticana** 96
di Paola Manoni
- I-Tal-Ya books: il censimento digitale
dei libri ebraici d'Italia, un progetto di cooperazione
dell'Unione delle comunità ebraiche italiane,
della Biblioteca nazionale centrale di Roma
e della Biblioteca nazionale di Israele** 106
di Andrea De Pasquale
- La tecnologia 5G e i beni culturali:
percorsi di storie e architetture all'Aquila** 117
di Silvia Mantini, Fabio Graziosi,
Fabio Franchi, Stefano Boero

**EAGLE: storia di un'idea
dalle origini all'ingresso di EDF** 126
di Maria Letizia Caldelli, Silvia Orlandi

**Iste liber est... on-line.
Una rassegna di strumenti in rete per lo studio
e il censimento dei provenance marks** 138
di Lucrezia Signorello

**IDP - Illuminated Dante Project:
un archivio e database per la più antica
iconografia dantesca (secc. XIV-XV)** 150
di Ciro Perna

**Proposta per una valutazione automatica
della completeness dei metadati
nel contesto delle biblioteche digitali** 159
di Matteo Lorenzini, Marco Rospocher, Sara Tonelli

DOCUMENTI E DISCUSSIONI

**L'evoluzione del ruolo delle Biblioteche
Digitali Scientifiche (BDS)** 171
di Costantino Thanos

**33 million objects and counting:
a conversation about the Deutsche Digitale Bibliothek** 174
di Cosmina Berta, Gerke Dunkhase

**Evaluating a Semantic Portal for the
"Mapping Manuscript Migrations" Project** 178
di Toby Burrows, Nicole Bergk Pinto, Mahaut Cazals,
Alexandre Gaudin, Hanno Wijsman

EVENTI

Archivissima 2020 – Gli archivi in digitale 189
di Ilaria Scalmani

SEGNALAZIONI

**Un assaggio di Leopardi.
Il progetto BDL – Biblioteca Digitale Leopardiana** 193
di Gioele Marozzi

Saggi

Alphabeticata, il nuovo portale per la ricerca integrata: un salto di qualità per le biblioteche italiane

«DigItalia» 2-2020
DOI: 10.36181/digitalia-00010

Simonetta Buttò - *Direttore ICCU*

Alphabeticata, il nuovo portale progettato a cura dell'Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU), si avvia verso la fase finale di realizzazione. Questa nuova piattaforma permetterà di accedere alle banche dati gestite dall'ICCU e ai progetti ad esso afferenti tramite un unico sistema di navigazione e ricerca, valorizzando il lavoro di cooperazione ultratrentennale svolto dalla comunità delle biblioteche italiane nel Servizio Bibliotecario Nazionale e le sue numerose risorse digitali.

Il progetto per la realizzazione di un sistema integrato che consente di interrogare contemporaneamente tutte le basi dati gestite dall'ICCU, presentato anche sulle pagine di questa rivista al momento del suo lancio¹, si avvia all'ultimo giro di boa. Frutto di una lunga attività di studio e analisi, il progetto mira a realizzare l'integrazione fra le basi dati nazionali gestite dall'ICCU in un unico "catalogo generale" collegato alle risorse digitali presenti sulla Digital Library di Internet Culturale, al quale si potrà accedere tramite un portale di nuova concezione, *Alphabeticata*, pensato come piattaforma completa di recupero delle informazioni provenienti da database diversi, nati in tempi diversi, per fini diversi, usando soluzioni tecnologiche diverse.

Il lavoro certosino di reingegnerizzazione di questi sistemi, al fine di renderli interrogabili in simultanea, ha messo al centro della nuova architettura dell'informazione l'Indice di SBN, l'insieme di dati, cioè, che rappresenta il risultato tangibile del lavoro partecipato, ultratrentennale, svolto dalla comunità delle biblioteche italiane.

Sarà dunque l'Indice a funzionare sia come catalizzatore delle richieste degli utenti che come attrattore delle informazioni provenienti dagli altri database a partire da liste di autorità controllate e voci di soggetto.

Se infatti i metodi tradizionali di consultazione dei singoli cataloghi (ricerca sem-

¹ *Il Portale delle biblioteche e degli istituti culturali italiani: presentazione del progetto. Roma 11 aprile 2019, Sala Spadolini, MiBACT, «DigItalia. Rivista del digitale nei beni culturali», 14 (2019), n. 1, p. 9-28.*

plice e ricerca avanzata) saranno comunque sempre disponibili, chi sceglierà di esplorare il patrimonio culturale navigando nel nuovo portale avrà a disposizione chiavi di ricerca mirate a definire contesti (protagonisti, biblioteche, luoghi, ...), oppure dedicate ad alcune tipologie di materiali (libri, periodici, manoscritti, grafica, cartografia, musica, audiovisivi) che consentono di ricevere risposte esaurienti, provenienti da SBN e dalle basi dati specialistiche del *Censimento del libro italiano del XVI secolo* - EDIT16 e del *Censimento dei manoscritti delle biblioteche italiane* - ManusOnLine, integrate con le risorse digitali attualmente presenti nella biblioteca digitale di Internet culturale.

Il sistema si gioverà anche di due banche dati afferenti, entrambe gestite dall'ICCU, il portale tematico *14-18: Documenti e immagini dalla Grande Guerra*, un grande archivio di immagini che accoglie fonti documentarie e memorialistiche di varia natura, e il sito MOVIO, che conserva le mostre virtuali realizzate da numerose istituzioni culturali ed educative.

L'obiettivo del progetto che si avvia a conclusione è quello di offrire a tutti, in Italia e all'estero, un punto di riferimento innovativo ed efficace, tecnologicamente avanzato ma anche sostenibile, per conoscere, mettere in comune e utilizzare tutto il lavoro che è stato realizzato negli ultimi decenni dalle biblioteche italiane, sia per l'accrescimento e il miglioramento dei cataloghi, sia per la produzione di risorse digitali da mettere a disposizione degli utenti.

Alla base del sistema c'è dunque una visione complessiva del mondo delle biblioteche, che è un insieme complesso e interistituzionale, nel quale convivono realtà (e dunque porzioni di pubblico) molto differenti fra loro, tutte però ugualmente importanti, anzi indispensabili, per attuare una politica culturale aperta e di ampio respiro. Al contrario dei grandi motori di ricerca, Google per primo, le biblioteche non hanno il compito di arrivare a rappresentare per il pubblico il principale canale informativo o il principale canale per la circolazione del prodotto-libro.

Le biblioteche semmai sono state, e sono anche oggi, il principale canale per lo studio, a tutte le età, e il posto dove trovare libri e riviste che non è possibile, o sostenibile, acquistare e tenere in casa, grazie ai quali soddisfare le più diverse esigenze informative e conoscitive, dalla semplice curiosità fino alla ricerca scientifica, presenti in tutti gli strati della popolazione.

Nella nostra storia recente, caratterizzata da processi di disintermediazione sempre più rapidi, le biblioteche, luoghi di raccolta, conservazione e tutela dei documenti, si sono trovate di fronte alla necessità di definire uno spazio proprio – e significativo per la propria comunità di riferimento – per fornire, con i propri mezzi, anche risposte concrete a esigenze concrete di informazione, sulla memoria storica del territorio e sulla sua rappresentazione, ma anche sull'attualità, sui grandi temi che pervadono la società di oggi, come l'emergere di nuove forme di povertà, la comprensione dei flussi migratori, la contrazione delle politiche di welfare culturale, e – non da ultimo – anche per dare un contributo fattivo al processo in atto di trasformazione della

fruizione dei contenuti culturali e di lotta al *digital divide*, che rappresenta la nuova frontiera in tutte le società, anche le più evolute, del nostro tempo².

In un contesto dominato da una dimensione “orizzontale”³ delle pratiche culturali, tipica della nostra epoca, lo spazio proprio delle biblioteche, da sempre caratterizzato da una solida cultura dell’accesso, si estende sempre di più anche nella direzione della produzione di contenuti culturali di qualità elevata, in sinergia con le università, le scuole, gli istituti e le associazioni presenti sui territori, nella valorizzazione delle proprie attività e nella creazione di nuove forme di fruizione.

Oggi si osserva da più parti che la complessità verticale (e dunque la capacità di approfondire), tipica della cultura del libro, sia stata decisamente archiviata dalla complessità tutta orizzontale della rete⁴ in un *hic et nunc* sterminato, un grande adesso, che pone sullo stesso piano informazioni verificate e *fake news*, documenti storici e opinioni di singoli, immagini isolate e immagini contestualizzate, favorendo subdolamente l’innalzamento del livello di passività che caratterizza molte forme di consultazione della rete.

Invece uno degli obiettivi più qualificanti per chi opera nelle biblioteche e gestisce un patrimonio immenso, reale e virtuale, sta proprio nel mettere in campo strategie e strumenti mirati, per restituire al pubblico, sia pure in forme nuove, contenuti strutturati e complessi, in grado di stimolare lo sviluppo del pensiero, la riflessione critica, e accrescere le capacità e le competenze, la formazione personale e professionale di ciascuno di noi⁵.

² Il 23 settembre 2020 il Parlamento italiano ha approvato in via definitiva la ratifica della *Convenzione quadro del Consiglio d’Europa sul valore del patrimonio culturale per la società*, presentata dal Consiglio d’Europa a Faro il 27 ottobre 2005 e fondata sul presupposto che la conoscenza e l’uso dell’eredità culturale rientrano pienamente fra i diritti umani. In particolare (cfr. l’art. 13, *Eredità culturale e conoscenza*) la Convenzione di Faro promuove la comprensione del patrimonio culturale in rapporto alle comunità che lo hanno prodotto (cfr. art. 14, *Eredità culturale e società dell’informazione*) e invita i paesi firmatari a sviluppare l’uso delle tecnologie digitali per migliorare l’accesso all’eredità culturale, abbattendo gli ostacoli che ne limitano la fruizione. Per il testo completo della Convenzione cfr.: <https://www.beniculturali.it/mibac/multimedia/MiBAC/documents/1492082511615_Convenzione_di_Faro.pdf>.

³ Sul tema, cfr. il recente volume di Giovanni Solimine e Giorgio Zanchini, *La cultura orizzontale*, Roma-Bari: Laterza, 2020.

⁴ Cfr. Gino Roncaglia, *L’età della frammentazione: cultura del libro e scuola digitale*, Bari-Roma: Laterza, 2018, in particolare p. 196-197.

⁵ Cfr. anche il *Nuovo Manifesto per le biblioteche digitali* pubblicato dall’AIB e attualmente in bozza, che significativamente collega il tema della diffusione della conoscenza e della formazione permanente all’azione delle biblioteche sul territorio (Principio n. 7, *Le biblioteche digitali interagiscono con le proprie comunità*): «Le biblioteche digitali basano il loro rapporto con gli utenti sul metodo dell’interazione, favoriscono la partecipazione attiva degli utenti alla propria comunità, anche attraverso la creazione di nuove risorse, e cercano di parlare il loro linguaggio. Supportano inoltre l’attivazione di percorsi dedicati allo sviluppo della competenza informativa degli utenti stessi, intesa come capacità di ricerca indipendente, di analisi, valutazione critica, riutilizzo efficace delle diverse fonti informative, e di comprensione di come l’informazione è prodotta e valutata, anche al fine di produrre nuova conoscenza», <<https://www.aib.it/struttura/commissioni-e-gruppi/gruppo-di-lavoro-biblioteche-digitali/2020/82764-nuovo-manifesto-per-le-biblioteche-digitali/>>.

Questo tipo di “lavoro culturale”, che le biblioteche svolgono al fianco di altre istituzioni culturali, non rappresenta solo una operazione di *advocacy*, ma consente anche il migliore contatto con i bisogni delle comunità, per ampliare gli inter-

essi e collegarli a quelli espressi da altri protagonisti della vita culturale e sociale del territorio, estendendo la rete delle “relazioni sociali” delle biblioteche.

Attraverso il nuovo Portale *Alphabetic*a le biblioteche italiane che partecipano alla crescita, qualitativa e quantitativa, dei servizi informativi nazionali, avranno una vetrina tutta nuova per presentare al pubblico i documenti che possiedono, i servizi che offrono e la cultura che producono, mentre il pubblico potrà scoprire un altro modo per porre domande, altre vie per scoprire il patrimonio, incrociando dati provenienti da domini disciplinari differenti, e nuovi percorsi per ampliare le proprie ricerche.

In questo contesto di integrazione e di servizio, anche le biblioteche cosiddette “di conservazione”, sia statali che degli enti locali, che registrano in questi ultimi decenni un pericoloso slittamento in direzione di una progressiva musealizzazione, sia delle loro sedi, sia – ed è più grave – delle loro funzioni, potranno dare un contributo di elevata qualità, fon-



Figura 1. Prototipo per la presentazione dei risultati di ricerca

dato sul valore di manoscritti, prime edizioni, carteggi, materiali archivistici, repertori che rappresentano le indispensabili fonti primarie con le quali confrontare, e talvolta confutare, la dilagante letteratura secondaria oggi presente sia nei formati tradizionali che in rete.

Il portale *Alphabetic*a presenterà i risultati in maniera ordinata, rispettando le preferenze espresse, limitando drasticamente sia il “rumore” tipico dei motori di ricerca largamente utilizzati per l’interrogazione simultanea, sia la casualità della loro disposizione.

Anche le risorse digitali collegate ai risultati della ricerca emergeranno in maniera chiara e trasparente direttamente dalla loro relazione con i dati di qualità rappresentati dai record del catalogo, riducendo la dispersione granulare e consentendo all’utente di non perdere di vista il contesto di riferimento.

Sul nostro territorio, che accoglie mille realtà differenti, è oggi disponibile infatti una grande quantità di risorse e contenuti digitali, estremamente utili per la valorizzazione del nostro patrimonio culturale e – come si è detto – anche per la didattica, l’apprendimento, la formazione professionale, come pure per l’accesso alla documentazione di fonte pubblica, che richiede però una gestione consapevole e finalizzata per essere utilizzabile nel tempo da diverse comunità di ricerca e gruppi di utenti e servire a produrre conoscenza.

Purtroppo il proliferare di basi dati molto specialistiche, realizzate da università ed enti di ricerca, di *repository* di interesse locale, di teche digitali di piccole dimensioni, alcuni realizzati utilizzando software proprietari, senza attenzione per il dialogo con sistemi analoghi e per l’interoperabilità con i servizi nazionali, altri resi invisibili dall’occultamento in siti web scarsamente conosciuti, o instabili, ha contrassegnato negativamente la loro sostenibilità nel tempo: una teca digitale priva di una utenza significativa diventa immediatamente del tutto irrilevante nella vastità della rete, e in questo modo progetti anche di grande interesse e utilità per il mondo della ricerca sono annegati nel vasto mare del web, perché poveri di utenti, o per mancanza di manutenzione, alimentazione, sviluppo, contribuendo a una dispersione di risorse informative e impedendo di fatto il libero accesso ai contenuti.

È dunque diventato un dovere, in una economia di scala, recuperare tali risorse attraverso una campagna mirata di sensibilizzazione all’interno della comunità delle biblioteche e migliorarne l’accessibilità condividendo, in una logica di integrazione e di servizio, una infrastruttura di accesso ai contenuti aperta alla pubblicazione di risorse digitali culturali di qualità, provenienti dalla cooperazione SBN, ma anche realizzate al di fuori di essa, attraverso l’adozione di procedure controllate per la contemporanea interrogazione di sistemi diversi.

Il modello concettuale che è alla base del progetto è pensato per il futuro ed è

dunque flessibile per sua definizione⁶: sono infatti previsti collegamenti complementari, da realizzarsi attraverso convenzioni mirate con altri enti e istituzioni culturali, che per le loro finalità istituzionali hanno prodotto e diffuso contenuti digitali di alto valore culturale, in qualità di gestori di catalogo, aggregatori di primo livello, *content provider*.

La flessibilità dell'architettura informativa che sta alla base del progetto rappresenta una parte importante della sua sostenibilità nel futuro, un modo di mettere in contatto quanti, nei decenni scorsi, soprattutto nel settore pubblico, hanno sviluppato idee innovative e costruito sistemi avanzati per la conoscenza e la valorizzazione del patrimonio, fornendo uno stimolo per lavorare sempre di più insieme, senza steccati e senza disperdere preziose risorse ed energie.

L'ampia disponibilità di risorse relative al patrimonio culturale presenti nella comunità di SBN, ma anche al di fuori di essa, impone infine l'adozione di strumenti di visualizzazione e restituzione di contenuti in grado di utilizzarle al meglio e di porle in relazione tra loro, ampliando significativamente i contesti culturali.

A garanzia della interoperabilità fra le collezioni e di una modalità di fruizione avanzata delle immagini il portale *Alphabetic* sarà dotato del protocollo IIIF-International Image Interoperability Framework, uno degli strumenti di ultima generazione in grado di arricchire i contenuti grazie a un set di API- Application Programming Interface che fornisce funzionalità avanzate di visualizzazione, condivisione, annotazione.

Questo progetto non si sarebbe potuto avviare se nel corso degli ultimi quindici anni, l'ICCU non avesse sviluppato una linea di attività interamente dedicata alla partecipazione alle reti europee per il *Cultural Digital Heritage*, le *Digital humanities* e le *e-infrastructures*, le piattaforme digitali per la ricerca, che costituiscono oggi un fattore unificante di considerevole portata, una grande opportunità contro la frammentarietà esistente e un esempio di condivisione e reciproco arricchimento.

È in questo contesto che l'Istituto ha compiuto passi fondamentali per migliorare la reperibilità delle risorse, armonizzare le politiche di gestione dei dati e adottare infrastrutture digitali in grado di interoperare con altre infrastrutture attraverso l'adozione di procedure comuni per l'integrazione dei sistemi, adottando nuovi flussi di lavoro, sviluppando nuove competenze.

Queste esperienze costituiscono il presupposto, logico e pragmatico, per la costruzione di reti solide tra istituzioni, enti di ricerca, università, dedicate alla cura di dati trasversali a molte e diverse discipline, e per ampliare le opportunità d'uso del patrimonio culturale digitale da parte degli utenti, per comprenderne meglio le esigenze e migliorare i servizi offerti.

⁶ Cfr. anche *Nuovo Manifesto per le biblioteche digitali* cit., Principio n. 18 (*Le biblioteche digitali hanno modelli flessibili*): «I modelli di biblioteche digitali devono essere flessibili, aggiornabili in funzione delle innovazioni tecnologiche e aperti alle sinergie».

Un sistema cooperativo basato sulla partecipazione per offrire servizi, dunque fondato esattamente sugli stessi presupposti che l'ICCU ha adottato a livello nazionale oltre trenta anni fa dando vita al Servizio bibliotecario nazionale.

Alphabetic, the new portal designed by the Central Institute for the Union Catalogue of Italian Libraries and Bibliographic Information (ICCU), is about to be completed. This platform will allow access to the several databases managed by ICCU and its related projects through a single query, promoting the work carried out by the community of Italian libraries in over thirty years of cooperation within the National Library Service and its numerous digital resources.

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2020

Sistema di ricerca integrato: un nuovo catalogo di servizi per le biblioteche

«DigItalia» 2-2020
DOI: 10.36181/digitalia-00011

Luigi Cerullo — Maria Cristina Mataloni

ICCU

Il Sistema di Ricerca Integrato (SRI) costituisce un articolato progetto di integrazione delle basi dati gestite dall'ICCU e al contempo mira ad offrire all'utente finale un servizio di restituzione originale, destinato a soddisfare le esigenze di un pubblico più vasto, costituito non solo da specialisti del settore.

All'interno di questo progetto se ne innestano altri complementari, SBNteca e SBNcloud, che hanno lo scopo di offrire alla comunità SBN strumenti tecnologicamente avanzati e di sicuro ausilio rispetto alla crescente domanda di servizi digitali, mai come oggi fondamentali per gli utenti finali.

Il modello di integrazione

L'idea di fondo da cui muove il progetto¹ nasce dall'esigenza di costruire un modello di integrazione dei dati di catalogo che consenta di superare le criticità legate alla presenza di piattaforme di ricerca e restituzione separate per ciascuno dei principali servizi informativi che l'ICCU garantisce (dal catalogo collettivo del Servizio Bibliotecario Nazionale alle banche dati specialistiche rappresentate dai censimenti nazionali, dall'aggregatore dedicato alle risorse digitali alle altre banche dati di corredo); al contempo si intende "ricucire lo strappo" determinatosi tra le informazioni catalografiche e le risorse digitali raccolte dall'aggregatore Internet Culturale.

Integrazione dell'informazione

In estrema sintesi l'integrazione è concepita come correlazione e sistematico riferimento a un oggetto informativo primario, rappresentato dal record bibliografico del Catalogo collettivo SBN, arricchito dal legame con la copia digitale presente nel sistema aggregatore Internet Culturale. Il dataset dell'Indice SBN, alla base del Catalogo collettivo, viene integrato sia da *collegamenti* a schede relative allo stes-

¹ Per un approfondimento di natura più tecnica, cfr. Luigi Cerullo, *ISS Project: The Integrated Search System in the National Bibliographic Services*, in: *Digital Libraries and Multimedia Archives. 14th Italian Research Conference on Digital Libraries, IRCDL 2018, Udine, Italy, January 25-26, 2018*, ed. by G. Serra, C. Tasso, Cham: Springer, 2018, (Communications in Computer and Information Science; 806), p 219-224.

so documento, presenti nel repertorio specialistico EDIT16² dedicato al libro italiano del 16° secolo, sia da schede di manoscritti provenienti dalla banca dati ManusOnLine³ (MOL), una tipologia documentaria non gestita nel Catalogo collettivo. In quest'ultimo caso ad essere recuperate sono le descrizioni interne dei manoscritti il cui profilo descrittivo meglio si concilia con il modello di restituzione dati del Catalogo collettivo.

Accanto all'integrazione dell'informazione riferibile ai "documenti", il progetto ha previsto lo sviluppo di strumenti e procedure finalizzate a costruire un *cluster* di record di authority per i nomi di persone ed enti collegati ai documenti. Anche in questo caso il modello di clusterizzazione fa perno sulla individuazione di una scheda informativa di riferimento, rappresentata parimenti dal record SBN che si arricchisce in tal modo di collegamenti ai repertori specialistici che continuano, dal canto loro, a conservare una propria rappresentazione della entità. Qualora l'entità non sia presente nel sistema di riferimento, la procedura si fa carico di integrare l'archivio Nomi di SBN con i nuovi record.

Una volta costituito, tale authority integrato continuerà ad essere alimentato attraverso l'ordinaria attività di catalogazione svolta nelle tre banche dati coinvolte (Indice SBN, EDIT16 e MOL), grazie alla "integrazione" del livello applicativo di cui si darà una breve panoramica nel paragrafo seguente.

Quanto descritto poc'anzi consente di raggiungere uno degli obiettivi primari di progetto: la possibilità di richiamare, direttamente dalla scheda di una entità (ente, persona o famiglia) - presente nel punto d'accesso unico per Nomi, esplorabile attraverso una innovativa interfaccia di ricerca guidata di cui si dirà nel prosieguo - l'insieme dei titoli riferibili a tale entità indipendentemente dalla loro provenienza, e di filtrare con semplicità i beni correlati scegliendo tra due relazioni primarie, semanticamente chiare, "documenti di" e "documenti su".

Integrazione applicativa

La sostenibilità di questo modello di integrazione dell'informazione è garantita dalla contestuale reingegnerizzazione dei gestionali EDIT16 e MOL, dotati della capacità di condividere i propri record - già in fase di creazione dell'informazione a cura degli operatori accreditati - con l'Indice SBN attraverso l'uso dei servizi rappresentati dal protocollo applicativo SBNMarc⁴.

² http://edit16.iccu.sbn.it/web_iccu/ihome.htm.

³ <https://manus.iccu.sbn.it/>.

⁴ <https://www.iccu.sbn.it/it/SBN/evoluzione-dellindice-sbn/protocollo-sbn-marc/index.html>.

I nuovi EDIT16 e MOL si configurano di fatto, nel nuovo assetto, come Poli speciali del Servizio Bibliotecario Nazionale, in grado di contribuire al Catalogo collettivo mantenendo le proprie specificità e autonomia.

Parallelamente allo sviluppo delle nuove applicazioni gestionali, la riorganizzazione del complesso sistema di aggregazione e fruizione delle risorse digitali dell'ICCU, rappresentato da Internet Culturale, ha permesso di sviluppare una nuova architettura dei flussi digitali articolata in tre linee di intervento - dettagliate nel prossimo paragrafo - che nel loro complesso consentono di raggiungere l'obiettivo di rappresentare il contenuto digitale come estensione del Catalogo, attraverso il sistematico riferimento delle "copie digitali" dei beni alle schede bibliografiche che li rappresentano.

Digital library distribuita

Internet Culturale continuerà a svolgere la sua funzione di infrastruttura dedicata alla raccolta e indicizzazione delle copie digitali rese disponibili da Teche remote, ma non disporrà più di un proprio portale di pubblicazione e di strumenti di ricerca e fruizione destinati all'utenza. Nell'architettura del nuovo sistema informativo il suo ruolo sarà quello di rendere disponibili al sistema di integrazione e indicizzazione centrale tutte quelle risorse digitali che siano riconducibili a una scheda catalografica già presente nel sistema di Catalogo, comunicando anche tutte quelle informazioni tecniche e d'uso che hanno la funzione di arricchire il *core* bibliografico e potenziare i suoi strumenti di ricerca sul patrimonio digitale.

Allo scopo di ampliare il ventaglio dei possibili fornitori esterni di contenuti digitali, l'aggregatore permetterà l'acquisizione anche dei soli metadati descrittivi, delegando alla Teca remota il compito di rendere visibile i propri contenuti digitali. Perché i metadati descrittivi siano facilmente allineabili alle schede catalografiche presenti in SBN e quindi nel Catalogo integrato del nuovo sistema informativo, l'Istituto pubblicherà nell'ambito di questo progetto una versione evoluta della Teca digitale italiana - strumento già in uso presso diversi provider di Internet Culturale - che consentirà, attraverso opportune mappature, la generazione automatica di tali metadati e la semplificazione dei processi necessari ad alimentare una propria teca digitale a partire dai dati di catalogo disponibili nei formati standard più diffusi.

Il nuovo Aggregatore digitale (questo è il nome tecnico corrispondente al suo rinnovato ruolo) si occuperà di gestire il "residuo" degli oggetti digitali non già associati agli elementi catalografici dei sistemi MOL, Editweb e SBNCloud, ovvero tutto ciò che i provider di oggetti digitali di SRI hanno al di fuori dei sistemi suddetti.

Allo stesso tempo disporrà di uno storage locale per gli oggetti digitali aggregati in modo da ovviare alle criticità dell'attuale sistema legate al reperimento dinamico degli oggetti digitali dalle Teche remote, consentendo in tal modo l'implementazione delle specifiche IIIF⁵.

SBNTeca come Cloud centralizzato

Una delle linee progettuali chiave nell'economia complessiva del progetto è costituita dallo sviluppo di un sistema centralizzato di gestione del patrimonio digitale/digitalizzato. I servizi resi disponibili in tale sistema sarebbero orientati sia al recupero del "digitale sommerso", conservato dagli istituti presenti sul territorio, sia al trattamento di nuovi *asset* prodotti da campagne di digitalizzazione e/o produzione di contenuti digitali e multimediali.

La componente centrale di tale sistema è rappresentata da un DAM (Digital Asset Manager) i cui servizi saranno integrati negli ambienti gestionali specialistici di EDIT16 e MOL, e nella piattaforma SBNCloud destinata a potenziare il mondo della cooperazione SBN.

Tale componente opera dunque in modalità *multi-tenant* e come estensione funzionale pienamente integrata in tali ambienti gestionali, consentendo il recupero delle informazioni tecniche e gestionali sugli oggetti digitali direttamente nei contenuti catalografici.

Ecosistema informativo

La piattaforma di pubblicazione del nuovo sistema informativo è costituita da un network di Portali, ciascuno dei quali dà accesso a servizi di ricerca e tipologie di contenuti destinati a diverse comunità di utenti caratterizzate da esigenze informative diversificate.

EDIT16 e MOL disporranno di nuovi portali dotati di sistemi di ricerca ed esplorazione dei risultati basati su faccette e di una più razionale rappresentazione delle informazioni bibliografiche. Il caso di Manus è particolarmente emblematico del cambio di paradigma in questo senso, nella più articolata rappresentazione interna della scheda manoscritto che consente ora una più chiara individuazione delle unità testuali e del loro raggruppamento in unità codicologiche.

Grazie poi all'integrazione dei servizi del DAM della SBNTeca nei sistemi gestionali dei repertori specialistici, anche i loro portali disporranno della rappresentazione diretta

⁵ In termini di IIIF Image API, il livello garantito è "2". Cfr. <<https://iiif.io/api/image/3.0/>>.

dei contenuti digitali (digitalizzazioni parziali o integrali del documento fisico) attraverso il visualizzatore centrale dell'ecosistema, Mirador, basato sul protocollo IIIF.

Accanto alla reingegnerizzazione e rivisitazione funzionale delle piattaforme specialistiche di ricerca, il progetto ha previsto un analogo intervento sul portale del Catalogo del Servizio Bibliotecario Nazionale.

Il portale principale di progetto: Alfabetica

Grazie al core bibliografico rappresentato dal Catalogo generale e le sue logiche di integrazione, sommariamente descritte nei paragrafi che precedono, si intende superare il modello tradizionale di ricerca e restituzione bibliografica (quello di un OPAC, per quanto potenziato), in favore della costruzione di un modello di ricerca guidata da un rigoroso e tuttavia flessibile sistema di tassonomie.

Alfabetica – questo il nome del nuovo servizio – classifica tutte le risorse (i documenti) e le entità che nei cataloghi vi sono collegate (Nomi di enti, luoghi, persone) secondo le voci di una tassonomia doppia incrociata⁶.

Una tassonomia flessibile di questo tipo consente in primo luogo una classificazione multipla dei medesimi documenti grazie alla presenza di suddivisioni tipologiche ricorrenti in più categorie del sistema, ma allo stesso tempo considera una categoria tipologica (intesa come classificazione diretta di tipologie documentarie mappabili a priori con la classe) come “perimetro tematico” in grado quindi di richiamare tipologie documentarie, estranee come tali alla classe di interesse, in virtù del solo contenuto rappresentato. Individuare questo contenuto in assenza di chiavi semantiche esplicite (prodotto di una catalogazione semantica) è compito particolarmente complesso che ha comportato un lungo lavoro di analisi dei dati e poi di sperimentazione dei risultati di ricerca ottenuti in fase di implementazione del sistema di indicizzazione centrale.

SBNCloud

Il “framework di servizi” delineato nei paragrafi precedenti non si limiterà alla sola parte informativa, destinata al pubblico, ma si estenderà ai servizi gestionali attraverso lo sviluppo del nuovo ambiente integrato SBNCloud.

⁶ Utilizza un vocabolario controllato di parole per classificare i suoi oggetti, e lo organizza in due insiemi, l'insieme delle categorie (tassonomia verticale) e l'insieme dei tag (tassonomia orizzontale). Le categorie mappano i tipi di documenti con i canali di ricerca e rappresentazione dei contenuti del portale Alfabetica e identificano l'ambito di manifestazione primaria delle risorse; i tag (il contenuto informativo “parla di”), invece, portano il contenuto altrove classificando gli stessi beni (le loro manifestazioni secondarie basate sui tematismi individuati) con categorie e suddivisioni altre.

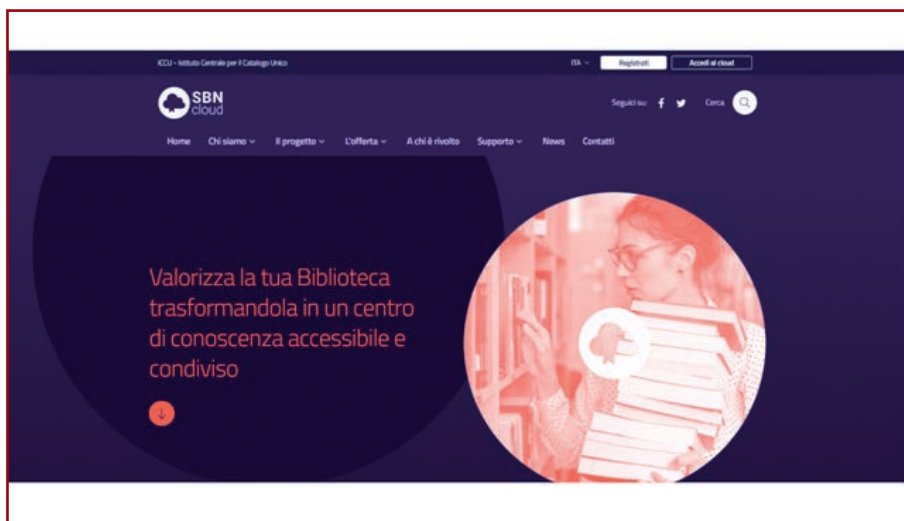


Figura 1.

SBNCloud costituisce la piattaforma di aggregazione dove saranno disponibili tutti i servizi e le informazioni relative al patrimonio di ogni biblioteca che voglia aderire a SBN usufruendo dei vantaggi, economici e gestionali, derivanti da un Cloud “software as a service” erogato direttamente dall’Istituto. Un radicale cambio di prospettiva che sposta l’enfasi sul servizio offerto e non più sul prodotto.

SBN: nuove esigenze crescono

La struttura stellare della rete SBN, con l’Indice al centro e i Poli in periferia, è stata pensata e mantenuta per rendere più agevole l’adesione a SBN da parte delle innumerevoli biblioteche di diversa tipologia diffuse su tutto il territorio italiano. Queste avrebbero così trovato nei poli, strutture preposte all’accoglienza in SBN, un presidio che si sarebbe occupato di erogare corsi di formazione, indispensabili per garantire una corretta adesione alla catalogazione partecipata, in grado di sobbarcarsi gli oneri dovuti alla gestione dell’hardware e software necessari al colloquio con l’Indice. Una struttura che avrebbe favorito la circolazione delle informazioni utili alla cooperazione, raggiungendo in maniera rapida e capillare le biblioteche collegate al Polo.

Questa organizzazione ha svolto egregiamente il proprio compito per molti anni, crescendo in numero (dal 1996 al 2020 i Poli sono passati da 33 a 104 e le biblioteche aderenti da 419 a 6.590)⁷.

⁷ Dati al 14 dicembre 2020. Per consultare il numero aggiornato di poli e biblioteche si veda la pagina <<https://www.iccu.sbn.it/it/SBN/poli-e-biblioteche>>.

Da qualche tempo, tuttavia, la continua diminuzione del personale in servizio e i consistenti e costanti tagli agli enti pubblici hanno minato l'efficienza dei Poli nelle loro prerogative istituzionali, proprio nel momento in cui più forte si sentiva l'esigenza, da parte delle biblioteche, di trovare sostegno nella rete SBN. I tagli di cui si parlava colpiscono, infatti, indiscriminatamente tutte le strutture territoriali, dalle piccole biblioteche di quartiere alle pubbliche statali.

L'ICCU, come fornitore di servizi e coordinatore della rete SBN, sta da tempo riflettendo, insieme ai partner SBN, sulla necessità di un profondo rinnovamento e cambiamento del servizio stesso, che sia in grado di raccogliere le nuove esigenze e rispondere in maniera adeguata, fornendo alle strutture territoriali il giusto sostegno, cercando di sollevarle il più possibile da quelle attività tecniche indispensabili alla partecipazione alla rete SBN che al contempo sono costose in termini di risorse economiche e professionali. In tal modo soprattutto i Poli potrebbero orientare le proprie energie alle attività di informazione, formazione, indirizzo, controllo, indispensabili al buon funzionamento dei servizi e al miglioramento della qualità dei dati dell'Indice.

L'applicativo SBNWeb

Per la catalogazione partecipata e per la gestione dei servizi bibliotecari la comunità SBN ha a disposizione da anni, gratuitamente, l'applicativo SBNWeb, sviluppato dall'ICCU che ne garantisce tuttora la manutenzione.

L'Istituto ha recentemente pubblicato i codici sorgenti di SBNWeb e dell'OPAC di Polo sul *repository* di *code hosting* GitHub⁸, inaugurando così una nuova politica di distribuzione dei software che progetta e sviluppa per le biblioteche, gli istituti culturali e il mondo della cooperazione SBN. Il software è stato rilasciato con licenza *GNU AGPL v3*⁹ secondo quanto stabilito dalle *Linee guida su acquisizione e riuso di software per le pubbliche amministrazioni*¹⁰ dell'AGID¹¹.

Va detto, comunque, che SBNWeb è un applicativo andato in esercizio ormai più di 15 anni fa e la sua tecnologia risente del tempo trascorso. Inoltre, i continui adeguamenti alle esigenze di una eterogenea tipologia di biblioteche e la volontà di rendere molti servizi parametrizzabili hanno nel tempo reso il software sempre più complesso e, per le piccole realtà bibliotecarie, piuttosto sovradimensionato.

⁸ <https://github.com/IstitutoCentraleCatalogoUnicoBiblio?tab=repositories>.

⁹ <https://www.gnu.org/licenses/agpl-3.0.html>.

¹⁰ https://www.agid.gov.it/sites/default/files/repository_files/lg-acquisizione-e-riuso-software-per-pa-docs_pubblicata.pdf.

¹¹ Agenzia per l'Italia Digitale: <<https://www.agid.gov.it/>>.

The screenshot shows the 'Sintetica biblioteca' (Synthetic library) interface. At the top, there is a navigation bar with 'SBN cloud', 'Menu', 'Polo sviluppo CSW', 'Seleziona biblioteca', and 'ITA'. On the right, it shows 'super user' and a 'Logout' button. Below the navigation bar, the title 'Sintetica biblioteca' is displayed. The main content is a table with the following columns: Polo, Codice SBN, Codice ISIL, Denominazione, Indirizzo, Tipologia biblioteca, Centro sistema, and Funzioni. The table contains four rows of data, each with a search icon in the 'Funzioni' column.

Polo	Codice SBN	Codice ISIL	Denominazione	Indirizzo	Tipologia biblioteca	Centro sistema	Funzioni
AGR	24	IT-AG0078	Biblioteca della Provincia regionale di Agrigento Gaspare Ambrosini	piazzale Aldo Moro 1 92100 Agrigento	Specializzata		<input type="button" value="Q"/>
		IT-AG0134	Biblioteca del Polo territoriale universitario di Agrigento. Università degli studi di Palermo	Via Ugo La Malfa 92100 Agrigento	Istituto di insegnamento superiore		<input type="button" value="Associa"/> <input type="button" value="Q"/>
		IT-PA0671	Biblioteca del Polo didattico di Agrigento. Università degli studi di Palermo	Palermo	NON SPECIFICATA		<input type="button" value="Associa"/> <input type="button" value="Q"/>
CSW	01	IT-AG0005	Biblioteca dell'Archivio di Stato di Agrigento	Via Mazzini 185 92100 Agrigento	Specializzata		<input type="button" value="Associa"/> <input type="button" value="Q"/>

Figura 2.

Dal software al servizio

Il progetto SBNCloud nasce dall'esigenza di dare una risposta concreta alle criticità emerse nel tempo e brevemente descritte poc'anzi, sia per quanto riguarda l'aspetto amministrativo che le necessità tecnologiche.

Il nome stesso del progetto richiama il nuovo assetto su cui si baserà la prossima architettura, che verrà realizzata con software completamente open-source. Si tratta di una vera e propria rivoluzione rispetto alla concezione di servizio portata avanti dall'ICCU nei confronti degli utenti (sia interni che esterni) della rete SBN: offrire alle diversissime tipologie di biblioteche dei servizi scalabili in base alle altrettanto diverse esigenze cui tali servizi sono destinati.

Non più l'offerta di un applicativo monolitico, parametrizzabile sì, ma sempre complesso e ricchissimo uguale per tutti, bensì un servizio erogato in funzione di ciò che ciascuna entità SBN richiede effettivamente.

Quando si parla di entità ci si riferisce ai Poli tradizionali, alle singole biblioteche o anche ad aggregazioni di biblioteche (pensiamo alle tantissime reti territoriali, sistemi metropolitani, sistemi di comunità montane, già presenti e operanti all'interno di alcuni Poli).

In estrema sintesi, il progetto mira a:

- rendere meno onerosa sia in termini economici che operativi la gestione tecnica per Poli e biblioteche SBN. Soltanto le istituzioni di grandi dimensioni e meglio strutturate dispongono e disporranno delle risorse (non solo economiche ma innanzitutto tecnico-gestionali) necessarie a gestire in autonomia un sistema integrato orientato alla valorizzazione del patrimonio e all'erogazione dei servizi all'utenza. Le biblioteche di medie e piccole dimensioni rischiano l'emarginazione da

SBN proprio a causa dell'insostenibilità di costi e competenze;

- proporre tale architettura come servizio SaaS (Software as a Service) multi-tenant direttamente erogato dall'ICCU, in grado di ospitare anche singole biblioteche che trovano difficoltà a entrare in un Polo già costituito;
- dare maggior impulso a quei servizi divenuti via via più richiesti dall'utenza finale, come quelli relativi alla gestione del digitale in tutti i suoi aspetti, dagli strumenti di metadatazione alla gestione dei servizi di fruizione del digitale (DAM). Per quest'ultimo aspetto, SBNCloud utilizzerà il modulo apposito creato nell'ambito del progetto SRI (sistema SBNteca).

Il modello gestionale del progetto di sviluppo

La complessità del progetto ha richiesto l'adozione di un modello AGILE e di un approccio SCRUM¹² nello sviluppo delle componenti software dell'ecosistema che ha visto coinvolto sostanzialmente tutto il personale dell'Istituto. Il progetto del Sistema di ricerca integrato è stato suddiviso in dieci sotto-progetti (Work-package), ciascuno dei quali coordinato da un referente d'area per l'ICCU e da una figura corrispondente nella organizzazione del Raggruppamento temporaneo di imprese (RTI) aggiudicatario dell'appalto:

WP1 - Reingegnerizzazione di EDIT16 gestionale

WP2 - Reingegnerizzazione di MOL gestionale

WP3 - Sistema di importazione-come-riconoscimento (sviluppo del cluster d'autority)

WP4 - Reingegnerizzazione del backend di Internet Culturale

WP5 - Sviluppo del sistema SBNTECA

WP6 - Sviluppo del motore di ricerca integrato

WP7 - Sviluppo del Portale del Servizio Bibliotecario Nazionale (OPAC)

WP8 - Sviluppo del Portale di EDIT16

WP9 - Sviluppo del Portale di MOL

WP10 - Sviluppo del Portale principale: Alfabetica

Non meno complessa è l'articolazione del progetto di sviluppo complementare denominato SBNCloud e che vede coinvolto in primo luogo il personale dell'area SBN e dell'area Servizi.

I lavori si concluderanno nell'arco del 2021 con la progressiva pubblicazione di tutti i servizi tra la primavera e l'autunno.

¹² [https://it.wikipedia.org/wiki/Scrum_\(informatica\)](https://it.wikipedia.org/wiki/Scrum_(informatica)).

The Integrated Search System (ISS) represents a complex integration project of the databases managed by the ICCU and at the same time provide end user with an original research service, intended to meet the needs of a wider public, not only by sector specialists. Within this project there are sub-projects, the SBNTeca and SBNCloud, which aim to offer the SBN community technologically advanced tools, designed to provide a secure support in the face of growing demand for digital services.

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2020

Non solo ebook: la biblioteca digitale per l'e-learning

«DigItalia» 2-2020
DOI: 10.36181/digitalia-00012

Anna Maria Tammaro

Università di Parma. Dipartimento di Ingegneria ed Architettura

La biblioteca digitale ed i sistemi e-learning sono entrambi sistemi socio tecnologici la cui visione ha avuto origine dalla scienza cognitiva. Malgrado ciò, attualmente esiste un gap tra le comunità delle biblioteche digitali e dell'e-learning. Attraverso un'analisi sistematica della letteratura e della documentazione vengono descritti gli approcci teorici e le buone pratiche della convergenza tra biblioteche digitali ed e-learning. È necessaria una migliore comprensione dei bisogni formativi - non come sono percepiti dalla comunità delle biblioteche digitali ma come sono sperimentati nel campo dalle comunità di apprendimento - questo contribuirà a colmare il divario tra le due comunità delle biblioteche digitali e dell'e-learning. Inoltre, offrire risorse educative aperte sarà un passo strategico utile per rafforzare il coinvolgimento delle biblioteche digitali nell'apprendimento. Portando nuove applicazioni nel campo dell'insegnamento, le biblioteche digitali potrebbero agire come innovatrici nel campo dell'apprendimento.

Introduzione

La biblioteca digitale è stata identificata da Bush nel 1945¹ come un'infrastruttura per la creazione della conoscenza. Memex, come Bush ha chiamato il sistema informativo che ha orientato la nascita della biblioteca digitale, è un deposito centralizzato di risorse informative, dotato di sistemi avanzati di recupero dell'informazione. Licklider venti anni dopo ha dedicato il suo libro a Bush, evidenziando la continuità della sua visione di biblioteca digitale con quella di Bush, in un contesto tecnologico che nel 1965 tuttavia non è più centralizzato ma distribuito in linea. Licklider², come primo direttore di DARPA, che poi diventerà Internet, prevede tutte le tecnologie che potranno essere applicate alle biblioteche del futuro per trasformarle da depositi di risorse a protesi per pensare: "pro-cognitive utility net". La visione di Bush e di Licklider si fonda sulle scienze cognitive³ e risponde al bisogno della comunità di ricerca di avere un ambiente favo-

¹ Vannevar Bush, *As we may think*, «Atlantic monthly», July (1945), p. 101-108, <<https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/>>.

² Joseph C. R. Licklider, *Library of the future*, Cambridge (Mass.): MIT Press, 1965.

³ Le scienze cognitive sono interdisciplinari e quindi è difficile definirne le discipline costituenti. Prevale recentemente un approccio comportamentale e lo psicologo cognitivo Legrenzi le definisce come: "... il campo di studio di tutto ciò che ha a che fare con le capacità creative dell'uomo e con gli artefatti da lui creati", Paolo Legrenzi, *Prima lezione di scienze cognitive*, Roma: Laterza, 2002, p. 3.

revoles alla creazione di conoscenza. La realtà delle biblioteche digitali, a partire dallo sviluppo del web negli anni '90, ha visto la continuazione della visione della biblioteca digitale come infrastruttura per la creazione di conoscenza, anche se sono state perseguite anche altre visioni di biblioteca digitale, più vicine al concetto della biblioteca tradizionale, come la realizzazione di biblioteche nazionali digitali guidate dalle istituzioni culturali (GLAM Archivi, biblioteche, musei, gallerie). La biblioteca digitale oggi di cui scriviamo in questo articolo è un sistema informativo maturo che è andato oltre la visione di "protesi per pensare" fino a diventare un ambiente che facilita la creazione di conoscenza in modo collaborativo. Le biblioteche digitali sono definite come sistemi socio-tecnologici, dove le persone interagiscono con il computer e con altri attori sociali per la creazione di conoscenza. Nancy Van House, Ann P. Bishop e Barbara P. Buttenfield⁴ introducono questo concetto sociale della biblioteca digitale:

«We are interested in understanding the social aspects of DLs, not just social impacts but the web of social and material relations in which DLs are embedded. This book originated in the work of its editors and authors in designing, evaluating, and simply trying to understand DLs and their uses. Initially, this book was to be about DL evaluation, but it rapidly became apparent to the authors and editors that we were concerned with much more. We began with the belief that a good DL is useful. Like a traditional library, a useful DL fits the needs, activities, and contexts of the people who use it, as well as those of the people who create it, operate it and contribute to its content. The more we delved into DLs and their social worlds, the more we found ourselves drawn into questions, not just about DLs but about documents, collections, and classification; activity, work, and knowledge; politics and values; institutions; and identity, organizations, and communities».

Se le biblioteche digitali sono ambienti per creare conoscenza, possono essere anche ambienti di supporto all'e-learning?

E-learning

Anche la visione dell'e-learning si basa sulle scienze cognitive e risponde al bisogno di facilitare l'apprendimento, che fa parte del processo più ampio della creazione di conoscenza. Gordon Pask nel 1975 ha applicato la sua "Teoria delle conversazioni"⁵ - nata per spiegare la creazione della conoscenza - all'apprendimento. La visione di Pask era concentrata sul discente che viene facilitato dalla tecnologia disponibile ad usare risorse educative, senza limitazioni di tempo e di spazio. Pask ha definito l'apprendimento come una conversazione tra docente e discente, allar-

⁴ Nancy A. Van House – Ann P. Bishop – Barbara P. Buttenfield, *Introduction: Digital Libraries as Sociotechnical Systems*, in: *Digital library use: Social Practice in Design and Evaluation*, ed. by N. A. Van House, A. P. Bishop, B. P. Buttenfield, Cambridge (Mass.): MIT Press, 2003, p. 1.

⁵ Gordon Pask, *Conversation and learning: a cybernetic theory and methodology*, Amsterdam - Oxford - New York: Elsevier, 1975.

gata all'interazione con il contenuto: le risorse educative devono essere strutturate in modo granulare e devono essere riusabili.

L'e-learning è chiamato oggi in vari modi (come formazione a distanza, online learning, web based learning) ed il concetto identifica la situazione indicata da Pask: il discente è a distanza dal docente ed ha a disposizione un ambiente per interagire con le risorse ed i diversi attori caratterizzato dalle tecnologie.

L'e-learning inizialmente si è concentrato sulla tecnologia disponibile per la fruizione di contenuti. Nel 1959 i ricercatori dell'Università dell'Illinois hanno sviluppato il primo sistema di e-learning al mondo, chiamato PLATO (Programmed Logic for Automatic Teaching Operations), collegato al computer mainframe ILLIAC, anch'esso sviluppato presso l'Università. La definizione di "e-learning" della Treccani⁶ si concentra sull'aspetto tecnologico:

«Nella tecnologia dell'informazione, complesso di mezzi tecnologici messo a disposizione degli utenti per la distribuzione di contenuti didattici multimediali. L'e-learning si distingue da altri processi di formazione a distanza perché basato su una piattaforma tecnologica, cioè su un sistema informatico che gestisce la distribuzione e la fruizione dei contenuti formativi».

La ricerca sull'e-learning non si limita all'aspetto tecnologico ma è interdisciplinare e soprattutto si basa sulle teorie dell'apprendimento sviluppate da Dewey, Vygotskij e Piaget, per citare solo i più importanti.

La teoria che più ha contribuito allo sviluppo dell'e-learning è basata sul concetto dell'apprendimento lungo tutta la vita del filosofo Dewey⁷ nel contesto di una visione democratica di un accesso uguale per tutti alla formazione: «I believe that education is a process of living and not a preparation for future living».

L'approccio pedagogico più importante su cui si è concentrata la ricerca dell'e-learning è il costruttivismo, che si basa sul costruttivismo sociale di Vygotskij⁸ e sulla teoria costruttivista evolutiva di Piaget⁹. In questo approccio, gli studenti si impegnano nella costruzione di conoscenza, anche creando un manufatto digitale, in un ambiente educativo individuale o di gruppo, in stile laboratorio guidati da guide esperte. Il costruttivismo è legato ad una didattica innovativa, centrata su chi apprende, basata sulla didattica attiva (active learning) e sulla didattica sociale (social learning). Nell'approccio costruttivista, la natura dell'apprendimento richiede l'efficacia di ambienti formativi che vengono programmati insieme ai contenuti

⁶ <http://www.treccani.it/enciclopedia/e-learning/>.

⁷ John Dewey, *The Early Works of John Dewey 1882-1898. 4: Early essays and the study of ethics, 1893-1894 : a syllabus*, London: Feffer & Simons, 1971, p. 87; John Dewey, *The Early Works of John Dewey 1882-1898. 5: Early essays, 1895-1898*, London: Feffer & Simons, 1972.

⁸ Lev Semenovič Vygotskij, *Mind in Society: The Development of Higher Psychological Processes*, Cambridge (Mass.): Harvard University Press, 1978.

⁹ Jean Piaget, *Psychology and Epistemology: Towards a Theory of Knowledge*, New York: Grossman, 1971.

per favorire l'apprendimento attivo, l'apprendimento sociale e l'apprendimento onnipresente.

In sintesi, Haythornthwaite ed Andrews¹⁰ affrontano le implicazioni sociali dell'e-learning, gli effetti trasformativi dell'approccio partecipativo e l'interazione tra tecnologia ed aspetti sociali che supporta e dirige l'e-learning. Gli autori presentano nuove prospettive sull'argomento, fornendo una base teorica a nuove pratiche di apprendimento derivata dalle discipline di scienza della formazione, scienza della comunicazione e scienza dell'informazione. Gli autori affermano che "apprendimento" e "reti" si uniscono in una gamma di configurazioni, sia di processo che di prodotto, poiché l'apprendimento è stato inquadrato variamente: come una relazione di rete sociale, un risultato di rete di relazioni nella forma di comunità di apprendimento e di conoscenza, e un risultato dell'influenza degli ambienti formativi sull'apprendimento.

La ricerca nell'ambito dell'e-learning persegue l'innovazione tecnologica e pedagogica insieme, mantenendo un forte collegamento tra ricerca e pratica. Viene rappresentata un'ampia gamma di ambienti formativi in questa ricerca; qualsiasi contesto in cui avviene l'apprendimento è un terreno aperto. In questo ambito, ci concentriamo sul ruolo della biblioteca digitale come ambiente formativo.

Convergenza delle biblioteche digitali con l'e-learning

La biblioteca digitale e l'e-learning sono sistemi socio-tecnologici in cui la tecnologia facilita i processi della creazione di conoscenza e dell'apprendimento. Il processo di apprendere è intersecato con il processo della creazione della conoscenza, la comunicazione scientifica ed anche la ricerca dell'informazione.

Marchionini e Maurer nel 1995¹¹ hanno pubblicato il primo studio sul ruolo della biblioteca digitale per l'e-learning ed hanno individuato almeno tre ruoli:

1. Il primo è un ruolo pratico, come quello di condividere risorse costose inclusi bibliotecari professionisti;
2. Il secondo è un ruolo culturale, selezionando, organizzando e curando nel tempo risorse di qualità;
3. Il terzo è un ruolo sociale ed intellettuale mettendo insieme persone ed idee.

Marchionini e Maurer individuano come il maggiore vantaggio della biblioteca digitale quello di rendere facile l'apprendimento lungo tutta la vita, sia quello formale che quello informale e quello professionale (non formale).

¹⁰ Caroline Haythornthwaite – Richard Andrews, *E-learning theory and practice*, New York: Sage publications, 2011.

¹¹ Gary Marchionini – Hermann A. Maurer, *The roles of digital libraries in teaching and learning*, «Communications of the ACM», 38 (April 1995), n. 4, p. 67-75, <<https://doi.org/10.1145/205323.205345>>.

Roes¹² afferma nel 2001, durante un corso a Tilburg per bibliotecari digitali, che le biblioteche digitali devono necessariamente integrarsi coi sistemi e-learning, considerando i bisogni formativi di diverse tipologie di discenti coinvolti nell'apprendimento lungo tutta la vita. L'autore analizza i due diversi approcci perseguiti per la convergenza di biblioteche digitali ed e-learning negli Stati Uniti ed in Europa. L'approccio europeo è stato più tecnologico, cercando l'interoperabilità delle biblioteche digitali con i sistemi e-learning. L'approccio negli Stati Uniti si è concentrato soprattutto sui contenuti, costruendo ad esempio una biblioteca digitale nazionale di risorse educative dedicate a tutti i tipi di discenti e livelli di scuole come NSDL¹³.

Due biblioteche digitali educative hanno pubblicato le loro riflessioni sulla loro esperienza. Zia¹⁴ ha presentato la biblioteca digitale educativa NSDL come un ambiente di apprendimento e una rete di risorse e comunità per la formazione per la scienza, la matematica, l'ingegneria e l'educazione tecnologica. La biblioteca digitale come ambiente di apprendimento per l'autore deve avere diverse caratteristiche:

- è progettato per soddisfare i bisogni dei discenti, sia come individui che come gruppi;
- è costruito per consentire l'uso dinamico di un'ampia gamma di risorse per l'apprendimento principalmente in formato digitale;
- è "curato" attivamente per promuovere un accesso affidabile sempre e ovunque a collezioni e servizi di qualità, disponibili sia all'interno che all'esterno della rete.

In aggiunta nel 2004 lo studio di Tamara Sumner e Mary Marlino¹⁵ responsabili di un'altra biblioteca digitale educativa, la Digital Library for Earth System Education, evidenzia che le biblioteche digitali educative come ambienti di apprendimento hanno tre caratteristiche: sono sistemi cognitivi, sono depositi per risorse formative ed anche sono reti collaborative per la costruzione della conoscenza.

Le prime esperienze di biblioteche digitali educative hanno consentito di approfondire l'importanza della metodologia pedagogica, adottando il metodo del design educativo dei corsi o "instructional design". Saeed afferma che, sebbene le biblioteche digitali hanno il potenziale di rivoluzionare la tradizionale classe, devono essere accompagnate da metodologie didattiche per realizzare questo poten-

¹² Hans Roes, *Digital Libraries and Education: Trends and Opportunities*, «D-Lib Magazine», 7 (2001), n. 7-8, <<http://www.dlib.org/dlib/july01/roes/07roes.html>>.

¹³ NSDL National Science Digital Library, <www.nsdlib.org>.

¹⁴ Lee L. Zia, *Growing a National Learning Environment and Resources Network for Science, Mathematics, Engineering, and Technology Education: Current Issues and Opportunities for the NSDL Program*. «D-Lib Magazine», 7 (2001), n. 3, p. 65-73, <<http://www.dlib.org/dlib/march01/zia/03zia.html>>.

¹⁵ Tamara Sumner – Mary Marlino, *Digital libraries and educational practice: a case for new models*, in: *JCDL. Proceedings of the 4th ACM/IEEE-CS Joint conference on Digital libraries, June 2004*, New York: Association for Computing Machinery, 2004, p. 170-178, <<https://doi.org/10.1145/996350.996389>>.

ziale¹⁶. Ancora più importante è che i sistemi e-learning (VMS Virtual learning environment) ed i sistemi delle biblioteche digitali (DAMS Digital Asset Management System) vanno integrati e devono essere interoperabili per connettere le risorse educative e le attività didattiche previste nei corsi.

Recentemente l'importanza del disegno strutturato dei corsi (instructional design) come metodologia didattica è stato maggiormente compreso dalla comunità delle biblioteche digitali. Uno studio pubblicato nel 2013 da Cervone¹⁷ definisce le biblioteche digitali come ambienti formativi che si concentrano su determinati risultati formativi (learning outcomes) facilitando l'interazione tra docenti e discenti ed offrendo la disponibilità di attività ed esercizi di simulazioni:

«As we give consideration to both how digital library learning experiences adapt to the needs of today's students, as well as help us generate more meaningful learning outcomes, we need to expand our repertoire and adopt more dynamic learning environments, such as simulation exercises in the online environment, rather than stick to the tried and true models such as the guided tutorial model commonly found today. Guided tutorials are fundamentally based on the prescriptive model and do not actively foster an interactive learning environment. Simulations, on the other hand, promote an interactive learning environment that provides students with a framework for constructing their own knowledge and models for future learning».

Il processo di apprendere è quindi intersecato con la ricerca dell'informazione ma l'approccio costruttivista e della didattica attiva resta come gap tra i due processi. Gary Marchionini¹⁸, nel fascicolo inaugurale di una nuova rivista accademica dedicata proprio alla convergenza di scienza dell'informazione e teorie dell'apprendimento dal titolo "Information and Learning Sciences", ha passato in rassegna i progressi nella ricerca dell'informazione negli ultimi sessanta anni, riassumendo anche diverse teorie sulla creazione di senso e teorie sull'apprendimento e ha proposto un quadro concettuale per integrare la creazione di conoscenza e l'apprendimento, basato sulla creazione di senso come ponte fra la ricerca dell'informazione e il processo dell'apprendimento.

L'affermazione principale è che i programmi del computer e gli ambienti di informazione digitale come le biblioteche digitali hanno aumentato il numero e l'intensità delle sovrapposizioni tra creazione di conoscenza e processi di apprendimento. Inoltre, i processi di creazione di senso assumono un'importanza crescente come processo saliente e attività che può essere considerata un ponte tra conoscenza e apprendimento.

¹⁶ Saeed Rezaei Sharifabadi, *How digital libraries can support e-learning*, «The Electronic Library», 24 (2006), n. 3, p. 389-401, <<https://doi.org/10.1108/02640470610671231>>.

¹⁷ Frank Cervone, *Learning, Adaptation, and Digital Libraries*, «OCLC Systems & Services: International Digital Library Perspectives», 29 (2013), n. 4, p. 200-203.

¹⁸ Gary Marchionini, *Search, sense making and learning: closing gaps*. «Information and Learning Sciences», 120 (2019), n. 1/2, p. 74-86, <<https://doi.org/10.1108/ILS-06-2018-0049>>.

Il processo di creazione di senso coinvolge sotto-attività come la comprensione, l'interpretazione, e la riflessione sulle singole unità di informazione. La ricerca di informazione può dirsi quindi un'attività secondaria dell'apprendimento, e l'apprendimento è situato all'intersezione della ricerca di informazioni e della comprensione dell'informazione come attività sovrapposte. Poiché i processi di creazione di senso sono importanti soprattutto durante il risultato, la creazione di senso serve come componente della ricerca e può servire da ponte tra ricerca e apprendimento.

Vakkari¹⁹ nella sua revisione storica della ricerca nella Scienza dell'informazione come processo di apprendimento, discute l'importanza di includere l'elaborazione dei risultati della ricerca (sia per continuare l'iterazione successiva della ricerca o per iniziare a utilizzare i risultati per soddisfare il bisogno di informazioni) come una delle caratteristiche che più avvicina la ricerca dell'informazione alle teorie dell'apprendimento. Sottolinea inoltre la necessità di metriche orientate al processo di apprendimento per valutare la ricerca dell'informazione per questo aspetto, piuttosto che fare affidamento solo su misure dei risultati della ricerca tradizionali come il richiamo e la precisione.

Buone pratiche

La comunità di ricerca delle biblioteche digitali è quindi favorevole alla convergenza di biblioteche digitali ed e-learning, tuttavia gli esempi reali di biblioteche digitali che apertamente dichiarano di dare servizi per l'e-learning sono pochi.

Il Rapporto "Digital libraries in education: analytical survey" pubblicato dall'UNESCO Institute for Information Technologies in Education (IITE) ha sottolineato la necessità di coordinare lo sviluppo delle tecnologie educative con la crescita delle biblioteche digitali²⁰. Secondo l'UNESCO, i benefici che le biblioteche digitali possono apportare alla formazione sono molto ambiziosi.

Nella definizione data dall'UNESCO IITE, le biblioteche digitali sono finalizzate al supporto dell'intero ciclo di creazione della conoscenza:

« [...] an environment bringing together collections, services, and people to support the full cycle of creation, dissemination, discussion, collaboration, use, new authoring, and preservation of data, information, and knowledge»²¹.

Questa definizione dell'UNESCO mette insieme risorse, servizi e persone che interagiscono per il processo di creazione della conoscenza, e corrisponde alla descrizione di biblioteche digitali come ambiente formativo.

¹⁹ Pertti Vakkari, *Searching as learning: a systematization based on literature*, «Journal of Information Science», 42 (2016), n. 1, p. 7-18, <<http://journals.sagepub.com/doi/abs/10.1177/0165551515615833>>.

²⁰ UNESCO Institute for Information Technologies in Education (IITE), *Digital libraries in education: analytical survey*, Moscow, 2003, <<https://iite.unesco.org/publications/3214609/>>.

²¹ Ivi, p. 7

Nel 2008 UNESCO IITE ha promosso un corso²² per i docenti per renderli capaci di costruire biblioteche digitali specializzate ai fini del corso e/o della scuola. Il docente del corso è stato Ian Witten, creatore del software Greenstone per costruire biblioteche digitali (DAMS) e su cui il corso UNESCO si è basato. Questa esperienza tuttavia di sensibilizzare i docenti e renderli capaci di costruire biblioteche digitali non ha avuto seguito.

Nella comunità professionale delle biblioteche (non solo digitali), il fenomeno dell'e-learning ha avuto un impatto a cominciare dalle biblioteche accademiche²³. Un'esperienza importante sono gli Standard di ACRL²⁴ per la formazione a distanza pubblicati nel 2006 (rinnovati nel 2016).

I temi all'intersezione tra biblioteche e processo di apprendimento sono stati individuati in tre aspetti principali: competenze digitali, applicazione dei metodi del costruttivismo, valutazione dei risultati di apprendimento.

Il Media Information Literacy (MIL) è una capacità ritenuta oggi essenziale ed il primo compito delle biblioteche per la didattica è quello di cercare di incorporare i corsi per l'alfabetizzazione nei curricula universitari. Questa necessità richiede che i bibliotecari inizino ad usare il metodo costruttivista e la progettazione didattica in collaborazione coi docenti e seguendo le politiche istituzionali.

L'alfabetizzazione MIL è talvolta considerata un requisito formativo generico o invece specifico del corso di laurea; in altri casi può essere un corso opzionale scelto dallo studente. Allo stesso modo, la tipologia del corso varia dall'orientamento con lezioni brevi (one-shot) o dalla spiegazione specifica di determinati strumenti (tutorial), a corsi indipendenti della biblioteca o a moduli incorporati nei singoli corsi di laurea. Le modalità di formazione sono adeguate alle diverse tipologie e vanno da sessioni faccia a faccia fino alla guida in linea just-in-time ed ai tutorial web.

Nell'ambiente dell'e-learning, le biblioteche digitali sono considerate soprattutto come una combinazione di servizi e risorse, non tanto come un ambiente di apprendimento digitale. La motivazione può ritrovarsi nel fatto che spesso non sono applicati i metodi didattici, così mancano attività di apprendimento attivo, apprendimento basato su progetti, esercizi e simulazioni. La gamma di risorse educative disponibili inoltre è spesso limitata ad ebook mentre le biblioteche dovrebbero includere oggetti multimediali e risorse come curricula e altro materiale didattico, come piani delle lezioni, syllabus, insieme a strumenti come programmi per la modellazione e simulazione, sistemi di tutoraggio intelligenti, accesso a strumenti

²² <https://iite.unesco.org/pics/publications/en/files/3214563.pdf>.

²³ Si veda: Anna Maria Tammaro, *Le biblioteche di fronte alla sfida dell'e-learning*, «Biblioteche oggi», (2004), n. 9, p. 59-62, <<http://www.bibliotecheoggi.it/2004/20040905901.pdf>>; Anna Maria Tammaro, *Centri di risorse didattiche multimediali: problematiche e tendenze in Italia*, «Biblioteche oggi» (2002), n. 3, p. 6-16, <<http://www.bibliotecheoggi.it/2002/20020300601.pdf>>.

²⁴ Association of College & Research Libraries (ACRL), *Standards for Distance Learning Library Services*, <<http://www.ala.org/acrl/standards/guidelinesdistancelearning#changing>>.

scientifici remoti. La descrizione delle risorse nella collezione spesso non include metadati per la didattica.

Per la valutazione dei risultati formativi, le biblioteche dovrebbero fornire servizi per docenti come valutazioni e “revisioni tra pari” dei compiti inseriti dagli studenti. A studenti e docenti, la biblioteca digitale potrebbe offrire la capacità di interagire con i colleghi, di fornire archiviazione dei materiali didattici, usufruire di sistemi di raccomandazione e di diffusione selettiva dell’informazione.

Docenti, studenti e altri utenti, come chi studia autonomamente, potranno trarre vantaggio da queste attività e strumenti offerti dalle biblioteche per il processo educativo. In questo modo, le biblioteche digitali saranno molto di più della somma delle loro parti (risorse e servizi) e promuoveranno il cambiamento e l’innovazione nella didattica.

Open Education è un’altra area che offre opportunità alle biblioteche digitali. Vanessa Proudman di SPARC Europe ha pubblicato recentemente²⁵ i risultati di uno studio sull’attuale servizio di supporto all’Open Education (OE) nelle biblioteche accademiche, studio che è stato condotto per comprendere meglio la visione del servizio, in che misura le biblioteche hanno adottato le politiche OE, le strategie delle biblioteche per promuovere le OER (Open Education Resources) e il ruolo delle biblioteche nella gestione delle OER. Le aree più comuni in cui le biblioteche hanno assunto un ruolo guida nella fornitura di servizi di supporto sono state i corsi per l’alfabetizzazione dell’informazione, i servizi di supporto alla ricerca di OER, la gestione delle raccolte ed i rapporti con gli editori e gli aggregatori. Alcune biblioteche hanno avuto inoltre un ruolo di supporto per la consulenza ai docenti su copyright e licenze, la cooperazione per la co-creazione di OER e lo scambio di conoscenze e di supporto tecnico per le piattaforme. In particolare, è interessante notare che le biblioteche accademiche si stanno assumendo un forte ruolo di supporto nella co-creazione di OER. Infine, il rapporto di SPARC Europe evidenzia le principali sfide e opportunità chiave nel supportare OE come sono percepite oggi dalle biblioteche intervistate. In sintesi, le biblioteche digitali hanno l’opportunità di favorire l’apprendimento sia come strumenti che come ambienti formativi. Il potenziale delle biblioteche digitali per favorire l’apprendimento va oltre l’offerta di strumenti ed ebook, come descritto nella Tabella 1.

Gli strumenti della biblioteca digitale sono classificati come strumenti per l’informazione che includono i tradizionali servizi di accesso e strumenti per l’interazione che estendono i servizi alle connessioni necessarie per il processo dell’e-learning. Gli ambienti formativi predisposti dalla biblioteca digitale sono classificati secondo l’approccio più tradizionale delle biblioteche digitali centrato sulla collezione o nell’approccio costruttivista centrato su chi apprende.

²⁵ Vanessa Proudman, *Opening up Knowledge in Higher Education. Survey results: Supporting Open Education in European Libraries today SPARC Europe 2020*, <<https://zenodo.org/record/3903175#.X1PMcS1aZQI>>.

STRUMENTI	AMBIENTI FORMATIVI
STRUMENTI PER INFORMAZIONE <ul style="list-style-type: none"> – Home page della biblioteca – Accesso all’informazione online – Motori di ricerca – Depositi istituzionali – Banche dati – Document delivery 	COLLECTION CENTERED <ul style="list-style-type: none"> – Risorse digitali integrate o non integrate con piattaforme e-learning – Depositi di OER e altro materiale educativo (Lesson plan, casi di studio, ecc.) – Metadati per la didattica – Consulenza copyright e licenze di accesso
STRUMENTI PER INTERAGIRE <ul style="list-style-type: none"> – Videoconferenza – Forum, chat, email – Reference – Social media – App Web per la didattica 	LEARNER CENTERED <ul style="list-style-type: none"> – Analisi dei bisogni formativi e piano dei risultati formativi – Integrazione con piattaforme e-learning – MOOC, OER, corsi per Media Information literacy – Collaborazione coi docenti per il disegno del curriculum – Esercizi, simulazioni, attività formative – Approcci partecipativi con studenti – Verifica dell’apprendimento

Tabella 1. *Buone pratiche delle biblioteche digitali per l’e-learning*

Alcuni esempi di buone pratiche delle biblioteche digitali per l’e-learning possono essere utili per evidenziare le opportunità per migliorare l’apprendimento che possono nascere quando c’è una scelta consapevole di rendere disponibile un servizio di supporto.

British Library

La British Library²⁶ rende disponibile un accesso dal suo sito al programma “Learning at the library” indirizzato a docenti, studenti universitari, famiglie e adulti. Una pagina del sito è dedicata alle risorse educative descritte per disciplina, per età, per percorso consigliato.

La British Library offre un ulteriore esempio del modo in cui i professionisti delle biblioteche digitali percepiscono l’uso delle proprie risorse nell’apprendimento, cercando di comprendere i bisogni formativi osservando gli scenari di uso e facendo

²⁶ <https://www.bl.uk/learning>.

progetti pilota in fase di sviluppo. Gli studi sugli scenari di uso educativo delle biblioteche digitali sono spesso preceduti dalla preparazione di collezioni raccomandate di risorse digitali, che potrebbero integrare l'insegnamento in argomenti specifici. Vale la pena ricordare che la British Library ha costruito le sue soluzioni sulla base degli scenari di uso ed ha tenuto conto delle preferenze di accesso degli utenti più giovani che accedono alle risorse web da dispositivi mobili. La documentazione della British Library per la procedura di appalto riferisce:

«availability of reputable, accurate and syllabus relevant content, that is searchable and accessible through mobile channels, is increasingly important for both A-Level students and their teachers»²⁷.

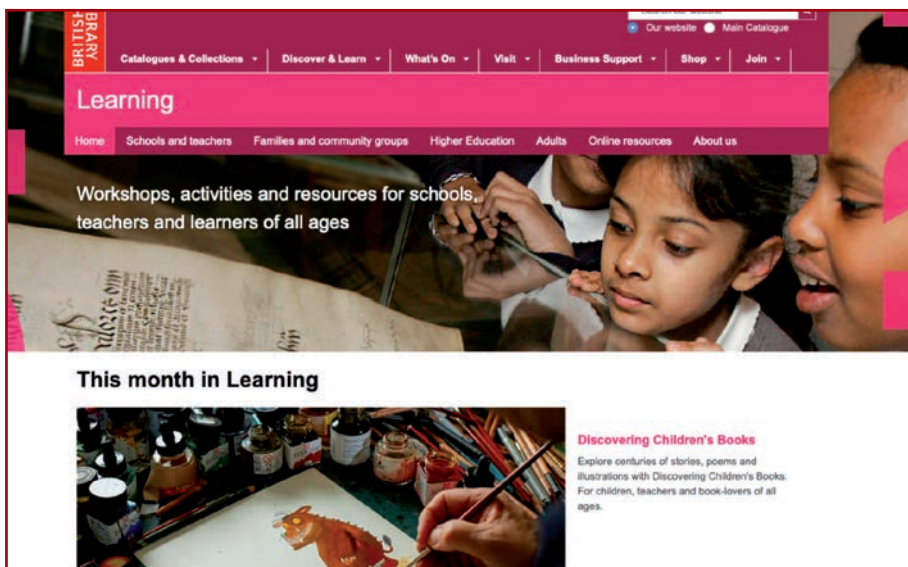


Figura 1. British Library

Europeana Space

Europeana ha reso disponibile un portale dedicato alla formazione, chiamato Europeana Space²⁸. Il portale raccoglie alcuni progetti pilota e dimostrativi delle possibilità di riuso del patrimonio culturale digitale, il MOOC²⁹ per i docenti, notizie sugli eventi organizzati periodicamente. Il portale Europeana Space realizza la Raccomandazione di Europeana for Education and Learning del 2015³⁰ in partico-

²⁷ Milena Dobrova – Galia Angelova – Gennady Agre, *Bridging the Gap between Digital Libraries and e-Learning*, «Cybernetics and Information Technologies», 15 (November 2015), n. 4, p. 92-110: 99.

²⁸ <https://www.europeana-space.eu/education/>.

²⁹ <https://www.europeana-space.eu/education/mooc/>.

³⁰ http://pro.europeana.eu/files/Europeana_Professional/Publications/Europeana%20for%20Education%20and%20Learning%20Policy%20Recommendations.pdf.

lare per quello che riguarda lo stimolo alla creatività e l'accesso aperto a risorse educative.

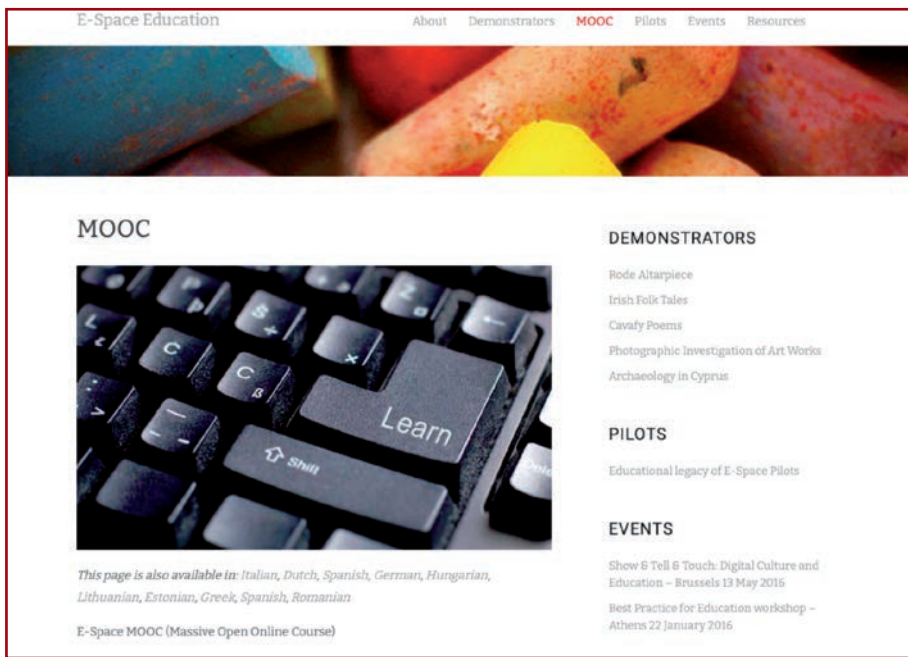


Figura 2. *Europeana Space*

Il metodo delle personas è stato utilizzato nella progettazione del sistema informativo per riassumere le caratteristiche del comportamento degli utenti tipici. L'evidenza è utilizzata per mettere insieme una descrizione di un insieme di caratteristiche tipiche di un utente che viene nominata e illustrata da un profilo che aiuta la "personificazione". Tale descrizione, chiamata persona, viene quindi utilizzata dai designer ed è particolarmente utile per i test interni del servizio. Europeana ha cercato di definire i bisogni formativi di diversi studenti e docenti attraverso l'individuazione di vari profili. La prima serie di personaggi sviluppati per gli scopi di Europeana è stata pubblicata nel 2009 e comprendeva, tra gli altri, l'insegnante Maria, lo studente Peter e la studentessa universitaria Julia. La successiva serie di persone di Europeana comprendeva Jukka etnomusicologa, Sarah commessa in un negozio, William appassionato di storia locale, e Terese professore universitario. Una dettagliata descrizione iniziale dei profili è stata fornita nel primo rapporto del progetto³¹ ed ulteriormente perfezionata nel 2011³².

³¹ Europeana Connect, *Milestone M3.2.3 Personas Catalogue*, 2009, <http://www.europeanaconnect.eu/documents/M3.2.3_eConnect_PersonasCatalogue_v1.0_20091228.pdf>.

³² Europeana Connect, *Milestone M3.2.3 Personas Catalogue Update*. 2011. <http://www.europeanaconnect.eu/documents/M3.2.3_eConnect_PersonasCatalogue_update.pdf>.

Historiana

Invece di trovare modi per integrare il materiale delle biblioteche digitali negli ambienti di apprendimento esistenti, le biblioteche digitali possono proporsi come un ambiente formativo che offre nuovi strumenti per uno specifico dominio educativo. Insieme al Progetto Euroclio, Europeana ha realizzato il portale Historiana³³ che raccoglie diversi servizi, risorse e strumenti per insegnare storia. Historiana è un ambiente formativo che offre molti strumenti per la didattica della storia.

I docenti hanno a disposizione uno strumento per creare lesson plan ed attività ed esercizi: Activity Builder. Possono anche cercare e ri-usare lezioni preparate da altri docenti e selezionare delle risorse educative predisposte da esperti. Possono naturalmente cercare tra le risorse di Europeana, ma in aggiunta hanno la possibilità di cercare partner, come ad esempio le biblioteche disponibili a collaborare per co-costruire in cooperazione un corso. Lo scopo alla base dello sviluppo di Historiana, viene così descritto:

«The goal set out for the History Education theme application is to stimulate the reuse of cultural heritage resources for history education through easy-to-find and free-to-use educational resources (sources, learning activities and tools) that are designed to stimulate historical thinking, multiperspectivity and active learning»³⁴.



Figura 3. *Historiana*

³³ <https://historiana.eu>.

³⁴ Europeana Foundation, *Transforming the World with Culture: Next Steps on Increasing the use of Digital Cultural Heritage in Research, Education, Tourism and the Creative Industries*, The Hague, 2015, <https://pro.europeana.eu/files/Europeana_Professional/Publications/Europeana%20Presidencies%20White%20Paper.pdf>.

CBook

Anche se la biblioteca digitale può offrire molto di più che ebook, questo non significa sottovalutare l'importanza di ebook interattivi e creati con approcci partecipativi. Il progetto Librare³⁵ ha adottato un approccio innovativo lavorando sul "libro", immaginando che i libri avranno un ruolo di primo piano nel mondo digitale, soprattutto superando la dicotomia carta-digitale attraverso l'interazione tra persone, oggetti, strumenti per facilitare creazione di conoscenza ed apprendimento. L'idea alla base del progetto di ricerca Librare (2015 - 2016) è stata quella di "trasformare" i volumi cartacei in oggetti digitali, seguendo il modello dell'Internet of Things, diventando oggetti tracciabili, raccogliendo sempre maggiori informazioni inserite dagli utenti. Parallelamente, in Librare è stata facilitata l'interazione tra libri cartacei e contenuti digitali in linea, consentendo una migliore integrazione delle attività che si svolgono sul web e fuori dalla scuola, offrendo attività educative alla didattica dell'aula capovolta (flipped classroom). Attraverso un insieme di strumenti (applicazioni web come Librare, Librando, First Life e Cbook) insegnanti e studenti delle scuole primarie e delle scuole medie hanno potuto socializzare la propria esperienza di lettura, geo-referenziando e commentando insieme i testi.

Queste attività didattiche sono state realizzate seguendo l'approccio costruttivista del laboratorio (Living Labs) dove studenti, docenti, cittadini sono stati coinvolti nei processi dell'apprendimento. Librare ha coinvolto studenti e docenti di tutti i gradi delle scuole (primaria e secondaria) di Torino e in altre parti d'Italia: un nucleo di circa 5 classi ha svolto tutte le attività proposte dal progetto. Innanzitutto è stata predisposta una piattaforma web come ambiente formativo per la gestione dei flussi di dati e delle interazioni che avvengono nelle Biblioteche scolastiche (book crossing). In secondo luogo, le attività di lettura sono state stimolate, individualmente o in collaborazione, con libri cartacei o edizioni online. La piattaforma offriva ai discenti e docenti due applicazioni per dispositivi mobili (Librare e Librando), un sistema di mappatura geografica (FirstLife) e un'app web per la lettura collaborativa online (il Cbook³⁶).

Il progetto Librare ha quindi affrontato con creatività alcune delle tematiche delle biblioteche digitali per l'e-learning, immaginando che i libri abbiano un ruolo da protagonisti anche nel mondo digitale.

Librare ha anche affrontato molte sfide: le scuole (per lo più primarie, ma anche alcune scuole secondarie e scuole professionali) non erano totalmente dotate di connessione wi-fi in aula e di tablet; alcune scuole secondarie avevano una biblio-

³⁵ Una descrizione del progetto Librare in: Anna Maria Tammaro – Laura Ciancio – Rosanna De Rosa – Eleonora Pantò – Fabio Nascimbeni, *Digital Libraries in Open Education: The Italy Case*, in: *Digital Libraries and Archives. IRCDL 2017. Communications in Computer and Information Science*, (v. 733), Cham: Springer, 2017, <https://doi.org/10.1007/978-3-319-68130-6_3>.

³⁶ <https://cbook.it>.

teca importante ma altre non ne avevano affatto, oppure una biblioteca molto povera e non gestita; studenti e famiglie erano molto preoccupati per la loro privacy e la possibilità di essere rintracciati; il progetto ha gestito con molta attenzione i dati e sono state messe in atto anche molte azioni correttive (creazione di gruppi chiusi). Cbook in particolare è stato molto apprezzato da docenti e studenti, ma la scarsità di titoli di risorse digitali (soprattutto classici italiani come Divina Commedia, Promessi Sposi) ne ha limitato l'adozione ma ha aumentato la consapevolezza dei diritti di proprietà intellettuale.

Conclusioni

In conclusione, mi sembra utile fare riferimento all'esperienza che abbiamo avuto durante la pandemia. Senza alcuna preparazione, durante il lockdown dei primi mesi del 2020, università e scuole sono passate da una modalità di insegnamento in presenza ad una a distanza. Nell'urgenza ci si è limitati ad una fruizione a distanza degli stessi corsi in presenza, e questa modalità è stata chiamata Emergency Remote Teaching (ERT), da non confondere con l'e-learning. Anche le biblioteche digitali hanno ottenuto rinnovata attenzione durante la pandemia e sono state pubblicizzate per la gran quantità di ebook a disposizione. Tuttavia, davvero crediamo sia sufficiente per i discenti fruire la lezione in linea senza nessuna strutturazione dei contenuti? Inoltre siamo sicuri che per chi apprende sia sufficiente trovare abbondanza di materiale in linea? In che modo questo contenuto può adattarsi perfettamente ai diversi contesti di utilizzo come la creazione di conoscenza e l'apprendimento? La pandemia ha reso evidenti dei nodi problematici, tra cui la mancanza di infrastrutture per accedere, e la necessità di capacità sia dei docenti che degli studenti. Tra le storture evidenziate, in particolare per l'impegno delle biblioteche digitali nell'e-learning, si può aggiungere l'esperienza della necessità di un ambiente formativo adeguato ai bisogni di studenti e docenti, che le biblioteche digitali hanno soddisfatto solo parzialmente. Le biblioteche digitali avrebbero potuto offrire molto di più che gli ebook ma un insieme strutturato di servizi e strumenti. Le lezioni da imparare dall'esperienza della pandemia sono molte, tra cui l'importanza di adeguare le biblioteche digitali al raggiungimento dei risultati formativi e della necessità della cooperazione con la comunità di docenti e discenti.

L'attenzione delle biblioteche digitali è attualmente quella di facilitare la scoperta di risorse educative e migliorare l'integrazione e la contestualizzazione degli oggetti digitali in ambienti formativi come l'e-learning. Sebbene queste siano funzioni ausiliarie molto utili, non sono fondamentali per il processo di apprendimento in cui la raccolta di esercizi, attività e simulazioni, così come l'aggiunta di esempi di livello diverso di difficoltà e qualità, sono molto necessari. Questo significa che le biblioteche digitali dovrebbero imparare dall'e-learning sia a strut-

turare contenuti che a disegnare ambienti formativi adeguati al raggiungimento dei risultati formativi.

La complessità dell'uso dei contenuti delle biblioteche digitali all'interno della didattica è dovuta al fatto che non possiamo parlare di usi educativi senza qualificare ulteriormente il tipo di ambienti educativi che abbiamo in mente e l'applicazione delle metodologie del disegno strutturato dei corsi (instructional design) e del costruttivismo.

Tuttavia, nonostante l'attenzione che la ricerca sulle biblioteche digitali ha riservato all'e-learning, le biblioteche digitali come ambiente per l'apprendimento "competono" ora con molti altri servizi e piattaforme pubbliche e private. L'integrazione delle biblioteche digitali nei contesti delle piattaforme e-learning e il rapporto con Open Education è stato ed è tuttora difficile. La mancanza di metadati educativi nell'organizzazione della raccolta - tra gli altri elementi - è un chiaro esempio della difficoltà delle biblioteche digitali ad assumere un ruolo educativo.

Forse la ragione di ciò è da evidenziare nella debolezza delle biblioteche digitali per la pedagogia? Seguendo la distinzione operata da Lynch³⁷ nel 2008, è evidente che le biblioteche digitali non svolgono un ruolo educativo perché non sono in grado di educare, ma solo di dare un servizio di supporto per l'apprendimento permanente.

In termini pratici, all'opposto Dobrevà, Angelova e Agre ritengono che le biblioteche digitali devono svolgere un ruolo educativo ed indicano le aree di massima importanza e urgenza: il raggiungimento di una chiara comprensione dei bisogni formativi dei contenuti nell'e-learning e il rilascio di risorse educative più aperte³⁸. La migliore comprensione dei bisogni formativi - non come sono percepiti dalla comunità delle biblioteche digitali ma come sono sperimentati nel campo dalle comunità di apprendimento - contribuirà secondo gli autori a colmare il divario tra le due comunità delle biblioteche digitali e dell'e-learning. Inoltre, offrire risorse educative più aperte sarà un passo strategico utile per rafforzare il coinvolgimento delle biblioteche digitali nell'apprendimento.

Portando nuove applicazioni nel campo dell'insegnamento, le biblioteche digitali potrebbero agire come innovatrici nel campo dell'apprendimento. Tuttavia sembra di poter affermare che resta ancora molta strada per realizzare questo potenziale e per integrare le biblioteche digitali come ambienti formativi negli attuali ambienti di e-learning.

³⁷ Clifford Lynch, *Digital Libraries, Learning Communities, and Open Education*, in: *Opening up education: the collective advancement of education through Open Technology, Open Content and Open Knowledge*, ed. by T. Iiyoshi and M. S. Vijay Kumar, Cambridge (Mass.): MIT Press, 2008, p. 105-118, <<http://mitpress.mit.edu/books/chapters/0262033712chap7.pdf>>.

³⁸ Milena Dobrevà - Galia Angelova - Gennady Agre, *Bridging the Gap between Digital Libraries and e-Learning*, cit.

The digital library and e-learning systems are both socio-technological systems whose vision originated in cognitive science. Despite this, there is currently a gap between the digital library and e-learning community. The theoretical approaches and best practices of the convergence between digital libraries and e-learning are described through a systematic analysis of literature and documentation. A better understanding of training needs is needed - not as they are perceived by the digital library community but as they are experienced in the field by the learning communities, this will help bridge the gap between the two communities of digital libraries and e-learning. Furthermore, offering open educational resources will be a useful strategic step to strengthen the involvement of digital libraries in learning. By bringing new applications to teaching, digital libraries could act as innovators in the field of learning.

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2020

Dal Catalogo generale dei beni culturali al *knowledge graph* del patrimonio culturale italiano: il progetto ArCo

«DigItalia» 2-2020
DOI: 10.36181/digitalia-00013

Chiara Veninata

ICCD

Le attività dell'ICCD sono da sempre indirizzate ad una maggiore condivisione e valorizzazione sia dei modelli di strutturazione della conoscenza sul patrimonio culturale sia dei dati prodotti nelle campagne di catalogazione. Negli ultimi anni l'ICCD ha concentrato le proprie attività sull'analisi e sull'applicazione delle potenzialità offerte dal semantic web e dai suoi strumenti. Uno dei risultati è il progetto ArCo, il grafo della conoscenza del patrimonio culturale italiano, costituito da una rete di ontologie e da oltre 169 milioni di triple riferite a oltre a 800 mila schede catalografiche. ArCo si basa sui dati del Catalogo generale dei beni culturali dell'Istituto centrale per il catalogo e la documentazione del MiBACT e sui dati dei suoi archivi fotografici. ArCo è distribuito congiuntamente con uno SPARQL endpoint, un software per convertire i record di catalogo in RDF e una ricca suite di materiale di documentazione (test, valutazione, istruzioni, esempi ecc.).

Premessa

L'Istituto centrale per il catalogo e la documentazione (ICCD) gestisce il Catalogo del patrimonio archeologico, architettonico, paesaggistico, storico artistico e demoetnoantropologico, realizzato ai sensi dell'art. 17 del Codice dei beni culturali. Il Catalogo nasce in Italia alla fine dell'Ottocento, quando il nascente Stato italiano dà avvio, nel 1875, alle prime rilevazioni sistematiche sul territorio con la finalità di censire il patrimonio storico artistico del nuovo stato italiano¹.

Il termine catalogazione fa generalmente riferimento a qualsiasi intervento volto a classificare, sulla base di caratteristiche comparabili, determinati oggetti. Esso

¹ Con la prima legge di tutela del 1902 si comincia a parlare di "Catalogo unico dei monumenti ed oggetti aventi pregio d'arte o di antichità". Ogni riferimento al Catalogo scompare nella seconda legge di tutela n. 364 del 20 giugno 1909, per ricomparire invece nel r.d. n. 1889 del 1923 in cui si richiama la necessità di compilare, per finalità sia di conoscenza che di conservazione, un catalogo dei monumenti e delle opere d'interesse storico, artistico e archeologico di proprietà statale. Anche se nulla viene normato sulla catalogazione nella legge n. 1089 del 1939, le soprintendenze continuano a catalogare il patrimonio: le schede così prodotte confluiscono nel 1969 nell'Archivio del nuovo Ufficio Centrale per il Catalogo, istituito quell'anno presso il Ministero della Pubblica Istruzione, trasformatosi in Istituto Centrale per il Catalogo e la Documentazione (ICCD) nel 1975, a seguito dell'istituzione del Ministero dei Beni Culturali.

comprende una serie di attività riconducibili alla registrazione, elencazione ed enumerazione organica di attributi di un oggetto e tuttavia non si esaurisce in essa. Il sistema della catalogazione del patrimonio culturale è infatti da sempre basato su due presupposti, l'uno rivolto alle specificità dell'oggetto di indagine, l'altro ad un recupero integrato delle informazioni pertinenti all'oggetto volto a garantirne la comunicazione nella dinamica spazio-tempo-contesto che lo caratterizza e che lo rende parte del patrimonio culturale.

Per garantire coerenza a tali presupposti, l'ICCD ha elaborato nel tempo un articolato "sistema di conoscenza": principi di metodo, strumenti e regole per acquisire le informazioni sui beni secondo procedure e criteri omogenei in tutto il territorio nazionale. La componente fondamentale di questo sistema è rappresentata dalle schede di catalogo, modelli descrittivi costituiti da una sequenza predefinita di voci, che raccolgono in modo formalizzato le notizie sui beni, seguendo un percorso conoscitivo che guida il catalogatore e al tempo stesso controlla e codifica i dati sulla base di precisi parametri e vocabolari. In relazione ai vari tipi di beni - mobili, immobili e immateriali - l'ICCD ha elaborato, ad oggi, trenta diverse schede di catalogo riferite a nove settori disciplinari (beni archeologici, beni architettonici e paesaggistici, beni demoetnoantropologici, beni fotografici, beni musicali, beni naturalistici, beni numismatici, beni scientifici e tecnologici, beni storici e artistici). Nel tracciato di ogni singola scheda si esplicitano informazioni identificative, descrittive, tecnico scientifiche, storico critiche, geografiche e di contesto, che consentono di porre in relazione il bene al territorio e agli altri beni secondo una prospettiva spazio temporale.

I dati del Catalogo generale del patrimonio culturale vengono governati attraverso il SIGECweb - Sistema informativo generale del catalogo - che gestisce circa tre milioni di schede di Catalogo che, pur nella loro numerosità, costituiscono una goccia nel mare del patrimonio culturale italiano e non consentono di realizzare, da sole, una attendibile ed esaustiva "mappa del patrimonio culturale"². Dal 2001 - con la riforma costituzionale che affida anche alle Regioni la valorizzazione del patrimonio culturale - le Regioni partecipano al processo di catalogazione del patrimonio culturale. È un sistema policentrico in cui si è dato vita a sistemi informativi regionali senza che, tuttavia, si sia mai avviato un organico e aggiornabile processo di comunicazione con il sistema informativo generale del catalogo gestito dall'ICCD³.

² Cfr. Laura Moro, *Il Catalogo del patrimonio culturale italiano. Nuova centralità e prospettive future*, «Economia della Cultura», 25 (2015), n. 3-4, p. 419.

³ Seguendo l'articolato sistema di relazione tra Stato e Regioni, a partire dagli anni Settanta, si arriva agli articoli 148 e 149 del D. Lgs. 112 del 1998 che definisce i rispettivi compiti nell'ottica delle distinte funzioni di tutela e valorizzazione dei beni culturali, distinzione poi confermata dalla riforma del Titolo V della costituzione (legge costituzionale 18 ottobre 2001, n. 3). L'Accordo nazionale Stato-Regioni del 2001 consolida un sistema policentrico in cui le Regioni avviano i loro sistemi informativi con riferimento alle necessità del territorio, in connessione con il sistema informativo generale del catalogo gestito, per il Ministero, dall'ICCD. Nel 2004, il Codice dei beni culturali all'art.

I sistemi, centrali e locali faticano ancora a scambiarsi i dati e la mancata interoperabilità tra di essi determina l'assenza di un Catalogo nazionale dei beni culturali, completo e aggiornato, ma solo la presenza delle varie e difformi "sue articolazioni"⁴.

Le possibilità offerte dalle tecnologie del web semantico

Con l'obiettivo di esplorare strade più innovative per lo scambio di dati tra sistemi informativi e per il loro arricchimento reciproco, l'ICCD ha di recente volto la sua attenzione alle possibilità offerte dal cosiddetto web semantico. In particolare la diffusione di modelli ontologici⁵ e di linked open data (LOD) sembrano offrire uno scenario in cui potrebbero trovare soluzione alcuni dei problemi suesposti. Si fa riferimento, in particolare, all'utilizzo di strumenti in grado di abilitare la descrizione, la pubblicazione e la condivisione dei dati attraverso database aperti e distribuiti (*triplestore*) in cui il singolo dato viene immediatamente reso disponibile sul web dall'istituzione responsabile, connotato semanticamente attraverso il ricorso a metadati espressi secondo formalismi (ontologie) comprensibili anche dai motori di ricerca più evoluti e collegabile facilmente ad altri dati attraverso relazioni significative, perché anch'esse riferite a modelli concettuali comprensibili dai software⁶. A livello internazionale, a partire dal 2010, si è manifestata una certa tendenza a promuovere i linked open data come una pratica standard di produzione e pubblicazione dei dati sul web, in linea con le raccomandazioni contenute nel *Library Linked data Incubator Group Final Report* rilasciato alla fine del 2011 dal *Library Linked data Incubator Group* del W3C⁷ e con gli indirizzi di Europeana⁸ e della

17 recepisce questo processo partecipativo delle regioni ma continua a fare perno sul Ministero, come garante dell'unitarietà della tutela, per assicurare la catalogazione del patrimonio culturale e definirne le metodologie con il concorso delle Regioni. Il quadro che ne emergeva era che Stato, Regioni ed enti pubblici territoriali catalogassero ciascuno i beni di loro proprietà e facessero affluire i dati prodotti al Catalogo nazionale, di fatto mai realizzato.

⁴ Cfr. art. 17 D. Lgs. 42/2004.

⁵ Le ontologie sono la formulazione di schemi concettuali esaustivi e rigorosi nell'ambito di un dominio dato. Tali schemi concettuali devono essere condivisi il più possibile nell'ambito della propria comunità di interesse e devono essere formalizzati secondo un linguaggio comprensibile anche alle macchine, che dovranno elaborare automaticamente quei dati consentendo di effettuare su di essi ragionamenti e query anche complesse.

⁶ Cfr. Commissione di coordinamento SPC, *Linee guida per l'interoperabilità semantica attraverso i linked open data*, 2012, disponibile all'indirizzo: <https://www.agid.gov.it/sites/default/files/repository_files/documentazione_trasparenza/cdc-spc-gdl6-interoperabilitasemopendata_v2.0_0.pdf>.

⁷ Disponibile all'indirizzo: <<http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/>>. Il *Final Report* forniva un'analisi dettagliata dei benefici derivanti dall'utilizzo delle tecnologie del *semantic web* nel settore dei beni culturali, sia per ciò che concerne gli utenti e i ricercatori sia per quanto riguarda gli stessi istituti culturali nei termini di una maggiore diffusione, comprensione e valorizzazione del loro immenso patrimonio informativo pubblicato sul "nuovo web".

⁸ Cfr. Harry Verwayen - Martijn Arnoldus - Peter B. Kaufman, *The Problem of the Yellow Milkmaid. A Business Model Perspective on Open Metadata*, Europea White Paper N. 2, novembre 2011. La

stessa direttiva 2013/37/UE sulla PSI (Public Sector Information), recepita anche in Italia nel 2015⁹. Negli ultimi dieci anni si stanno compiendo rapidi passi nella direzione della pubblicazione in formato linked open data dei dati relativi alle collezioni, soprattutto in ambito biblioteconomico e museale o per specifici progetti che riguardano il *cultural heritage* in maniera trasversale. Tali esperienze hanno determinato una crescente consapevolezza sulle potenzialità che i linked open data offrono alle istituzioni culturali per “aprire” i propri dati e pubblicarli secondo modalità innovative, che consentono l’arricchimento reciproco delle informazioni di partenza collegate ad altre fonti pubblicate secondo i medesimi paradigmi¹⁰. In questa direzione, oggi si inizia a parlare, anziché di “*cataloguing*”, di “*catalinking*”, secondo la felice espressione di Roy Tennant¹¹. Per l’ICCD, individuato l’obiettivo, si è trattato di comprendere come procedere alla pubblicazione dei linked open data del Catalogo senza sacrificare nulla della capacità, evoluta nel tempo, di esprimere complesse architetture della conoscenza sul patrimonio culturale.

ricerca focalizzava l’attenzione su costi/benefici dell’utilizzo dei linked data nel settore delle gallerie, biblioteche, archivi e musei (GLAMS). Venivano affrontati punti cruciali legati al rilascio come open data dei beni culturali, ovvero la licenza d’uso e la perdita del controllo delle risorse “liberate”: un nuovo modo di valorizzare le risorse culturali potrebbero essere costituito dallo sviluppo di servizi innovativi legati alle risorse culturali, applicazioni e servizi a favore dei ricercatori.

⁹ La direttiva 2013/37/UE interviene in materia di riutilizzo dell’informazione nel settore pubblico, attraverso la modifica della direttiva 2003/98/CE, la c.d. direttiva PSI (Public Sector Information), ed è finalizzata a favorire il riutilizzo dei dati delle pubbliche amministrazioni dell’Unione europea estendendo l’ambito di applicazione anche alle istituzioni culturali (biblioteche, musei e archivi) in precedenza escluse. In Italia è stata recepita con Decreto legislativo, 18/05/2015 n° 102 pubblicato in G.U. il 10/07/2015.

¹⁰ Una prima analisi dei progetti c.d. LOD-LAM (Linked Open Data for Libraries, Archives and Museums) è stata condotta con la *Review on linked open data sources*, effettuata nell’ambito del Progetto Athena Plus dell’ottobre 2013; essa è stata successivamente integrata e aggiornata da una ricerca di OCLC condotta tra luglio e agosto 2014 e poi ripubblicata con alcune correzioni nel 2015. I risultati della prima indagine del 2014, portata a termine nel 2017, sono stati successivamente confrontati con la revisione della medesima indagine condotta sempre da OCLC tra il 17 aprile e il 25 maggio 2018. All’indagine del 2018 hanno risposto 81 istituzioni che hanno segnalato un totale di 104 progetti, rispetto alle 71 istituzioni che ne avevano segnalati 112 nel 2015. Dei suddetti 104 progetti, solo 42 erano già stati descritti in precedenza. Le istituzioni italiane coinvolte nell’indagine OCLC del 2018 erano solo 4: si tratta della Biblioteca della Camera dei deputati, di Casalini Libri (SHARE-VDE group), del Coordinamento delle Biblioteche Speciali e Specialistiche di Torino (CoBIS) e dell’Università degli Studi Roma TRE. Le attività di ricerca condotte dalla scrivente nel 2018 arricchiscono notevolmente l’indagine sul fronte italiano. Cfr. Chiara Veninata, *Linked open data e ontologie per la descrizione del patrimonio culturale: criteri per la progettazione di un registro ragionato*, disponibile all’indirizzo <<https://iris.uniroma1.it/handle/11573/1325813#.XmuFM6hKg2w>>.

¹¹ Cfr. su Twitter lo storify dell’intervento *From linked data to shared knowledge* di Roy Tennant, Jon Voss e Ingrid Mason, alla conferenza di Information Online 2013 dal titolo “*Be different. Do differently*”. Information Online 2013 disponibile all’indirizzo <<https://storify.com/ALIAonline/keynote-roy-tennant-jon-voss-and-ingrid-mason>>.

La scelta del modello ontologico per il knowledge graph del patrimonio culturale

L’Agenzia per l’Italia Digitale (AGID) mira ad individuare alcuni ecosistemi digitali¹², in ciascuno dei quali una rete di ontologie per la pubblica amministrazione definita OntoPiA garantirebbe integrazione tra i dati prodotti dai soggetti che contribuiscono all’ecosistema. In questa prospettiva, l’ICCD ha inteso aprire i dati del Catalogo generale pensando ad essi come un tassello fondamentale dell’ecosistema digitale “Beni culturali e turismo”. Una volta scelto il paradigma dei linked open data, il Catalogo può assumere un nuovo ruolo come fondamento del *knowledge graph* sul patrimonio culturale, cui gli altri detentori di informazioni (catalografiche e non) sul patrimonio sono liberi di contribuire, in una architettura della conoscenza aperta e condivisa.

Al fine di garantire una ottimale pubblicazione delle informazioni contenute nelle varie tipologie di schede catalografiche, a novembre 2017 è stato avviato il progetto ArCo, con cui l’ICCD e il Semantic Technology Laboratory (STLab) dell’Istituto di Scienze e Tecnologie della Cognizione (ISTC) del CNR¹³ mirano a valorizzare il patrimonio artistico e culturale italiano attraverso l’individuazione (o la creazione, come vedremo) di una rete di ontologie che fornisca un’architettura della conoscenza dei beni culturali e consenta la pubblicazione di dati secondo il paradigma linked open data e che sia il più possibile integrata con OntoPiA.

Una pratica raccomandata da buona parte della letteratura sugli standard - in particolare in ambito *semantic web* e linked open data - è quella che prevede che, prima di creare una nuova ontologia, occorra valutare il riuso di ontologie disponibili per un determinato dominio di conoscenza. Ciò agevola l’interoperabilità semantica tra diverse basi di conoscenza.

Nell’individuare il modello ontologico più adeguato a render conto della complessità della metodologia catalografica e della ricchezza dei dati presenti nelle schede, si sono analizzate alcune ontologie del dominio¹⁴ dei beni culturali ed in particolare CIDOC-CRM, EDM, Cultural-ON, Dublin Core, DBPedia e Schema.org.

¹² Cfr.: <https://docs.italia.it/italia/piano-triennale-ict/pianotriennale-ict-doc/it/2019-2021/07_ecosistemi.html>.

¹³ Il team del CNR che lavora sul progetto ArCo è composto da Aldo Gangemi, Valentina Presutti, Luigi Asprino, Valentina Corriero, Andrea Nuzzolese, Margherita Porena, Alessandro Russo e Ludovica Marinucci. Per ICCD, responsabili scientifici del progetto sono Chiara Veninata e Maria Letizia Mancinelli.

¹⁴ La locuzione “di dominio” fa riferimento alla seguente distinzione a seconda delle finalità e dei confini tematici delle ontologie:

1) Upper-level ontologies (dette anche ontologie pesanti o ontologie fondazionali). Un’ontologia di alto livello rappresenta il tentativo di modellare l’intera realtà in un’unica tassonomia, da cui le specifiche ontologie ereditano al fine di classificare i diversi ambiti del sapere. Hanno un livello di astrazione molto elevato e sono spesso frutto di competenze multidisciplinari che coniugano le tecnologie dell’informazione con la filosofia analitica e l’epistemologia. Modellano i concetti fondazionali comuni a tutti i domini (spazio, tempo ecc.) e perciò sono adatte a scambiare informazioni attraverso domini diversi, dove i problemi di ambiguità terminologica acquistano fondamentale rilevanza. Le ontologie fondazionali sono sviluppate proprio allo scopo di facilitare la negoziazione dei significa-

Nell'ambito dei beni culturali, CIDOC-CRM è un'ontologia creata per la modellazione, l'integrazione e lo scambio di informazioni eterogenee riguardanti i beni culturali ed è stata ottenuta da un lavoro di oltre dieci anni da parte del CIDOC Documentation Standards Working Group¹⁵. CIDOC-CRM si pone in realtà a cavallo tra le ontologie c.d. fondazionali e le ontologie di dominio. L'ontologia è incentrata sulle relazioni tra l'oggetto culturale fisico e le relazioni tra esso e gli eventi che lo hanno visto coinvolto, nonché con le entità che hanno interagito con esso (persone, organizzazioni). CIDOC-CRM ha un approccio fortemente evento-centrico, focalizzandosi soprattutto sull'aspetto temporale delle attività e degli eventi, messi in relazione con la partecipazione di attori, luoghi, date e oggetti in generale. Ogni evento può essere concepito come un incontro tra risorse, in cui specifiche proprietà definiscono il tipo di partecipazione ad esso.

CIDOC-CRM non dà nessuna indicazione circa l'obbligatorietà di certe classi o proprietà, lasciando massima libertà nella modellazione dei dati, in modo da potersi adattare a qualunque requisito espressivo richiesto nei vari ambiti culturali per differenti oggetti culturali. Un'altra caratteristica di CIDOC-CRM è la struttura poli-gerarchica delle classi (e delle proprietà) che consente una grande libertà ed espressività nella modellazione dei dati. Il fatto che la questione terminologica venga demandata alla possibilità di utilizzare propri vocabolari, integrandoli con il modello ontologico, consente un ulteriore arricchimento semantico dei dati, senza sacrificare l'unicità del modello ontologico. I vocabolari stessi possono essere inoltre collegati tra loro per esplicitare dipendenze, corrispondenze e relazioni tra i termini in generale. Ad oggi CIDOC-CRM risulta una delle ontologie più utilizzate in ambito museale¹⁶.

ti dei termini di un certo dominio attraverso diverse comunità scientifiche o linguistiche.

2) Ontologie di dominio (*domain ontologies*). Si tratta di schemi concettuali che rappresentano la conoscenza di un dominio applicativo più o meno specifico, come possono essere gli archivi, le biblioteche, le opere d'arte, i beni archeologici ecc. Le ontologie utilizzate nei sistemi informativi sono di solito ontologie di dominio, ma di fatto i sistemi informativi basati su queste ontologie funzionano bene fintanto che le persone e le macchine condividono lo stesso vocabolario e il significato dei termini, come accade in comunità chiuse. Questo tipo di ontologie risulta essere debole relativamente a possibili fraintendimenti semantici quando gli utilizzatori del sistema informativo non condividono le stesse assunzioni sul significato dei termini del vocabolario. Per essere in grado di facilitare la negoziazione di significati, è estremamente utile che le ontologie di dominio contengano riferimenti semantici alle ontologie fondazionali, che possano chiarire i termini più specifici usati nelle ontologie di dominio.

¹⁵ A partire dal dicembre 2006 il modello concettuale è diventato uno standard ISO, e intende diventare il riferimento per la rappresentazione semantica di informazioni relative al patrimonio culturale. Nel 2014 l'ontologia è stata tecnicamente rivisitata a cura della commissione competente ISO / TC 46, "Informazione e documentazione", sottocomitato SC 4, "interoperabilità tecnica", in collaborazione con il Consiglio Internazionale dei Musei - Comitato per la Documentazione (ICOM CIDOC).

¹⁶ Nel tempo sono stati creati ulteriori CRM; nati come estensioni di CIDOC: FRBRoo, Bibliographic References, PRESSoo: Periodicals, CRMInf: Argumentation Model, CRMsci: Scientific Observation Model, CRMdig: Digital Provenance Model, CRMgeo: Spatiotemporal Refinement. È stato da poco rilasciato anche il profilo CRMarcheo a cura del team del Progetto Ariadne.

L'Europeana Data Model (EDM) è l'evoluzione – in chiave ontologica – dell'Europeana Semantic Elements (ESE), il modello di dati con cui nel portale Europeana viene gestita l'importazione dei dati relativi alle descrizioni di oggetti del patrimonio culturale. Tale ontologia fa riuolo diretto delle ontologie Dublin Core, DC Terms, OAI-ORE e stabilisce relazioni semantiche con CIDOC, Dolce-Lite, Abc, DCTems.

Cultural-ON è l'ontologia sviluppata nel 2015 dal MIBACT, insieme all'Istituto di Scienze e Tecnologie della Cognizione del Consiglio Nazionale delle Ricerche (ISTC-CNR) per la rappresentazione secondo il paradigma dei linked open data delle informazioni relative agli Istituti e luoghi della cultura e agli eventi culturali. Si articola in 56 classi e 90 proprietà e stabilisce relazioni semantiche con le ontologie Dolce, CIDOC, ORG, Event, Ticket ontology, LOD, FOAF, Fabio, Cito, Geonames, GoodRelation, PRO, SAN e con un set di ontologie di ISTAT. Di recente è stata rifattorizzata¹⁷ per integrarla nella rete di OntoPiA.

Dublin Core terms e refinements (DC e DCTerm) costituiscono dei data model leggeri che hanno lo scopo di descrivere attraverso un set minimo di metadati, una grande varietà di risorse in formati diversi. Dublin Core è estremamente diffuso a livello internazionale e il suo utilizzo è considerato un ottimo punto di partenza per l'interoperabilità tra sistemi diversi. Il nucleo (DC) è costituito da 15 proprietà cui è possibile affiancare dei "raffinamenti" (DCT) che consentono una descrizione delle risorse ancora più granulare¹⁸.

DBPedia ontology è una ontologia leggera e cross-domain, creata sulla base degli infobox più comunemente utilizzati all'interno di Wikipedia. L'ontologia mira a rappresentare i dati di DBPedia, che nasce per estrarre contenuti strutturati dalle informazioni create in vari progetti Wikipedia. Questa informazione strutturata assomiglia a un *open knowledge graph* (OKG), ovvero un tipo di database aperto che memorizza le conoscenze in una forma leggibile dalla macchina e fornisce un mezzo per raccogliere, organizzare, condividere, cercare e utilizzare dati e informazioni.

Schema.org è un vocabolario nato nel 2011 come tentativo di descrivere sul web qualsiasi tipologia di risorsa adoperando un linguaggio condiviso, per facilitare le ricerche. Fondatori del progetto Schema.org sono stati Google, Microsoft, Yahoo e Yandex. Nonostante la sua diffusione sia prevalente sui siti di e-commerce, da diver-

¹⁷ In ingegneria del software, il *refactoring* è una tecnica per modificare la struttura interna di porzioni di codice senza modificarne il comportamento esterno, applicata per migliorare alcune caratteristiche non funzionali del software quali la leggibilità, la manutenibilità, la riusabilità, l'estensibilità del codice nonché la riduzione della sua complessità. Cfr Martin Fowler, *Refactoring*, <https://refactoring.com/>

¹⁸ Il documento relativo alla trasposizione in RDF dello standard è una specifica aggiornata e autorevole di tutti i metadati gestiti dalla Dublin Core Metadata Initiative (DCMI). Sono inclusi i quindici termini del Dublin Core Metadata Element Set, che sono stati anche pubblicati come IETF RFC 5013 [RFC5013], ANSI / NISO Standard Z39.85-2007 [NISOZ3985] e ISO Standard 15836: 2009 [ISO15836].

so tempo il suo utilizzo si sta affermando in ambito internazionale anche per la descrizione del patrimonio culturale per aumentarne la visibilità sul web. In seno al W3C è stato creato lo Schema Bib Extend Community Group, col compito di preparare proposte per estendere Schema.org per una migliore rappresentazione e condivisione delle informazioni bibliografiche. L'iniziativa è complementare al lavoro che la Library of Congress conduce su BIBFRAME, ed è finalizzata non a rappresentare la ricchezza presente nei dati MARC ma a fare in modo che i dati bibliografici possano essere marcati in un modo che i motori di ricerca possano comprenderli e quindi utilizzarli appropriatamente. Sebbene infatti Schema.org non sia stato progettato per sostituire gli standard delle biblioteche, l'ontologia definisce un *reasonably coherent commonsense model* con classi e proprietà rilevanti per la descrizione delle risorse bibliografiche tra cui lavoro creativo, persona, autore, regista, luogo, organizzazione, editore, data di copyright, libro, ISBN e così via. Schema.org è proposto come un'impalcatura concettuale di massima a cui possono essere collegate ontologie più dettagliate progettate dalle varie comunità degli specifici domini.

A lungo il riuso di una ontologia, raccomandato tra le buone pratiche nell'implementazione dei linked data, è stato inteso come riuso diretto, ovvero adozione di un'ontologia già sviluppata per la modellazione dei propri dati. Nella pratica si tratta di fare riferimento in una ontologia A ad un'ontologia B tramite richiamo della sua URI. L'ontologia B contiene definizioni il cui significato diviene così parte integrante del significato dell'ontologia A. Poiché la proprietà è transitiva, se l'ontologia B ha importato a sua volta una terza ontologia C, anche tale ontologia C viene ad essere parte integrante dell'ontologia A. Ciò implica che, per un riuso diretto consapevole, è bene conoscere tutte le catene di importazione cui l'ontologia importata è eventualmente vincolata.

Esiste anche una seconda via per fare riferimento a ontologie preesistenti e ritenute pertinenti al proprio dominio, ed è quella che prevede il riuso indiretto delle ontologie connettendo esplicitamente la propria ontologia ad altre ontologie significative attraverso meccanismi di allineamento semantico (*ontology mapping*). Questa modalità presenta un duplice vantaggio, ovvero quello di mantenere autonomo il proprio schema concettuale da eccessive dipendenze da schemi concettuali mantenuti da agenti terzi su cui non si ha alcun potere di controllo e rendere la propria ontologia totalmente autoconsistente nel modellare un determinato dominio, con indubbi vantaggi anche per gli eventuali riutilizzatori dell'ontologia creata.

Con questa consapevolezza e poiché nessuna delle ontologie esaminate riusciva ad esprimere la complessa semantica sottesa alle trenta normative catalografiche rilasciate da ICCD, il team di ricerca ICCD-CNR ha stabilito di procedere al disegno di una nuova ontologia, pubblicata con il nome di "ArCo - Architettura della Conoscenza". ArCo si presenta come una rete di ontologie rilasciate dall'ICCD insieme all'Istituto di Scienze e Tecnologie della Cognizione (ISTC) del CNR tra il 2018 e il 2020 per la strutturazione della conoscenza per i beni culturali.

La rete di ontologie di ArCo – Architettura della Conoscenza

ArCo intende rispecchiare l'analiticità strutturale delle schede catalografiche elaborate dall'ICCD nonché valorizzare il contesto dei beni culturali e restituire la complessità dell'oggetto "bene culturale", rafforzando la ricchezza semantica dei dati tipicamente catalografici.

Le ontologie di ArCo si compongono di più moduli. Il modulo ArCo importa sia Cultural-ON che altre ontologie di OntoPiA. Gli altri moduli finora rilasciati riguardano la "context-description", il "cultural-event", la "denotative-description", il "catalogue record" e la "location".

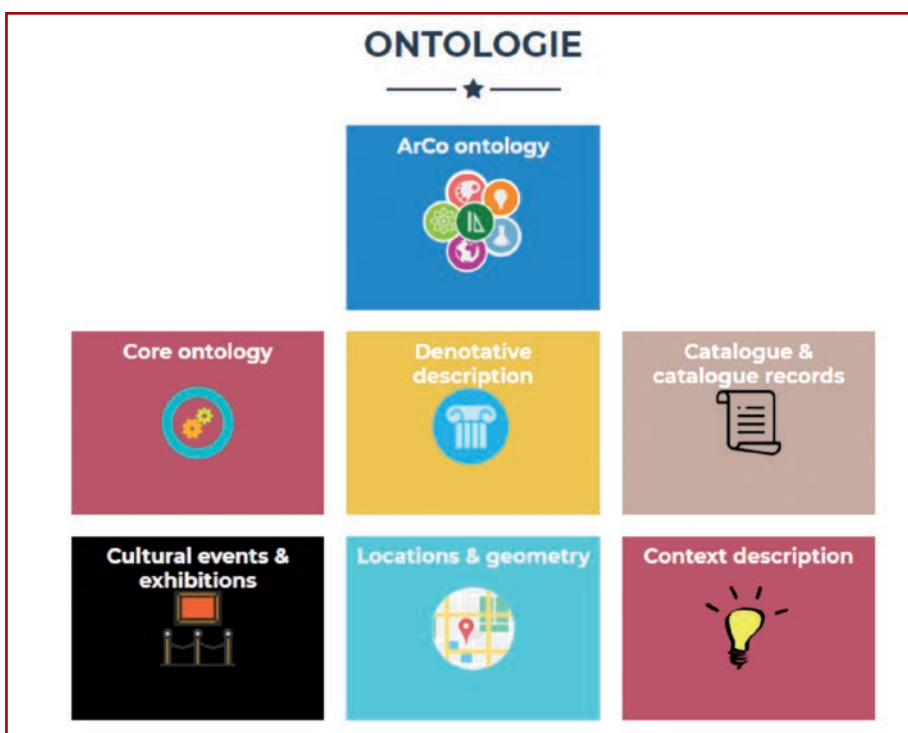


Figura 1. Le ontologie del progetto ArCo

ArCo viene rilasciato con licenza CC-BY-SA 4.0 sia su GitHub che sul sito ufficiale dei dati aperti del MiBACT. È accessibile online e può essere anche scaricato e installato localmente con un *docker*¹⁹. ArCo include:

- una rete di ontologie che modellano il dominio dei beni culturali con un alto livello di dettaglio;

¹⁹ *Docker* è un progetto *open-source* che consente di rilasciare una applicazione all'interno di un contenitore software fornendo anche una virtualizzazione del sistema operativo di Linux, evitando l'installazione e la manutenzione di una macchina virtuale. Cfr. <https://www.docker.com/>.

- un dataset di dati LOD che conta 169 milioni di triple, che descrivono oltre 800.000 beni culturali;
- un software per convertire automaticamente i record di catalogo conformi alle normative ICCD in LOD conformi ad ArCo, che consente aggiornamenti automatici e frequenti e facilita il riutilizzo delle ontologie;
- una documentazione dettagliata che riporta: (i) i requisiti ontologici espressi sotto forma di user story, le cosiddette domande di competenza, (ii) i modelli ontologici con diagrammi ed esempi di utilizzo;
- una serie di esempi che i potenziali utenti possono utilizzare come materiale di formazione: vi sono domande poste in linguaggio naturale e tradotte in corrispondenti query SPARQL, che possono essere testate direttamente con lo SPARQL endpoint del sito dati.beniculturali.it;
- una suite di test, implementata come file OWL e query SPARQL, utilizzata per convalidare il grafo della conoscenza di ArCo;
- un file di allineamento esterno tra le classi e le proprietà di ArCo e le classi e le proprietà delle ontologie di CIDOC, EDM, Schema, OntoPiA per dimostrare l'interoperabilità semantica tra i modelli ontologici.

La rete di ontologie ArCo è il risultato di un progetto di eXtreme Design (XD), una metodologia consolidata basata sul riutilizzo di modelli di progettazione ontologica (*ontology design patterns*, ODP).

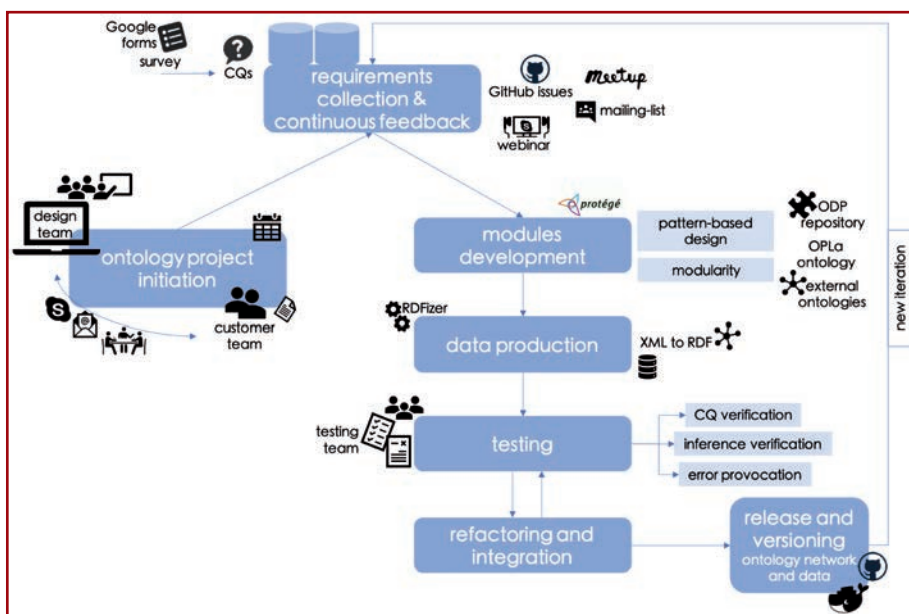


Figura 2. La metodologia eXtreme Design implementata per il knowledge graph di ArCo

Tale metodologia prevede un ciclo iterativo di test e feedback che coinvolge diversi attori:

- (i) un gruppo addetto alla progettazione dell'ontologia;
- (ii) un team di test, disgiunto dal team di progettazione, che si occupa di sperimentare l'ontologia;
- (iii) un team di attori interni ed esterni a ICCD, che esplicita i requisiti che vengono tradotti dal team di progettazione e dal team di test in impegni ontologici che guidano lo sviluppo dell'ontologia. Per rilevare eventuali incoerenze nell'ontologia, è stato eseguito regolarmente un ragionatore, Hermit, durante la fase di modellazione.

Tutte le proposte di miglioramento e le segnalazioni di malfunzionamento sono state raccolte su GitHub, su cui ArCo viene pubblicato in modo incrementale. Le comunità di sviluppatori e sperimentatori hanno interagito attraverso apposite mailing list, nonché partecipando a incontri in presenza e virtuali. Parallelamente allo sviluppo di moduli ontologici, sono sempre stati prodotti anche i dati in formato RDF²⁰ e a ciascuna risorsa RDF è stato assegnato un URI²¹. I dati sono stati infine collegati automaticamente a dati esterni usando il software LIMES, sulla base di euristiche stringenti che hanno garantito solo allineamenti verificabili.

Nella modellazione di ArCo, l'ICCD – insieme al CNR – ha coinvolto fin dall'inizio numerosi potenziali sviluppatori di applicazioni che usano le ontologie e i LOD prodotti nel settore dei beni culturali, includendo un programma c.d. di *early adoption*. Attraverso incontri periodici svolti con lo strumento dei *meetup*²² e il rilascio regolare e progressivo delle ontologie e dei dati – anche se in formato instabile – su Github, ICCD e STLab hanno voluto inaugurare una nuova stagione di partecipazione con gli *early adopter*, che hanno contribuito alla raccolta dei requisiti delle ontologie e allo sviluppo e miglioramento del progetto fornendo continui feedback.

²⁰ Ciò è avvenuto tramite lo sviluppo del software RDFizer: il suo componente principale è il convertitore XML2RDF, che prende in entrata un file XML conforme agli standard di catalogazione ICCD e lo trasforma in file in formato RDF attraverso fogli di stile XSLT.

²¹ L'URI per gli individui è <https://w3id.org/arco/resource/> con il prefisso data:. L'ID di ogni individuo è preceduto dal nome del suo tipo (ad esempio Agent, CulturalProperty ecc.). Per ogni tipo, sono stati attribuiti manualmente un insieme di elementi che costituiscono una possibile chiave (ad esempio AUTN, il nome dell'autore). Poiché i valori di questi elementi sono spesso stringhe, è stata rimossa la punteggiatura dai valori, sono stati convertiti in lettere minuscole, concatenati e ordinati in ordine alfabetico (ad esempio michelangelo-buonarroti). Infine è stato calcolato un MD5 checksum sulla stringa risultante, utilizzata come ID dell'URI, ad esempio dati: Agent/dcd4ca7b54dd3d7dac083dd4c54a9eef.

²² I meetup sono riunioni informali in cui si incontrano persone che hanno in comune un interesse culturale, tecnologico, sociale, ricreativo ecc. (dall'inglese "to meet up" che significa "incontrarsi per caso"). Si veda: <<https://www.meetup.com/it-IT/Meetup-Web-Progetto-ArCo-Architettura-della-conoscenza>>.

Conclusioni e prospettive

La rete di ontologie ArCo è stata dunque utilizzata per la modellazione dei dati del Catalogo generale dei beni culturali, gestita da ICCD e per la modellazione dei dati degli archivi fotografici di ICCD. Essa è stata utilizzata anche per la pubblicazione di LOD del Museo del Paesaggio sonoro²³ di Riva presso Chieri (TO) e per la pubblicazione dei LOD delle opere d'arte da parte dell'Istituto per i beni artistici culturali e naturali (IBACN) della Regione Emilia-Romagna. Anche la Regione Campania ha dichiarato di volere usare ArCo per i propri linked data nell'ambito del progetto Move to Cloud²⁴.

L'attività con StLab del CNR si è collocata in una fase in cui il Catalogo generale dei beni culturali online mostrava chiaramente alcune debolezze, riconducibili in buona parte ad una interfaccia in cui poca attenzione era data all'esperienza utente. La maggior parte degli utenti del Catalogo era infatti riconducibile a funzionari della pubblica amministrazione. Per ampliare l'utenza del sito e sperimentare i benefici del nuovo paradigma di pubblicazione dei LOD, ICCD ha deciso di ripensare il proprio sito di pubblicazione dei dati catalografici e ha promosso una *call to co-creation*, un contest a premi con cui sono stati coinvolti gli *early adopters* nella progettazione di modalità di fruizione innovative del sistema del Catalogo generale dei beni culturali. Alla gara hanno partecipato cinque aziende, tra cui una giuria di esperti ha selezionato la proposta vincitrice. Dalle ricerche fatte all'interno della piattaforma scelta e grazie ai LOD, l'utente potrà ottenere non solo la consultazione della risorsa digitale e dei dati descrittivi sul singolo elemento o gruppo di beni ricercato, ma anche la ricostruzione del contesto nel quale essi si collocano, evidenziando le relazioni esistenti fra gli elementi del patrimonio, i soggetti che li riguardano, i luoghi che ne sono lo scenario, le persone a cui sono legati, arricchendo le fonti di partenza con informazioni di qualità utili a fini conoscitivi, educativi, di ricerca, oltre che di valorizzazione. I dati del Catalogo mostreranno tutti i collegamenti verso altre basi di conoscenza (Regioni che pubblicano in LOD i dati sul proprio patrimonio, i LOD della Fototeca Zerì, Wikidata, Geonames, Europeana ecc.), ottenuti dinamicamente tramite procedure di riconciliazione semantica²⁵. I medesimi meccanismi sono stati utilizzati anche per collegare automaticamente i dati delle schede di catalogo con i dati presenti in quei *repository* a loro volta collegati a Wikidata (Enciclopedia Treccani, British Museum, VIAF ecc.). Infine, la riconciliazione è stata effettuata anche verso dati "interni", ovvero presenti nelle descrizioni inventariali degli archivi fotografici dell'ICCD, con l'obiettivo di dare or-

²³ <http://museopaesaggiosonoro.org/sound-archives-musical-instruments-collection-samic/>.

²⁴ Cfr. <<https://cultura.regione.campania.it/movetocloud>>.

²⁵ Con tale termine si intende fare riferimento ai meccanismi di allineamento basati sul concetto di identità (o di somiglianza) tra risorse nell'ambito del *semantic web*. Due risorse identiche a e b, condividono gli stessi attributi e le stesse relazioni. Se la risorsa a viene aggiornata (nei suoi attributi e nelle sue relazioni), di fatto si aggiorna anche la risorsa b. Con lo stesso significato, si parla anche di data linking, interlinking, data matching.

ganicità al complesso di dati catalografici e immagini fotografiche sul patrimonio che finora sono stati gestiti da ICCD su due banche dati separate.

Il passo successivo sarà l'integrazione dei LOD del Catalogo con ulteriori LOD sul patrimonio culturale, ad esempio della Regione Veneto²⁶, della Regione Emilia-Romagna²⁷, della Regione Campania²⁸ e gli open data della Regione Lombardia²⁹ e della Regione Sardegna³⁰: ciò consentirà, da una parte, di sperimentare l'interoperabilità tra i dati, superando i limiti dell'interoperabilità tra i sistemi, prefigurate già nel 2012 dalle *Linee guida per l'interoperabilità semantica*; dall'altra, contribuirà ad accrescere il *knowledge graph* sul patrimonio culturale e a fare un passo notevole verso la realizzazione del Catalogo nazionale del patrimonio culturale italiano, almeno in termini quantitativi.

In termini qualitativi, ICCD ha avviato, insieme all'Università di Bologna, il progetto ArCo4Science che, insieme alle collaborazioni con le università di Ferrara, La Sapienza, Roma Tre, mira all'implementazione del *knowledge graph* del patrimonio culturale italiano, integrando i dati del catalogo con i risultati della ricerca sul patrimonio culturale svolta in ambito universitario. Ciò avverrà, in particolare, attraverso la definizione ed esecuzione del processo di acquisizione e produzione di *knowledge graph* collegati al Catalogo generale dei beni culturali anche attraverso estrazione automatica di conoscenza da testo e attraverso la produzione e acquisizione di immagini e ricostruzioni 3D per la costruzione di ambienti virtuali semantici basati sul *knowledge graph* ArCo.

ICCD always aimed at sharing the models for structuring knowledge on cultural heritage and the data produced in cataloging campaigns. In recent years, the ICCD has focused its activities on the semantic web and its tools. One of the results is the ArCo project, the graph of knowledge of the Italian cultural heritage, consisting of a network of ontologies and more than 169 million triples referring to over 800 thousand catalog cards. ArCo is based on the data of the General Catalog of Cultural Heritage of the Central Institute for the catalog and documentation of MiBACT and on the data of its photographic archives. ArCo is jointly distributed with a SPARQL endpoint, a software to convert catalog records into RDF and a rich suite of documentation material (tests, evaluation, instructions, examples, etc.).

²⁶ <https://www.culturaveneto.it/it/footer/ontologia>.

²⁷ <https://ibc.regione.emilia-romagna.it/servizi-online/lod>.

²⁸ <https://cultura.regione.campania.it/ecosistema>.

²⁹ <https://www.dati.lombardia.it/Cultura/Opere-d-Arte-conservate-nei-Musei/dsrv-9ish>.

³⁰ <http://dati.regione.sardegna.it/dataset>.

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2020

RIFERIMENTI BIBLIOGRAFICI

Valentina Anita Carriero – Aldo Gangemi – Maria Letizia Mancinelli – Andrea Giovanni Nuzzolese – Valentina Presutti – Chiara Veninata. *Pattern-based design applied to cultural heritage knowledge graphs*. (2019). ArXiv, <<https://arxiv.org/abs/1911.07585>>.

Valentina Anita Carriero – Aldo Gangemi – Maria Letizia Mancinelli – Ludovica Marinucci – Andrea Giovanni Nuzzolese – Valentina Presutti – Chiara Veninata. *ArCo: the Italian Cultural Heritage Knowledge graph*. In: *The Semantic Web – ISWC 2019. ISWC 2019. Lecture Notes in Computer Science*. Cham: Springer, 2019, p. 36–52. <<https://arxiv.org/abs/1905.02840>>.

Valentina Anita Carriero – Aldo Gangemi – Maria Letizia Mancinelli – Ludovica Marinucci – Andrea Giovanni Nuzzolese – Valentina Presutti – Chiara Veninata. *ArCo Ontology Network and LOD on Italian Cultural Heritage*. In: *Proceedings of the First International Workshop on Open Data and Ontologies for Cultural Heritage co-located with the 31st International Conference on Advanced Information Systems Engineering (CAiSE 2019), Rome, Italy, June 3, 2019*. <<http://ceur-ws.org/Vol-2375/>>.

Martin Doerr. *The CIDOC Conceptual Reference Module An Ontological Approach to Semantic Interoperability of Metadata*. «AI Magazine», 24 (2003), n. 3, p. 75-92.

Gordon Dunsire – Corey Harper – Diane Hillmann – Jon Phipps. *Linked data Vocabulary Management: Infrastructure Support, Data Integration, and Interoperability*. «Information Standards Quarterly (ISQ)», 24 (Spring/Summer 2012), n. 2/3.

Claudio Gnoli – Carlo Scognamiglio. *Ontologia e organizzazione della conoscenza: introduzione ai fondamenti teorici dell'indicizzazione semantica*. Lecce: Pensa Multimedia, 2008.

Francesca Tomasi – Marilena Daquino. *Linked Cultural Objects: dagli standard di catalogazione ai modelli per il web of data. Spunti di riflessione dalla Fototeca Zerì*. «Umanistica digitale», (2017), 1, p. 29-43.

Giorgia Lodi – Luigi Asprino – Andrea Giovanni Nuzzolese – Valentina Presutti – Aldo Gangemi – Diego Reforgiato Recupero – Chiara Veninata – Annarita Orsini. *Semantic Web for cultural heritage valorization*. In: *Data Analytics in Digital Humanities*, Cham: Springer, 2017, p. 3-37. <https://link.springer.com/chapter/10.1007/978-3-319-54499-1_1>.

Maria Letizia Mancinelli – Chiara Veninata. *Architettura della conoscenza: il sistema ICCD come modello per la descrizione dei beni culturali*. In: *Conferenza GARR 2018: selected papers. Data (R)evolution. Cagliari, 3-5 ottobre 2018*. Roma: Associazione consortium GARR, 2019, p. 130-134. DOI: 10.26314/GARR-Conf18-proceedings.

Laura Moro. *Il Catalogo del patrimonio culturale italiano. Nuova centralità e prospettive future*. «Economia della Cultura», 25 (2015), n. 3-4, p. 419-432.

Laura Moro. *Dalla somiglianza alla sinergia. La descrizione del patrimonio culturale dalla specificità all'integrazione digitale*. Intervento presentato in occasione dell'Incontro di studio tenutosi presso la Sala convegni dell'Archivio storico della Presidenza della Repubblica: *Dalla somiglianza alla sinergia. La descrizione del patrimonio culturale dalle specificità all'integrazione digitale*. Roma, 17 ottobre 2017. <<http://www.iccd.beniculturali.it/getFile.php?id=6083>>.

Sandra Vasco Rocca. *Beni culturali e catalogazione. Principi teorici e percorsi di analisi*. Roma: Gangemi, 2002.

Chiara Veninata. *I dati aperti di ICCD: verso l'interoperabilità semantica attraverso le ontologie e i linked open data*. In: *Atti del Convegno ArcheFoss. Cagliari 7-9 ottobre 2016*. In corso di pubblicazione.

Progetti

Il progetto inDICEs: misurare l'impatto della cultura digitale

«DigItalia» 2-2020
DOI: 10.36181/digitalia-00014

Sara Di Giorgio - ICCU
Claudio Prandoni - AedekaSrl

inDICEs è un progetto di ricerca e innovazione finanziato dalla Commissione Europea nell'ambito del programma Horizon 2020, e coordinato dall'ICCU. La ricerca condotta da inDICEs permetterà di sviluppare una metodologia scientifica per misurare e valutare l'impatto economico della digitalizzazione del patrimonio culturale, analizzando le modalità di accesso ai beni e ai servizi culturali in Europa. A partire da questo, il progetto elaborerà delle raccomandazioni rivolte ai responsabili del settore dei beni culturali e delle industrie culturali e creative, per fornire loro strumenti utili ad affrontare le sfide poste dall'avvento del mercato unico digitale al fine di far prosperare e diffondere la cultura europea. Infine, il progetto svilupperà una piattaforma partecipativa e un sistema di autovalutazione, utile agli istituti culturali per migliorare il proprio posizionamento strategico nel mercato unico digitale, favorendone perciò la trasformazione digitale.

Introduzione

L'Unione Europea ha da tempo individuato il ruolo delle industrie culturali e creative (ICC) quali propulsori di innovazione economica e sociale. I settori culturali e creativi contribuiscono, infatti, pienamente allo sviluppo economico dell'Unione, generando occupazione e crescita, e sono quindi fondamentali per il futuro dell'Europa. La cultura e le espressioni culturali concorrono al benessere, alla cittadinanza attiva, ai valori comuni, all'inclusione sociale, allo sviluppo del dialogo interculturale e a un ambiente mediatico libero, pluralistico e diversificato.

La rivoluzione digitale ha un impatto travolgente sulle ICC: i giganti digitali come Google, Amazon, Apple e Facebook offrono nuovi modelli di business basati su contenuti completamente nuovi, che hanno cambiato radicalmente, a livello locale e globale, il mercato dei beni culturali. Questa rivoluzione non è ancora finita e le industrie culturali e creative, in Europa, hanno bisogno di strutture che consentano loro di essere competitive in un'epoca di progresso tecnologico sempre più pervasivo. Per questo motivo le ICC sono state poste al centro della nuova agenda politica europea¹ in linea con i fabbisogni economici del settore e nella prospettiva indotta

¹ Si veda il Piano di lavoro per la cultura 2019-2022, adottato il 27 novembre 2018: <<http://data.consilium.europa.eu/doc/document/ST-13948-2018-INIT/en/pdf#http://data-consilium.europa.eu/doc/document/ST-13948-2018-INIT/en/pdf>>.

dai cambiamenti digitali, predisponendo il programma “Cultura” in modo da adeguare i fabbisogni del settore culturale e creativo all’era digitale, attraverso un approccio più pragmatico e globale.

Il settore dei Beni Culturali è un fattore chiave per lo sviluppo delle ICC in Europa, non solo perché fornisce l’accesso a grandi quantità di contenuti culturali riutilizzabili, ma perché può essere visto come un laboratorio di ricerca e sviluppo dell’ecosistema culturale e creativo attraverso il quale è possibile contribuire in modo significativo al progresso economico del settore delle ICC e alla società in generale.

Tuttavia è ancora necessario comprendere appieno il ruolo cruciale che il settore dei beni culturali può svolgere nel processo di digitalizzazione dell’ecosistema culturale e di sviluppo dell’economia in Europa, identificando i quadri di riferimento per favorire lo sviluppo del settore e per misurare l’impatto delle iniziative e delle politiche adottate. D’altra parte nei progetti di digitalizzazione musei, biblioteche e archivi sono spesso frenati dalle difficoltà che spesso si incontrano nella gestione dei diritti d’autore e necessitano di linee guida per capire appieno il quadro normativo nazionale ed europeo e per applicare le disposizioni legali, senza rinunciare però alle opportunità che si aprono grazie all’adozione di una politica di apertura e condivisione dei contenuti digitali. È necessario, inoltre, offrire agli istituti culturali degli indicatori idonei per misurare gli effetti delle politiche e delle iniziative adottate sugli individui, sulla collettività o sul territorio, in modo massimizzare il loro impatto.

In questo contesto si articola inDICES², un progetto di ricerca e innovazione finanziato dalla Commissione europea nell’ambito del programma Horizon 2020 in risposta al bando “Digitisation, Digital Single Market and European culture: new challenges for creativity, intellectual property rights and copyright”³ con l’obiettivo di fornire ai decisori politici e ai responsabili nel settore dei beni culturali e delle industrie culturali e creative (ICC) degli strumenti per comprendere appieno l’impatto sociale ed economico della digitalizzazione del patrimonio culturale, valutando in particolare l’importanza dei contenuti “aperti” per lo sviluppo di nuovi business model improntati sul ri-uso creativo dei beni culturali e analizzando i problemi che derivano dalla gestione del diritto d’autore.

La ricerca condotta da inDICES, coordinata dall’ICCU, permetterà di sviluppare una metodologia scientifica per misurare e valutare l’impatto economico della digitalizzazione del patrimonio culturale, analizzando le modalità di accesso ai beni e ai servizi culturali in Europa. Il progetto elaborerà inoltre uno studio comparativo della normativa nazionale ed europea sul diritto d’autore e delle raccomandazioni rivolte

² <https://indices-culture.eu/>.

³ Si veda il testo del bando al link: <<https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/dt-governance-13-2019>>.

a musei, biblioteche e archivi per aiutare gli istituti culturali a superare gli ostacoli allo scambio dei nuovi beni e servizi digitali e a cogliere le sfide che si pongono con il mercato unico digitale per far prosperare e diffondere la cultura europea.

Le raccomandazioni presenteranno buone pratiche e linee guida per lo sviluppo di strategie e business model innovativi per cogliere le nuove opportunità che si offrono grazie all'uso delle tecnologie digitali nei beni e servizi culturali e per stimolare l'interesse e la partecipazione di pubblici diversi anche alla cura delle risorse culturali, con particolare attenzione ai giovani e alle interazioni con le ICC e con il mondo della didattica e della ricerca. Assieme alle raccomandazioni, il progetto svilupperà un sistema di autovalutazione, utile agli istituti culturali per migliorare il posizionamento strategico nel mercato unico digitale e favorendone perciò la trasformazione digitale.

inDICEs realizzerà, inoltre, un osservatorio europeo online dotato di una piattaforma web partecipativa dove diverse categorie di stakeholder potranno informarsi, collaborare, discutere, contribuire allo sviluppo di policy e iniziative condivise e, successivamente, monitorare lo sviluppo delle strategie per l'accesso ai beni e ai servizi per il patrimonio culturale. Uno strumento di visualizzazione dei dati, basato su tecnologie di analisi intelligente di Big data, identificherà e visualizzerà le tendenze culturali emergenti nei canali social online e supporterà i responsabili degli istituti culturali e i decisori politici ad identificare la rilevanza dei contenuti

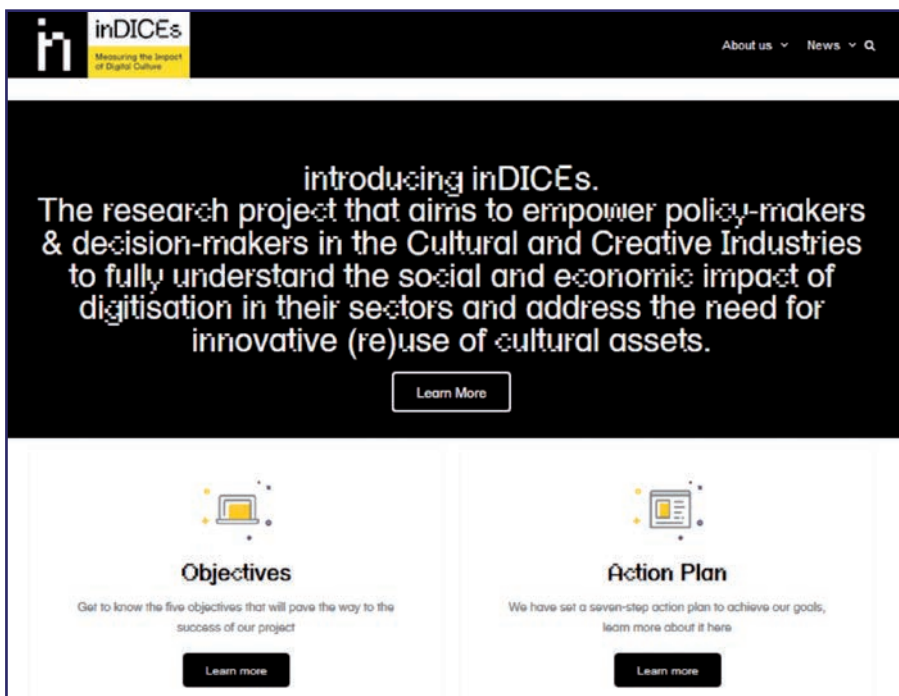


Figura 1. Home page del sito di progetto

generati dagli utenti, offrendo perciò degli indicatori affidabili per comprendere l'impatto cross-mediatico e la reputazione delle iniziative nell'ambito del patrimonio culturale e tracciando politiche e tendenze a lungo termine.

Grazie al progetto, i decisori politici potranno disporre di una solida struttura per valutare l'impatto del patrimonio culturale e di una piattaforma web per tenere traccia dell'evoluzione dell'impatto del patrimonio culturale nella società in Europa. Gli istituti culturali potranno prendere decisioni strategiche che consentiranno loro di migliorare il loro contributo verso le industrie culturali e creative e la società, partecipando attivamente al mercato unico digitale (Digital Single Market - DSM).

Il consorzio

inDICEs riunisce un consorzio internazionale di 14 organizzazioni, provenienti da 9 diversi paesi europei, che riuniscono un insieme di competenze multidisciplinari e complementari. Il partenariato, coordinato dall'Istituto Centrale per il Catalogo Unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU) comprende:

- Istituti di ricerca con una presenza consolidata nel campo dei beni culturali digitali, delle scienze sociali e della normativa sui diritti di proprietà intellettuale, quali: la Katholieke Universiteit Leuven, che partecipa a inDICEs con due gruppi di ricerca: il Centro per il diritto dell'informatica e della proprietà intellettuale e il Laboratorio di studi culturali digitali; la Fondazione Bruno Kessler, un istituto di ricerca italiano, leader internazionale per l'eccellenza scientifica nei campi dell'ICT, della sociologia e dell'economia; il PIN, il rinomato consorzio di ricerca creato dall'Università di Prato, con un solida esperienza nel campo delle applicazioni digitali ai Beni Culturali.
- Organizzazioni non governative con una sostanziale capacità di sensibilizzazione dell'opinione pubblica: la fondazione polacca Centrum Cyfrowe, impegnata in iniziative per sensibilizzare gli istituti pubblici ad utilizzare le licenze aperte come strumento chiave per la promozione e conservazione del patrimonio culturale; la Rete delle organizzazioni museali europee (NEMO), rappresentata dal Deutscher Museumsbund; la Fondazione Europea, operatore della piattaforma e dei servizi dell'infrastruttura Europea; Cluj Cultural Centre, una fondazione rumena che opera in ambito culturale sviluppando progetti per la trasformazione sociale e lo sviluppo urbano.
- Rappresentanti delle industrie culturali e creative e di network nazionali e paneuropee, tra cui: l'Istituto olandese per il suono e la visione; l'Associazione europea per il patrimonio culturale della moda; l'Associazione culturale Michael, una rete di istituti del patrimonio culturale; Capital High Tech, una società francese specializzata nello sviluppo di modelli di business innovativi e con una grande esperienza nel campo della misurazione dell'impatto socio-economico e dell'analisi della catena del valore.

- Società specializzate nello sviluppo di applicazioni innovative, quali: Platoniq, una società con sede a Barcellona che sviluppa piattaforme collaborative per permettere la partecipazione sociale e democratica dei cittadini e per co-progettare strumenti e metodologie digitali per la didattica, l'economia e l'innovazione sociale; WebLyzard, una società austriaca che si occupa di ricerca e sviluppo nel campo della Web intelligence applicata ai big data estratti dai social media per tracciare e visualizzare le tendenze.



Figura 2. Il consorzio inDICES

La metodologia di ricerca

Per comprendere gli effetti della rivoluzione digitale sui modi di produzione culturale e creativa e individuare dei parametri qualitativi e quantitativi e per misurarne l'impatto economico e sociale, è stato adottato un quadro di riferimento generale Cultura 3.0 (Sacco et al., 2017⁴) che distingue tre regimi di produzione e partecipazione culturale:

- Mecenatismo (Cultura 1.0), oggi applicato per lo più a settori non prevalentemente orientati al mercato come le arti visive, le arti dello spettacolo, i musei e i beni culturali;
- Industrie culturali e creative (Cultura 2.0), applicato a forme industrializzate di produzione culturale e creativa basate sulla distinzione strutturata tra produttori e pubblico;
- Piattaforme digitali aperte (Cultura 3.0), dove la produzione e la distribuzione di contenuti non è necessariamente mediata dal mercato e la distinzione tra produttori e utenti diventa sempre più sfumata.

⁴ Si veda: Pier Luigi Sacco — Emanuele Teti, *Cultura 3.0: un nuovo paradigma di creazione del valore*, «Economia & Management», (2017) 1, p. 79-95, <https://www.researchgate.net/publication/313843292_Cultura_30_un_nuovo_paradigma_di_creazione_del_valore>.

I settori non industriali come le arti visive e dello spettacolo, e in particolare i musei, le biblioteche e gli archivi (audiovisivi), sebbene meno orientati al mercato ed essenzialmente basati su sovvenzioni pubbliche per la loro redditività economica, svolgono un ruolo chiave nella conservazione e nella trasmissione della conoscenza e nell'esplorazione di nuove idee innovative. Infatti, la loro esposizione relativamente limitata al mercato sottolinea l'originalità e la creatività trasformativa che viene poi adottata e ulteriormente elaborata dai settori culturali e creativi più orientati al mercato.

L'impatto del Mercato Unico Digitale sul settore del patrimonio culturale, dove la riproducibilità digitale dei contenuti non fa parte del modello di core business ma è uno strumento per facilitare la circolazione e l'utilizzo spesso al di fuori del contesto di mercato, è un argomento particolarmente interessante e impegnativo e la letteratura disponibile è molto limitata. Il progetto inDICEs si pone quindi come un'esplorazione pionieristica in questo senso, coinvolgendo esperti economisti della cultura, per gettare le basi di un paradigma concettuale innovativo per analizzare la questione e le sue principali implicazioni politiche, compreso il rapporto tra il settore dei beni culturali e le ICC dal punto di vista del DSM.

inDICEs intende fornire strumenti e modelli per analizzare e misurare l'impatto sociale della cultura digitale, contribuendo alla democratizzazione delle risorse culturali e dimostrando i benefici sociali della cultura aperta. Ad oggi, infatti, i modelli per la valutazione d'impatto sono ancora troppo spesso l'eccezione piuttosto che la regola in campo culturale, e inDICEs mira a stabilire un nuovo standard attraverso un attento mix di tecniche qualitative e quantitative mirate a rispondere alle domande e alle esigenze specifiche di esperti e professionisti, anche attraverso la raccolta partecipativa di dati e la ricerca (approccio di co-creation).

L'analisi condotta dal progetto mira ad individuare i parametri rilevanti per effettuare una corretta valutazione della competitività degli istituti culturali che vogliono entrare nel Digital Single Market, con servizi e soluzioni innovativi. L'obiettivo principale del progetto è di fornire una metodologia per misurare l'impatto della digitalizzazione sull'accesso ai beni e servizi culturali europei. Tuttavia, l'elaborazione di tale metodologia richiede che le parti coinvolte nel processo di digitalizzazione siano pienamente informate delle opportunità offerte dal quadro normativo, sulla proprietà intellettuale che regola il processo di digitalizzazione. Biblioteche, musei e archivi devono essere in grado di sviluppare strategie competitive per posizionarsi all'interno del complesso quadro giuridico. Le linee guida che saranno sviluppate dal progetto offriranno un quadro chiaro per la comprensione e la gestione delle questioni legali sulla proprietà intellettuale e per la pubblicazione dei contenuti aperti e riutilizzabili.

La metodologia di ricerca, su cui si basa il progetto, è organizzata in tre fasi distinte: la raccolta dei dati, la creazione dei modelli, l'analisi e l'interpretazione. La figura qui sotto descrive sommariamente questa metodologia. I dati saranno raccolti sia da banche dati e archivi esistenti sia attraverso specifiche attività di raccolta dati derivati da piattaforme web e social media.

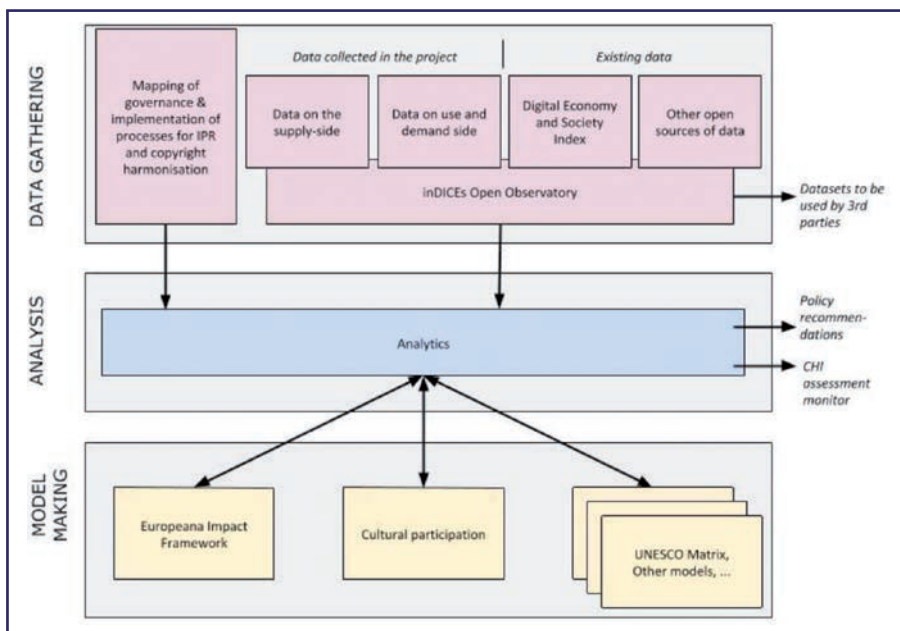


Figura 3. La metodologia di ricerca di inDICES

La fase di raccolta dati si svolgerà su più fronti:

- Mappatura delle legislazioni pertinenti in materia di Intellectual Property Rights (IPR) e diritto d'autore, in particolare quella relativa alla protezione del copyright, in tutte le legislazioni europee. Questa attività di raccolta e di mappatura della normativa permetterà di individuare analogie e differenze tra i vari paesi. La mappatura sarà organizzata scegliendo un numero selezionato di variabili sufficienti a descrivere la variazione osservata negli approcci nazionali, in modo da renderli direttamente comparabili. La mappatura e la classificazione si concentrerà in particolare sulle variazioni relative allo status giuridico dei "surrogati digitali" e saranno valutate le variazioni in termini di eccezioni e limitazioni che consentono il riutilizzo o l'uso creativo dei contenuti digitali.
- Raccolta di dati da fonti online e offline. La raccolta dei dati sarà effettuata scegliendo un numero adeguato di fonti online e offline da analizzare in base alle specifiche esigenze di ricerca. Ad esempio, inDICES raccoglierà dati dai social media per analizzare in che misura le persone su una specifica piattaforma producono e condividono un certo tipo di contenuti. In questo tipo di raccolta saranno considerate quattro diverse tipologie di dati:
 - a. Supply-Side data, ovvero dati generati da specifiche istituzioni e professionisti che operano sistematicamente nella produzione e diffusione di contenuti culturali e creativi.

b. Demand Side data, cioè dati generati da soggetti che sono principalmente impegnati nell'accesso e nella fruizione di contenuti culturali e creativi. È da notare che gli stessi soggetti possono posizionarsi sia sul lato dell'offerta che su quello della domanda, a seconda delle circostanze e del loro specifico ruolo all'interno dell'ecosistema culturale e creativo.

c. Dati relativi al cambiamento comportamentale. I dati raccolti dai social media saranno analizzati semanticamente. Ciò consentirà l'estrapolazione di informazioni sui tratti latenti della personalità (Big 5⁵) per analizzare come la "psicologia collettiva" della comunità di utenti sia influenzata dall'accesso digitale ai contenuti culturali e creativi e dal tipo di contenuto in particolare. Sarà quindi possibile seguire l'evoluzione del *sentiment* e gli orientamenti degli utenti come conseguenza dell'esposizione a determinati contenuti.

d. Dati sull'impatto economico. Particolare attenzione sarà prestata a specifiche fonti di dati e misurazioni già esistenti, come ad esempio l'indice di digitalizzazione dell'economia e della società⁶, Eurostat⁷, il portale Open Data dell'UE⁸, i dati UNCTAD sull'economia creativa⁹. Sarà effettuata una meta-analisi sistematica di queste fonti di dati statistici esistenti.

– Fonti letterarie e archivistiche. Verranno raccolti e analizzati i dati creati da istituti statistici e altre istituzioni di raccolta dati, nonché la letteratura scientifica riguardante gli effetti della rivoluzione digitale sulla produzione, diffusione e accesso ai contenuti culturali e creativi.

– Estrazione e visualizzazione dei metadati. La piattaforma InDICES grazie alla tecnologia di analisi di big data, permetterà la visualizzazione di cluster di dati, sulla base dei parametri messi a punto da InDICES.

La fase di modellazione si svilupperà partendo da una revisione della letteratura esistente nel settore beni culturali, con particolare riferimento agli effetti della rivoluzione digitale, e terrà conto delle specifiche problematiche sollevate durante la fase di raccolta dati. In particolare, il modello di InDICES sarà progettato analizzando i principali framework di riferimento a livello Europeo: European Impact Framework¹⁰, UNESCO Culture for Development Indicator¹¹ (CDIS) e Partecipazione Culturale 3.0. Questi framework aiuteranno a stabilire i metodi di analisi, la struttu-

⁵ [https://it.wikipedia.org/wiki/Big_Five_\(psicologia\)](https://it.wikipedia.org/wiki/Big_Five_(psicologia)).

⁶ Si veda: <<https://ec.europa.eu/digital-single-market/desi>>.

⁷ Si veda: <<https://ec.europa.eu/eurostat/home>>.

⁸ Si veda: <<https://data.europa.eu/euodp/it/data/>>.

⁹ Si veda: <<https://unctad.org/en/Pages/DITC/CreativeEconomy/Statistics-on-world-trade-in-creative-products.aspx>>.

¹⁰ Un gruppo di lavoro formato nell'ambito dell'Associazione del Network di European Impact Framework aggiornerà l'European Impact Framework grazie ai risultati ottenuti da InDICES. Si veda: <<https://pro.europeana.eu/project/impact-lite>>.

¹¹ Si veda: <<http://www.unesco.org/new/en/culture/themes/cultural-diversity/cultural-expressions/programmes/culture-for-development-indicators/>>.

ra logica, le relazioni e le limitazioni che definiranno i modelli di compilazione, accesso e analisi dei dati.

Infine, la fase di analisi e interpretazione sarà il risultato dell'applicazione coerente dei modelli elaborati ai dati disponibili al fine di ricavare indicatori sintetici per misurare i fenomeni più rilevanti. Diverse tecniche saranno impiegate per analizzare le diverse fonti di dati e le specifiche basi di dati create nelle fasi precedenti. Molte tecniche oggi consentono di fare inferenze sofisticate sulla struttura delle connessioni e dell'interazione tra attori diversi che operano in uno stesso campo. Ma anche i modelli socio-epidemici di contagio comportamentale sono sempre più utilizzati per descrivere la diffusione di determinati tipi di comportamenti in una data popolazione, a partire da un impulso originario. In questa fase saranno effettuate anche analisi quantitative e qualitative sulla relazione tra gli accordi istituzionali a livello nazionale ed Europeo in materia di protezione dell'IPR e del diritto d'autore e sui modelli relativi alla produzione, circolazione e accesso ai contenuti culturali e creativi. Quest'ultima fase prevede le seguenti attività:

- Il primo aspetto consiste nell'elaborazione di un toolbox che potrà essere messo a disposizione dei ricercatori, professionisti e decisori politici che operano non solo nel settore del patrimonio culturale, ma più in generale nell'intera sfera della produzione culturale e creativa. Verranno utilizzati strumenti tratti dalla psicologia sociale e dall'economia comportamentale, in particolare per lo studio degli atteggiamenti motivazionali e disposizionali delle persone verso fenomeni come la partecipazione culturale attiva e passiva, la partecipazione ai processi di co-creazione socialmente mediati e ad attività formative e didattiche.
- Il secondo aspetto delle attività di analisi è quello di fornire interpretazioni dei risultati, a partire da un'attenta analisi degli indicatori generati mediante l'applicazione del toolbox. Le interpretazioni saranno soggette a un attento esame e convalida da parte dell'Advisory Board del progetto, composto da otto esperti internazionali del settore.
- L'ultima parte dell'analisi sarà la preparazione di una serie di raccomandazioni politiche volte a migliorare la capacità del settore dei beni culturali di generare valore sociale ed economico facilitando la produzione, la diffusione e l'accesso a contenuti culturali e creativi digitali o digitalizzati. Queste raccomandazioni saranno testate e validate dall'Advisory Board che instaurerà in un dialogo costante con la Commissione europea e con altre organizzazioni internazionali attive nella promozione dello sviluppo economico e sociale attraverso la cultura come OCSE, Creative Commons e UNESCO.

Tutte le varie fasi descritte sopra si svolgeranno seguendo un approccio co-creativo, che prevede il coinvolgimento dei partner attraverso la piattaforma partecipativa sviluppata dal progetto. Seguendo questa metodologia sarà possibile misurare e valutare l'impatto della digitalizzazione su tutto il ciclo di vita dei contenuti, dal-

la produzione alla diffusione e all'accesso e all'eventuale riuso creativo. Ciò permetterà agli istituti culturali di valutare i vincoli legali sulle loro collezioni e di identificare i rischi e le strategie di mitigazione per partecipare attivamente allo sviluppo del Digital Single Market.

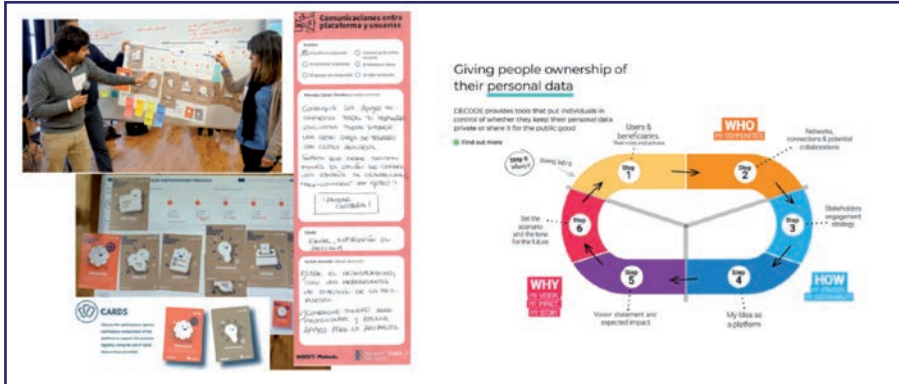


Figura 4. L'approccio di co-creation in inDICES

L'Osservatorio Aperto

inDICES svilupperà una piattaforma online che fungerà da Osservatorio Aperto (*Open Observatory*). Tale piattaforma metterà a disposizione dei decisori politici una serie di strumenti semplici da consultare ed utilizzare per supportarli nello sviluppo di strategie volte a favorire la trasformazione digitale di musei, biblioteche e archivi, mettendoli in grado di superare i problemi che si pongono nell'uso creativo di risorse culturali.

L'Osservatorio Aperto si rivolge ad un'ampia gamma di utenti (istituzioni culturali, decisori politici, agenzie di finanziamento, ricercatori, reti di professionisti ecc.) con lo scopo di sperimentare lo sviluppo di strategie e politiche condivise attraverso processi partecipativi. Si basa sull'analisi dei dati raccolti durante il progetto e

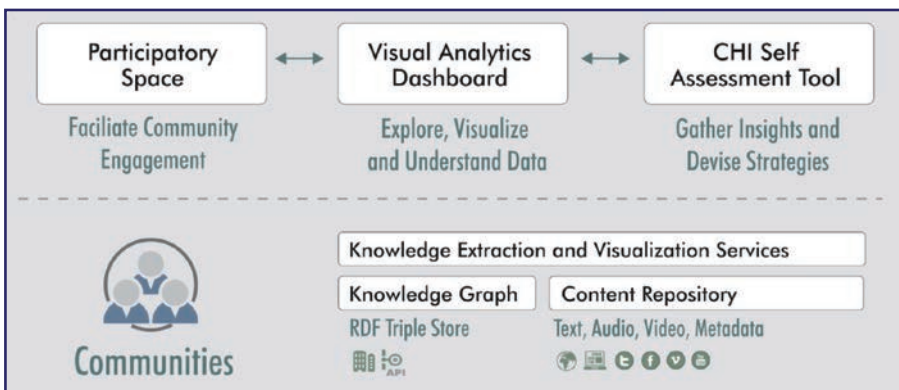


Figura 5. Le componenti dell'Osservatorio Aperto di inDICES

verrà sviluppato con un design creativo e funzionale che faciliterà il processo decisionale. Una delle caratteristiche più innovative dell'osservatorio è la creazione di un ambiente che favorisce la partecipazione e il dialogo continuo della comunità di esperti e utenti. La figura precedente illustra le tre componenti principali dell'Osservatorio Aperto: la *Visual Analytics Dashboard*, lo spazio partecipativo e lo strumento di autovalutazione per gli istituti culturali.

La *Visual Analytics Dashboard* è lo strumento che consentirà di esplorare l'archivio dei contenuti, tracciare le tendenze recenti, confrontare diversi set di dati e contestualizzarli lungo le dimensioni dei metadati per comprendere meglio gli schemi che stanno alla base. Questo strumento si basa sul framework di estrazione e visualizzazione dati sviluppato da WebLizard sfruttando tecnologie di analisi dei big data, che sono documentate in diverse pubblicazioni che trattano i temi dell'estrazione della conoscenza (Weichselbraun et al., 2017), e di elaborazione della conoscenza legata a procedure statistiche (Braşoveanu et al., 2016) e di analisi visiva (Scharl et al., 2017). Il dashboard di inDICES sarà arricchito grazie all'Observatory API Framework che permetterà l'accesso e il riutilizzo dei dati raccolti da terzi. Questo insieme di API saranno progettate per fornire accesso non solo ai dati raccolti, ma anche ad ulteriori metadati creati ad hoc dal progetto come entità, sentimenti, relazioni, annotazioni testuali e dati statistici.

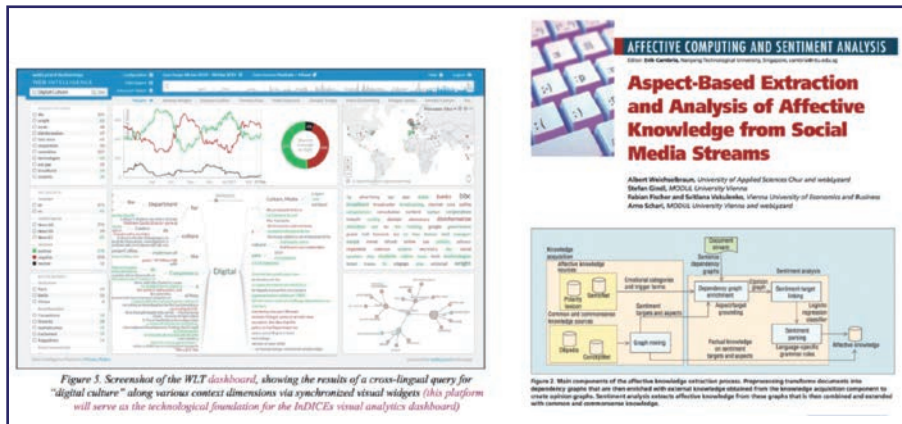


Figura 6. La *Visual Analytics Dashboard*

Lo spazio partecipativo¹² servirà per orchestrare il coordinamento delle attività di ricerca e di raccolta dati da parte dei partner del consorzio e sarà progettato in modo tale da facilitare il coinvolgimento della comunità e delle parti interessate, stimolando gli utenti a contribuire attraverso sondaggi, dibattiti e progetti collaborativi che promuovano l'innovazione nel settore e aiutino a formulare priorità

¹² <https://participate.indices-culture.eu/>.

condivise e a connettere diversi gruppi di stakeholders. Costruito utilizzando un software libero e aperto, lo spazio partecipativo fornirà un'infrastruttura trasparente per condividere esperienze, materiali e risorse di ogni tipo.

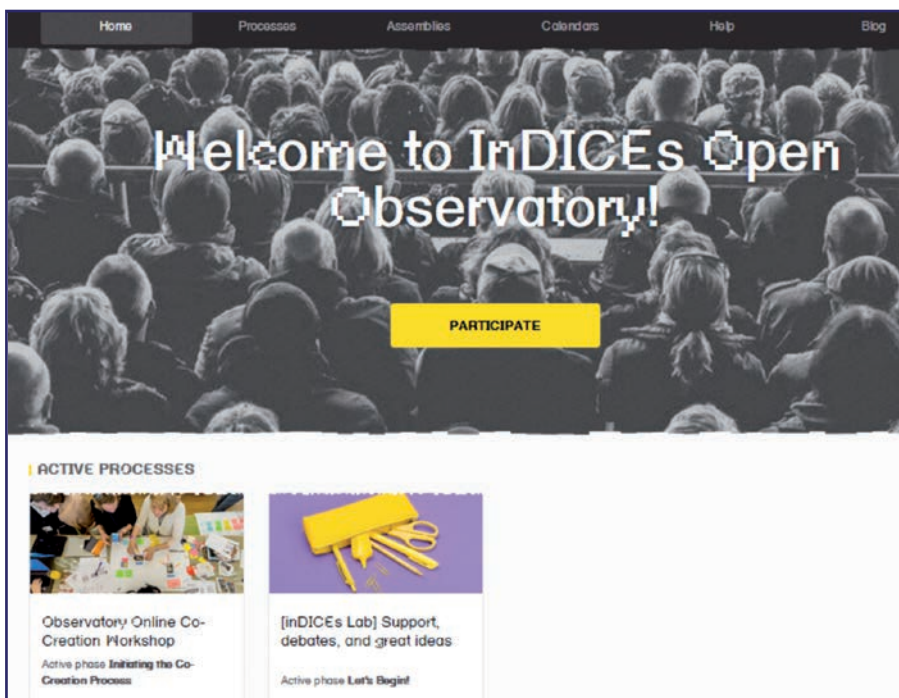


Figura 7. Lo spazio partecipativo di inDICES

Infine, verrà integrato nell'Osservatorio Aperto uno strumento di autovalutazione per biblioteche, musei e archivi per verificare la loro *readiness-to-market*. Questo strumento, basato sull'analisi delle politiche e dei criteri proposti per misurare il grado di innovazione digitale, permetterà agli istituti culturali e ai professionisti del settore di capire quali cambiamenti sono necessari nel loro lavoro per favorire un migliore accesso alla cultura e lo sviluppo di nuovi contenuti creativi.

inDICES e la crisi Covid-19

I nuovi scenari, introdotti forzatamente a causa della crisi originata dalla pandemia Covid-19, hanno avuto un impatto notevole sulle modalità di produzione, accesso e condivisione delle risorse culturali digitali, offrendo punti di riflessione interessanti e importanti di cui inDICES terrà conto.

In primo luogo, con l'esplosione della crisi Covid-19 e le conseguenze derivate dal blocco delle attività e dal distanziamento sociale sul funzionamento delle istituzioni culturali, si è potuto osservare il ruolo chiave della dimensione digi-

tale nell'accesso e ri-uso dei contenuti culturali e creativi. Il cambiamento forzato che la crisi ha provocato permette di analizzare in che modo l'esperienza culturale in ambito digitale abbia portato ad un impatto significativo sulle modalità di accesso alla cultura. La crisi Covid-19 rappresenta un esperimento naturale per inDICES, perché sono ben definite le fasi "prima" e "dopo" che possono essere analizzate comparativamente nelle loro conseguenze comportamentali e strutturali sul funzionamento degli ecosistemi culturali. In particolare, il progetto mira a capire se, nel nuovo contesto, un pubblico già caratterizzato da alti livelli di accesso e partecipazione culturale, incrementa ulteriormente l'accesso alla cultura attraverso la rete e, in tal caso, fino a che punto e in che modo, e come le istituzioni culturali stiano sviluppando nuove strategie di valorizzazione dei contenuti digitali. Inoltre, è utile comprendere se il nuovo scenario digitale attira anche un nuovo pubblico, intimidito dall'accesso fisico alle istituzioni culturali o più in generale scarsamente interessato alla fruizione culturale. Sarà istruttivo capire quali modalità di accesso questo secondo gruppo di utenti trova particolarmente utili e perché. Infine, sarà interessante capire se il nuovo scenario incoraggia lo sviluppo di modelli più proattivi, basati sulla produzione di contenuti generati dagli utenti su una scala più ampia e organizzata, e in che misura queste dinamiche sono mediate o favorite da specifiche strategie di engagement come la gamification.

Questi risultati possono servire per comprendere quali sono le tendenze a lungo termine nell'accesso culturale e per aiutare le istituzioni culturali a progettare interfacce e formati digitali più efficaci e di vasta portata.

In secondo luogo, durante l'attuale crisi Covid-19, assistiamo ad un aumento dei contenuti culturali digitali condivisi su internet. La maggior parte dei contenuti viene offerta da istituzioni culturali: musei, biblioteche ma anche archivi audiovisivi e cineteche. Inoltre, alcuni istituti culturali, ma anche editori e industrie creative, condividono parte delle loro opere fuori commercio o in pubblico dominio. Questa situazione evidenzia la volontà delle istituzioni e delle organizzazioni culturali di fornire accesso online alle loro collezioni su base gratuita e promuoverà, sicuramente, il modello di condivisione dei contenuti online anche dopo la crisi. Inoltre, i cittadini saranno sempre più consapevoli dell'importanza di avere accesso ai contenuti culturali in rete (gratuitamente o tramite abbonamento). Tuttavia, dal punto di vista del diritto d'autore, ciò comporta anche alcuni problemi. Innanzitutto, per quanto riguarda la territorialità intrinseca del regime del diritto d'autore. Anche se le istituzioni culturali sono disposte a condividere le loro collezioni online, nella maggior parte dei casi i confini territoriali rivestono ancora un ruolo chiave, rendendo inaccessibili alcuni contenuti per il pubblico situato in un paese diverso da quello in cui si trova l'istituzione. Ai fini del progetto inDICES, sarà interessante analizzare le difficoltà di fornire un accesso transfrontaliero ai contenuti digitali nell'UE e identificare i

miglioramenti che la nuova direttiva DSM presenta per favorire l'accesso di determinate opere.

A questo proposito, stiamo assistendo a come iniziative private come Google Arts sono in grado, attraverso accordi di licenza con istituzioni culturali, di fornire accesso a contenuti o "gallerie virtuali" senza incorrere nelle restrizioni territoriali. Verranno analizzati in che misura questi modelli di licenza permettano l'accesso transfrontaliero ai contenuti culturali. Tuttavia, la possibilità di condividere gratuitamente i contenuti online, in concomitanza della chiusura forzata dovuta dall'epidemia di biblioteche, musei e archivi, comporta che gli autori delle opere protette da copyright ricevano una remunerazione minima o nulla per la condivisione delle loro opere e che le istituzioni culturali stiano subendo grandi perdite economiche. Durante la crisi si è ampliata la partecipazione del pubblico, che accede e fruisce dei contenuti culturali digitali e li riutilizza a fini creativi. Poiché gli usi consentiti per i contenuti generati dagli utenti non sono armonizzati a livello di normativa europea, alcuni di questi usi potrebbero violare il copyright di opere preesistenti, a seconda della giurisdizione. In questo quadro il progetto inDICES, analizzerà l'impatto della normativa vigente sul riutilizzo dei contenuti culturali digitali, in cui fanno parte le collezioni di musei, biblioteche e archivi.

inDICES is a research and innovation project funded by the European Commission under the Horizon 2020 program. inDICES is developing of a scientific methodology to measure and assess the economic and social impact of digitisation of cultural heritage on the access to European cultural goods and services and their modes of production. The project will produce recommendations to policy-makers and decision-makers in the cultural heritage and CCI sector for the development of innovative strategies and business models to seize the new opportunities that are offered thanks to the use of digital technologies in cultural goods and services. Together with the recommendations, the project will develop a participation platform and a self-assessment system, useful for cultural institutions to improve their strategic positioning in the digital single market and therefore fostering their digital transformation.

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2020

RIFERIMENTI BIBLIOGRAFICI

- Braşoveanu et al.,
2016 Adrian M.P. Braşoveanu, Marta Sabou, Arno Scharl, Alexander Hubmann-Haidvogel, Daniel Fischl. *Visualizing Statistical Linked Knowledge for Decision Support*. «Semantic Web Journal», 8 (2016) n. 1, p. 113-137.
<<http://www.semantic-web-journal.net/system/files/swj1168.pdf>>.
- Sacco et al.,
2017 Pier Luigi Sacco, Guido Ferilli, Giorgio Tavano Blessi. *From Culture 1.0 to Culture 3.0: Three Socio-Technical Regimes of Social and Economic Value Creation through Culture, and Their Impact on European Cohesion Policies*. «Sustainability» 10-11 (2018), n. 3923.
<<https://www.mdpi.com/2071-1050/10/11/3923>>.
- Scharl et al.,
2017 Arno Scharl, David Herring, Walter Rafelsberger, Alexander Hubmann-Haidvogel, Ruslan Kamolov, Daniel Fischl, Michael Föls, Albert Weichselbraun. *Semantic Systems and Visual Tools to Support Environmental Communication*. «IEEE Systems Journal», 11 (2017), n. 2, p. 762-771. <<https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7244185>>.
- Weichselbraun et al.,
2017 Albert Weichselbraun, Stefan Gindl, Fabian Fischer, Svitlana Vakulenko, Arno Scharl. *Aspect-Based Extraction and Analysis of Affective Knowledge from Social Media Streams*. «IEEE Intelligent Systems», 32 (2018), n. 3, p. 80-88.

New Approaches to OCR for Early Printed Books

«DigItalia» 2-2020
DOI: 10.36181/digitalia-00015

Nikolaus Weichselbaumer - University of Mainz

Mathias Seuret - University of Erlangen-Nuremberg

Saskia Limbach - University of Milan

Rui Dong - Northeastern University

Manuel Burghardt - Leipzig University

Vincent Christlein - University of Erlangen-Nuremberg

Books printed before 1800 present major problems for OCR. One of the main obstacles is the lack of diversity of historical fonts in training data. The OCR-D project, consisting of book historians and computer scientists, aims to address this deficiency by focussing on three major issues. Our first target was to create a tool that identifies font groups automatically in images of historical documents. We concentrated on Gothic font groups that were commonly used in German texts printed in the 15th and 16th century: the well-known Fraktur and the lesser known Bastarda, Rotunda, Textura und Schwabacher. The tool was trained with 35,000 images and reaches an accuracy level of 98%. It can not only differentiate between the above-mentioned font groups but also Hebrew, Greek, Antiqua and Italic. It can also identify woodcut images and irrelevant data (book covers, empty pages, etc.). In a second step, we created an online training infrastructure (okralact), which allows for the use of various open source OCR engines such as Tesseract, OCRopus, Kraken and Calamari. At the same time, it facilitates training for specific models of font groups. The high accuracy of the recognition tool paves the way for the unprecedented opportunity to differentiate between the fonts used by individual printers. With more training data and further adjustments, the tool could help to fill a major gap in historical research.

OCR-D

In past decades, many libraries in Germany have started to digitise their holdings. The numerous digital copies of early printed books have been linked to the matching records in the respective national bibliographies, VD16, VD17 and VD18¹. Taken together, there are now hundreds of thousands of links to digital copies. This is already a significant achievement; yet it would mean a huge step forward for scholarship if these items were available for full-text searches and further processing. Therefore the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) organised a workshop in 2014 in

¹ <<http://www.vd16.de/>>; <<http://www.vd17.de/>>; <<http://www.vd18.de/>>.

which experts assessed how new developments in Optical Character Recognition (OCR) can be used to achieve this goal². OCR is a complex process that comprises more than the character recognition itself. It involves pre-processing (e.g. image de-noising, binarisation, etc.); layout analysis (recognising elements such as headings and illustrations); and post-processing (correcting errors).

The OCR-D project aims to create a conceptual and technical framework that allows for full text transformation of any digital copy in VD16, VD17 and VD18³. For this purpose, the individual steps of automatic text recognition are broken down, allowing users to trace them in the open-source software of OCR-D⁴. Users can then adapt the workflow to their specific needs, e.g. when the layout of particular early printed books need specific settings.

The OCR-D project was coordinated by four partners: the Berlin-Brandenburg Academy of Sciences and Humanities (BBAW), the Herzog-August Library Wolfenbüttel (HAB), the Berlin State Library (SBB) and the Karlsruhe Institute of Technology (KIT)⁵. In the first phase, the project identified major challenges for early printed books and funded eight module projects to address these issues.

Recognising Font Groups

Our module project addresses specific aspects of print in early modern Europe⁶. One of the biggest obstacles for OCR use in early printed books is the fact that OCR engines are usually trained with modern-day fonts. This ignores the large regional and stylistic variety in fonts used in printed texts before 1800 and leads to high error rates in text recognition. Our premise was the following: better OCR results can be created by being able to recognise automatically the font group used in an early printed book which would allow users to choose the best OCR-model for each text; in a second step users would also be able to train their own font-specific model for particularly difficult fonts (such as one of the fonts used by William Caxton, the first printer in England, see below).

² <https://www.dfg.de/en/index.jsp>.

³ Clemens Neudecker - Konstantin Baierer - Maria Federbusch - Matthias Boenig - Kay-Michael Würzner - Volker Hartmann - Elisa Herrmann, *OCR-D: An end-to-end open source OCR framework for historical printed documents*, in: *Proceedings of the 3rd International Conference on Digital Access to Textual Cultural Heritage*, 2019, p. 53–58, <<https://ocr-d.de/en/>>.

⁴ Since early 2020, a prototype of the OCR-D software is used in nine libraries, see: Konstantin Baierer - Matthias Boenig - Elisabeth Engl - Volker Hartmann - Clemens Neudecker - Reinhard Altenhöner - Alexander Geyken - Johannes Mangei - Rainer Stötzka, *OCR-D kompakt: Ergebnisse und Stand der Forschung in der Förderinitiative*, «BIBLIOTHEK-Forschung und Praxis» 44 (2020), n. 2, p. 1–13.

⁵ <<https://www.bbaw.de/en/>>; <<http://www.hab.de/en/home.html>>; <<https://staatsbibliothek-berlin.de/en/>>; <<https://www.kit.edu/english/index.php>>.

⁶ <https://www.buchwissenschaft.uni-mainz.de/forschung/modellrepositorium-ocr-d/#englisch>.

To this end, the project had three major objectives:

- train a neural network to recognise font groups commonly used in early modern books;
- build infrastructure to allow for easy training of all important open source OCR-engines;
- making our software publicly available for free.

Training a neural network requires a large amount of data. In our case, we needed images of book pages for which we knew which font group they showed. Thankfully, the TW (Typenrepertorium der Wiegendrucke) offered a good point of departure⁷. The TW has been developed over more than a century, listing almost every font ever used in incunabula (books printed between 1450 and 1500). This comprehensiveness made it ideal for our project. However, one has to keep in mind that the data in the TW is not complete. Most books were printed with two (or more) fonts, yet often the TW has only identified one of them. Moreover, it is not indicated in the records where specifically the font is used in the book, e.g. only on the title page or as a running header.

Such pages with multiple font groups created much noise for the network, which prevented it from learning accurately. Thus, it was necessary for a human expert to go through the images first and indicate manually which font was used on which page. For this, we made use of a very comprehensive list of fonts that Dr. Oliver Duntze from the TW project kindly shared with us. The list differentiates between the many Gothic fonts and Antiqua fonts and groups them together. We could then specifically look for certain font groups and selected images for them. On top of that, we accumulated as much data as possible from our cooperating libraries⁸.

We divided about 6,000 fonts of the TW into 6 groups: Textura, Rotunda, Bastarda, Schwabacher, Gotico-Antiqua and Antiqua. We then accumulated material for two additional font groups often found in early printed books: Greek and Hebrew. Finally we also included fonts which came into use early in the 16th century: Fraktur and Italic.

⁷ <https://tw.staatsbibliothek-berlin.de/>.

⁸ We were graciously supported by a number of libraries which provided us with data: the State Library in Berlin, the University Library of Cologne, the University Library of Erlangen, the State and University Library Göttingen, the University Library of Heidelberg, the State Library Munich, the State Library Stuttgart, the Herzog August Bibliothek Wolfenbüttel. We estimate that we were given access to more than 250 million digitised pages. We were especially thankful for the support of the British Library (London) which allowed us to work with their digitised incunabula collection. The British Library houses one of the largest and most diverse collections of books printed in the 15th century and the institution recently scanned their entire collection with outstanding image quality. Although the images are not online yet, the library sent us the data on a hard drive and allowed us to use the images for our research purposes.

After a first test, which showed promising but improvable results, we attempted to better understand the usefulness of these groups from a computer's point of view⁹. For this, we first trained a neural network on several thousand pages. Then, we randomly selected 18,000 unlabelled page images from the holdings of the Berlin State Library and extracted feature vectors for each of them. To do so, we applied the network on the whole surface of each page, and stored the 384-dimensional output of the penultimate layer of the network – a typical feature extraction approach. As such high-dimensional data cannot be directly displayed, nor be understood by a human, we projected it into a 2-D space using the t-distributed stochastic neighbor embedding (t-SNE)¹⁰. Each of the 18,000 images is displayed as a blue dot (see Fig. 1). By clicking on any of the dots (which then turns red), the corresponding page is displayed on the right side.

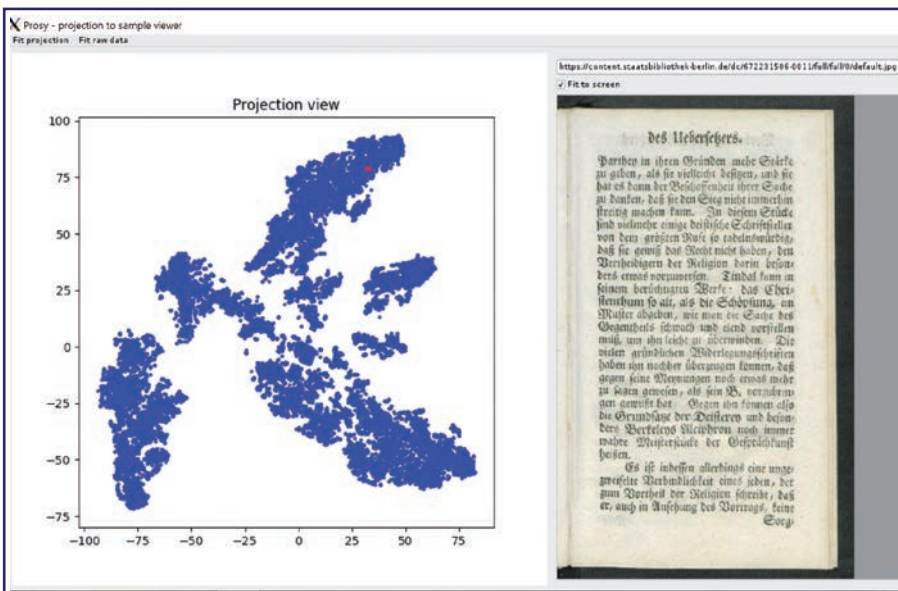


Figure 1. Clustering of 18k images from books in the Berlin State Library. The shown book page is located in the top part of the upper cluster (red dot). Screenshot of DICI, the tool used to check the test results. <<https://github.com/seuerm/dici>>

Once we had all individual pages displayed as blue dots, a human expert carefully checked thousands of pages for common features. It turned out that the clusters indeed matched some font groups, such as Antiqua, Rotunda, Textura, Bastarda and Fraktur (as seen in Fig. 2).

⁹ Nikolaus Weichselbaumer - Mathias Seuret - Saskia Limbach - Vincent Christlein - Andreas Maier, *Automatic Font Group Recognition in Early Printed Books*, in: *Digital Humanities im deutschsprachigen Raum (DHd) 2019. 6th International Conference 25-29 March 2019, Universitäten zu Mainz und Frankfurt*, p. 84-87, <<https://doi.org/10.5281/zenodo.2596095>>.

¹⁰ Laurens van der Maaten and Geoffrey Hinton, *Visualizing data using t-SNE*, *Journal of Machine Learning Research* 9 (2008), p. 2579-2605.

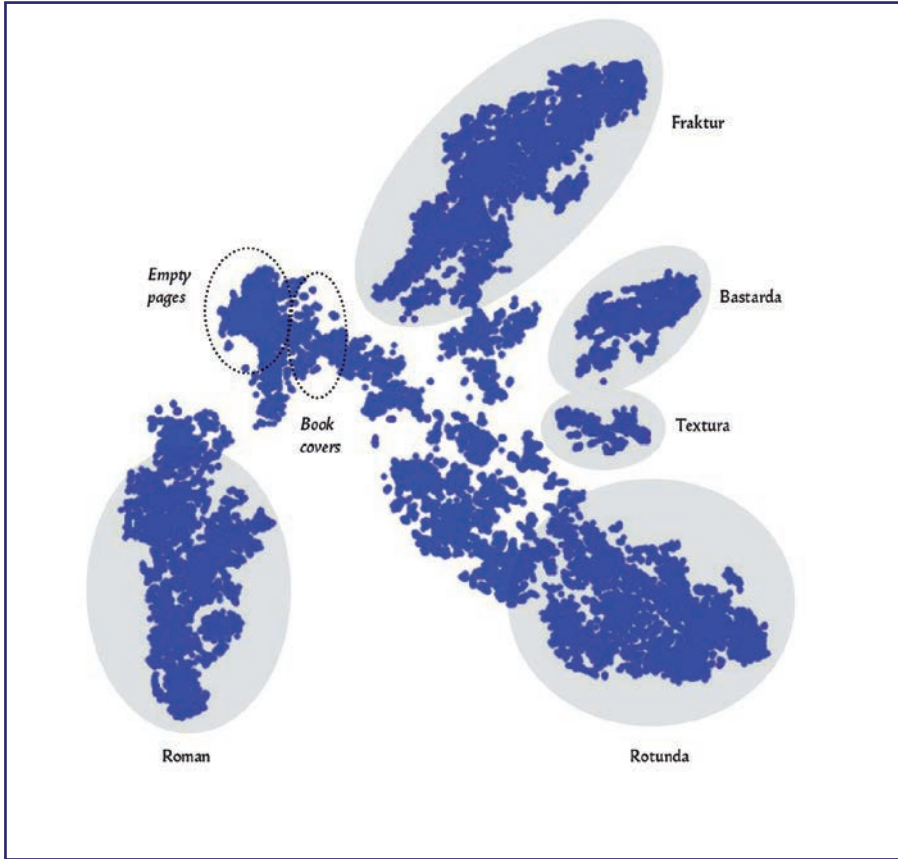


Figure 2. Results in t-sne, showing that certain font groups were clustered together

Other font groups were more difficult to find. Italic pages, which are not marked in Fig. 2, mostly appeared on top of the Antiqua cluster, along with many pages containing a mixture of Antiqua and Italic. Mixing these two font groups was quite common in the early modern period. Hence, there is no clear Italic cluster.

The results also showed that it was important to create more groups, which allowed us to teach the network to sort out “distractions”, such as empty pages or book covers. The network grouped them together already. However, if the network is not taught specifically that these images are different, i.e. do not contain a font group, it will attribute them to one of the selected font groups. Therefore, we created two additional groups. On the one hand, the group “other font” includes handwritten notes and non-Latin alphabets beyond Greek and Hebrew (such as Arabic and Farsi). On the other hand, we created the group “not a font” which includes blank pages, book covers, woodcuts, etchings, paintings, printer marks, ornaments and remnants of the digitisation process, such as colour charts, scanner beds and accidentally reproduced hands.



Figure 3. Samples of the font groups used in the project

Our next step was to prepare a large amount of data for training a deep neural network to identify font groups, since no such dataset existed before (see Fig. 3). We built it in an iterative way: we used information from the TW to identify incunabula containing the desired font groups and collected roughly 100 pages for each group. We used these pages for training a relatively small network, inspired by existing script identification methods¹¹. This network helped us to identify useful pages without any distractions out of the data provided by our partner libraries. This approach was significantly more efficient than a manual search.

We labelled the selected pages using two separate approaches: first, we used labelbox and uploaded a selection of images from our partner libraries¹². Then, we asked a number of participants to assign the correct label(s) and specify the main font if there was more than one. Second, we used the rest of the material our partners supplied to select pages manually with only a single font group. For this we prepared a batch of images from multiple libraries in one folder and extracted the “not a font” images in thumbnail view. We examined the remaining images and deleted every page with more than one font group. Finally, we transferred each image in the right single class folder. This second method proved to be much more efficient, since we could reduce loading times considerably and got significantly more data for underrepresented font groups, such as Textura and Rotunda. As a result our dataset consists mostly of pages with only one font group per page. We attempted to include pages from all editions of Gotico-Antiqua that were digitally available under free licenses in order to compensate for the form variety within this font group. This dataset is now freely available

¹¹ Florence Cloppet - Véronique Églin - Van Cuong Kieu - Dominique Stutzmann - Nicole Vincent, *ICFHR2016 Competition on the Classification of Medieval Handwritings in Latin Script*, in: *ICFHR, 2016*, p. 590–595; Florence Cloppet - Véronique Églin - Marlène Helias-Baron - Van Cuong Kieu - Nicole Vincent - Dominique Stutzmann, *ICDAR2017 Competition on the Classification of Medieval Handwritings in Latin Script*, in: *ICDAR, 2017*, p. 1371–1376.

¹² <https://labelbox.com>.

online¹³. While our dataset was growing, we tried various standard and specialised neural network architectures. We started with residual networks (ResNet) with 18 and 50 layers, greatly decreasing the amount of neurons in the various layers to prevent over-fitting¹⁴. We also tried VGG-16, and combinations of variational autoencoders and classifiers¹⁵. We found out that for small amounts of data, the design of architectures specific for this task tends to greatly improve the results. However, with the amount of data we labelled, differences between architectures almost vanished, and we settled on a DenseNet-121¹⁶.

A challenge, which is not entirely solved yet, is the case of pages with multiple font groups. Our labelled data contains many such pages, but processing them correctly with the network is not straightforward. The network attributes a “score” to each of its possible outputs. This score is an unbounded real number, and we decided that the highest value corresponds to the answer of the network. If several font groups are present on a page, then only one will be detected – typically the one covering the largest surface, as this leads to a higher score. Using a global score threshold from which a font group is detected is not directly feasible, as there is no simple way to distinguish between a misdetection (e.g. a medium score for Antiqua since there are also lines in Italic), and a correct detection of a small surface (e.g. a single word in Antiqua in the middle of a paragraph in Fraktur).

This issue could be tackled in two ways. First, if the training data contained location information for the different font groups, then the network could learn to provide reliable results even for small amounts of text, as localisation information could be used during the training phase. This would, however, require a vast amount of manual labour to prepare. In addition to simply indicating which font groups are present on a page, the experts preparing the data would have to manually draw rectangles or polygons over many words or parts of words in secondary fonts. This time-consuming procedure would have to be performed on thousands of pages.

¹³ Mathias Seuret - Saskia Limbach - Nikolaus Weichselbaumer - Andreas Maier - Vincent Christlein, *Dataset of Pages from Early Printed Books with Multiple Font Groups*, in: *HIP '19. Proceedings of the 5th International Workshop on Historical Document Imaging and Processing*, p. 1–6, <<https://doi.org/10.1145/3352631.3352640>>.

¹⁴ Kaiming He - Xiangyu Zhang - Shaoqing Ren - Jian Sun, *Deep residual learning for image recognition*. 2016, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, p. 770–778, <<https://doi.org/10.1109/CVPR.2016.90>>.

¹⁵ Karen Simonyan - Andrew Zisserman, *Very deep convolutional networks for large-scale image recognition*. 2014. *arXiv preprint* <[arXiv:1409.1556](https://arxiv.org/abs/1409.1556)>; Diederik P. Kingma - Max Welling, *Auto-encoding variational Bayes*. 2013. *Proceedings of the 2nd International Conference on Learning Representations*. *arXiv preprint* <[arXiv:1312.6114](https://arxiv.org/abs/1312.6114)>.

¹⁶ Gao Huang - Zhuang Liu - Laurens Van Der Maaten - Kilian Q. Weinberger, *Densely connected convolutional networks*, in: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, 2017, p. 2261–2269, <<https://doi.org/10.1109/CVPR.2017.243>>.

The second option we investigated is to command the network to indicate which areas of its input could lead to a specific font group output, and how much¹⁷. This is illustrated in the two images below, which we took from an incunable currently in the collection of the British Library (see Fig. 4). The intensity of the heatmap (or highlighting) indicates which specific regions of the page get related to a font group by the neural network. Thus, in the case of mixed content, the network is able to provide location information – even if it was never trained for this purpose.

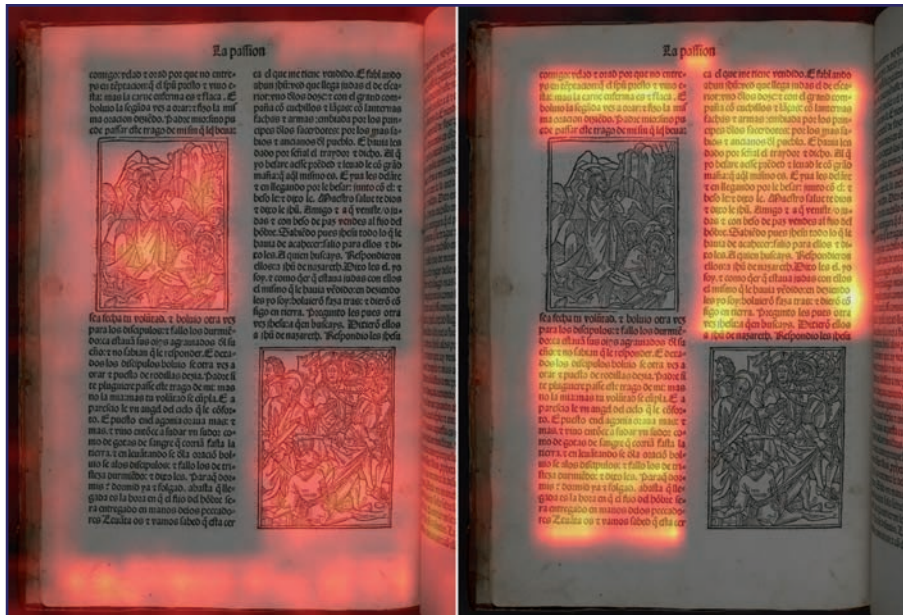


Figure 4. Heatmaps showing confidences for Not a font (left) and Rotunda (right). Sample page: *Jacobus de Voragine, Legenda aurea sanctorum (...)*, [Burgos: Juan de Burgos, about 1499], London, British Library: IB.53312. (GW M11511), p. 10

However, generating such heatmaps is time-consuming and therefore not well suited for processing large amounts of images, such as library collections. On top of that, storing the results for later use would require much disk space: when we store only one image plus color codes for labeling the pixels, the results may be difficult to evaluate. It would be better to store each heatmap as a single image. Yet, this takes up a lot of space. Thus, while these results are surprisingly good, we decided not to pursue this approach for our project. After all, our objective was to deal with large amounts of data, and using a slow, computationally expensive and

¹⁷ Using Grad-CAM, a method estimating where the source of a decision taken by a neural network is, see Ramprasaath R. Selvaraju - Michael Cogswell - Abhishek Das - Ramakrishna Vedantam - Devi Parikh - Dhruv Batra, *Grad-cam: Visual explanations from deep networks via gradient-based localization*, in: *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, 2017, p. 618-626, <<https://doi.org/10.1109/ICCV.2017.74>>.

power-consuming method is simply not acceptable, regardless of the accuracy achieved. In addition to these considerations, as the font recognition step can potentially be applied after the layout analysis, the labelling can be done with a finer granularity than page-level.

Training and Using OCR models

In the second part of our project, we created an online training infrastructure (okralact), which makes it much easier to train OCR models for different engines¹⁸. We focussed on the four commonly used OCR engines Tesseract, Ocropus, Kraken and Calamari. Figure 5 shows the overall structure of our online training system. Through a web-based front-end, the user can interact with the server-side component. It allows users to upload datasets, OCR models, or a configuration file containing the parameter setting to train, fine tune or evaluate an OCR model.

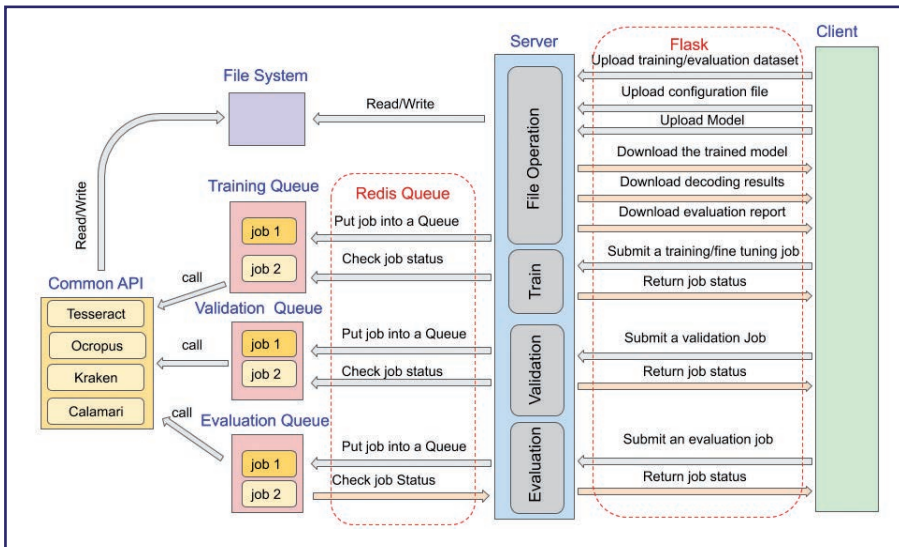


Figure 5. Framework of the okralact system

The server-side mainly contains a common API for the above mentioned OCR engines with unified training, decoding and evaluating interfaces (see Fig. 6). For example, a user can now upload one dataset and it is automatically split into a training and a validation set. We also added an evaluation component which compares results with the ground truth files to analyse the performance of the model.

¹⁸ Konstantin Baierer - Rui Dong - Clemens Neudecker, *Okralact - a multi-engine Open Source OCR training system*, in: *Historical Document Imaging and Processing (HIP) 2019, 5th International Workshop 21-22 September 2019, Sydney, Australia*, p. 25–30, <<https://doi.org/10.1145/3352631.3352638>>.

A very important feature of our system is the unification of the interfaces of the different OCR engines. This allows users to try out the engines without having to learn all of their individual specifics. This harmonisation was a long and complicated task, for several reasons. First, the documentation often does not provide accurate details about parameters, and thus we had to frequently dive into the source code of the engines to see exactly how the parameters are used (and what they correspond to). Second, in many cases the engines have parameters that share similar functions, but have different names and have to be used in a slightly different way (e.g. number of iterations and number of training samples), or in extremely different ways (e.g. defining the network architecture). And, third, some parameters are available only for some of the engines, so our system has to take this into account in order to allow full access to the engines' capabilities.

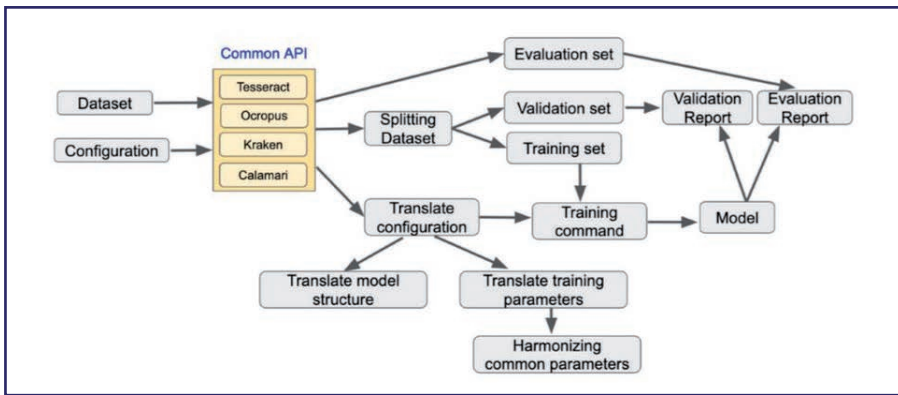


Figure 6. Workflow of the common API component

We also created the possibility for users to store and share their models in an online repository. The metadata stored with the models records the training set and parameters. It is also possible to evaluate the model with the help of a test set if the users would like to know how the model performs before using it.

Finally, we transcribed lines for three comparatively difficult font groups for which current OCR models have very high error rates – Schwabacher, Textura and Bastarda. Therefore, we adapted and used DIVAnnotation, a segmentation, labelling, and transcription tool. It is a semi-automatised modular annotation tool originally developed in 2018 by the DIVA group, from the University of Fribourg, Switzerland¹⁹. Our modifications of this tool mainly consisted in making its inter-

¹⁹ Mathias Seuret - Manuel Bouillon - Fotini Simistira - Marcel Würsch - Marcus Liwicki - Rolf Ingold, *A Semi-automatized Modular Annotation Tool for Ancient Manuscript Annotation*, in: *13th IAPR International Workshop on Document Analysis Systems (DAS)*, Vienna, 2018, p. 340-344, <<https://doi.org/10.1109/DAS.2018.80>>.

face significantly more intuitive, tune it to our specific needs, and fix several bugs. The adapted version of DIVAnnotation is very easy to use and allows the user to generate transcriptions of historical text very quickly (Fig. 7)²⁰.

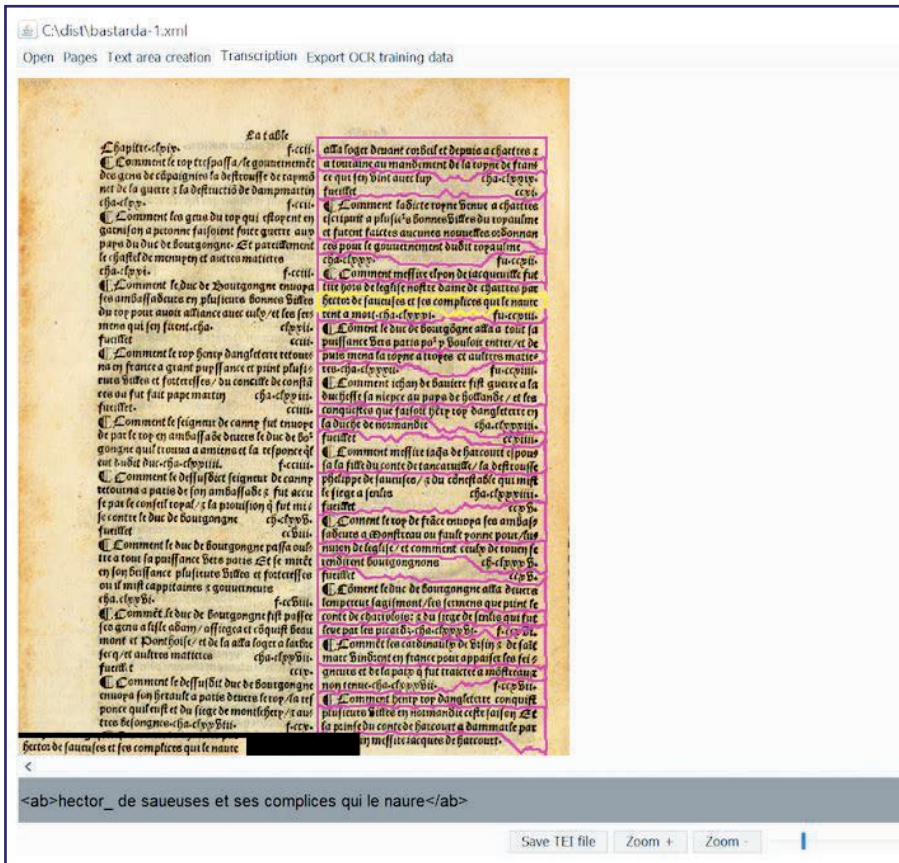


Figure 7. DIVAnnotation lite for transcribing historical documents

We also designed a test for a rather difficult rare book in the British Library²¹. For his *Dits des philosophes. The dictes or sayengis of the philosophres* (1477), the first English printer, William Caxton, used an exceptional Bastarda²². When using Tesseract, we received the following results: for the English model the error rate was 41% (41 out of 100 characters were wrong). The Fraktur model worked slightly better but still produced an error rate of 28%. A combination of the two models produced an error rate of 31%.

²⁰ <https://github.com/seuretm/divannotatio-lite>.

²¹ We were kindly invited by the British Library to present our results on 18 June 2019.

²² ISTC id00272000, <<https://data.cerl.org/istc/id00272000>>.

Next, we used DIVAnnotation to transcribe a few hundred lines and used only 92 lines to train a model for Kraken (the rest was kept for validation). We chose this low number of lines to demonstrate that users can already have good results without putting in too much effort. After all, users presumably do not want to transcribe thousands of lines before they can use OCR on their specific book. When we used our trained model on new text lines of the book, it produced an error rate of only 8%.

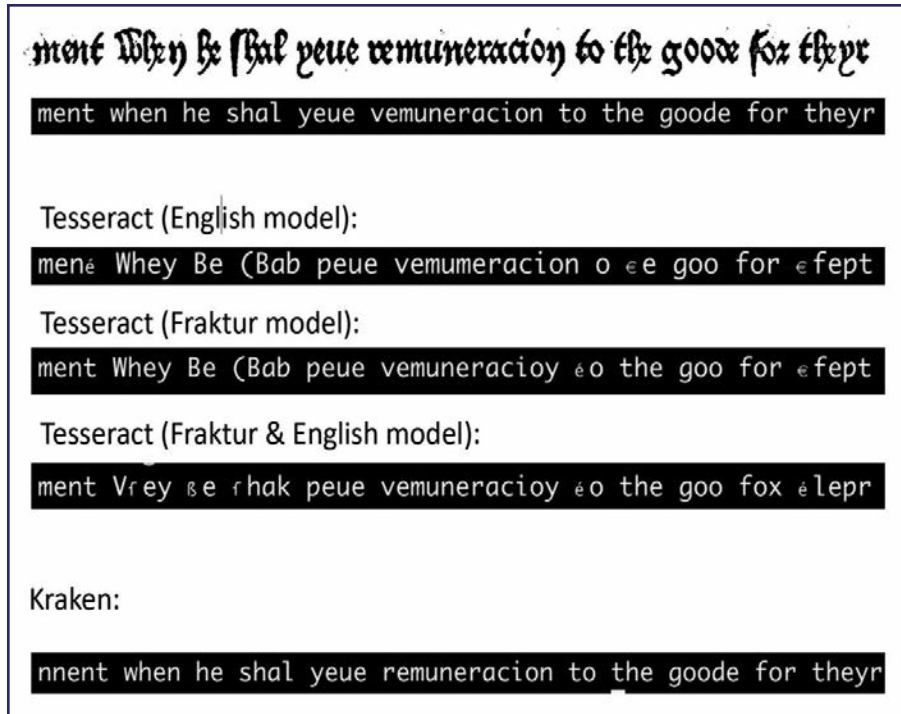


Figure 8. OCR results for a difficult Bastarda text, using Tesseract with various models and a pre-trained Kraken model. The textline and its transcription are at the top, the results of the OCR engines with their various models below

In the past two years, we have published the programs that we developed as open source and provided detailed notes on the code²³. Thus, our work can easily be used by other researchers and adapted to specific needs. The font group recognition system has already attracted great interest and has been adapted to some needs of the Transkribus team²⁴. A trained network is also provided with the recognition system, thus it can be used either as a part of the OCR-D framework, or as a stand-alone application.

²³ <https://github.com/seuretm/ocrd_typegroups_classifier/>; <<https://github.com/Doreenruiui/okralact>>.

²⁴ <https://read.transkribus.eu/2019/09/26/printed-vs-handwritten-text-lines-automatically-separated/>.

Furthermore, the dataset of our *Dataset of Pages from Early Printed Books* paper is freely available online²⁵. It consists of more than 35,000 pages of printed books dating from the 15th to the 18th century, which makes it one of the largest datasets in the field. It is also the very first dataset for font group identification in early printed books. Before, datasets almost exclusively consisted of two major font groups, i.e. Antiqua and Fraktur, with many other Gothic fonts being labelled wrongly as Fraktur as well as no differentiation between Antiqua and Italic. This created major issues for OCR. Hence, we accumulated a much more varied dataset, which represents the diversity of early modern fonts more accurately. Since its publication in August 2019, the dataset has attracted over 400 views and 150 downloads. The pages are randomly selected, unsorted, and therefore cannot be read consecutively to discourage any other use of the pages.

Conclusions

While the large-scale application of font group specific OCR is still hampered by a lack of training data, our project opened the door to this approach, that has remarkable potential. Yet font group recognition also opens new avenues for historical scholarship. We recently started to explore these possibilities by examining the emergence of the font group “Fraktur” in the 16th century and the development that lead to it becoming the most popular font choice for German texts for centuries. To address this issue, we used our tool on some 85,000 German books, printed between 1472-1800, which have been digitised by the Bavarian State Library, the German library with the by far largest collection.

The results largely confirmed previous (but rather anecdotal) observations, showing that Schwabacher, previously the most common font for German language printing, was gradually replaced by Fraktur over the run of the 16th century. In the course of the 17th and 18th century, Fraktur was used in around 9 out of 10 German language books. Contrary to previous assumptions, the 1790s show a slight uptick in the use of Antiqua for German language printing, demonstrating that the Antiqua-Fraktur-dispute had more of an impact than previously assumed²⁶. This small study showed how machine learning can help answer questions in the domain of book history that could not have been addressed with conventional methods. We are very optimistic that this tool has many further applications.

²⁵ <https://doi.org/10.5281/zenodo.3366685>.

²⁶ Nikolaus Weichselbaumer - Mathias Seuret - Saskia Limbach - Lena Hinrichsen - Andreas Maier - Vincent Christlein, *The rapid rise of Fraktur*, in: *Digital Humanities im deutschsprachigen Raum (DHd) 2020. 7th International Conference 02-06 March 2020, Universität Paderborn*, p. 229-231, <<https://doi.org/10.5281/zenodo.3666690>>.

I libri stampati prima del 1800 presentano molte difficoltà per l'OCR. Uno degli ostacoli principali è rappresentato dalla mancanza di diversità dei caratteri storici usati per lo sviluppo del sistema. Il progetto OCR-D, che ha visto la collaborazione di storici del libro e informatici, ha avuto il fine di affrontare questa carenza concentrandosi su tre problemi fondamentali. Il primo obiettivo è stato creare uno strumento che riconoscesse automaticamente le famiglie di caratteri nelle immagini di testi a stampa antichi. Ci siamo concentrati sulle famiglie di caratteri gotici comunemente usate nei testi stampati in Germania nel XV e XVI secolo: il più conosciuto Fraktur e le meno note Bastarda, Rotunda, Textura e Schwabacher. Lo strumento è stato sviluppato su 35.000 immagini, raggiunge un livello di precisione del 98% e riesce a distinguere non solo tra le famiglie di caratteri già menzionate ma anche tra ebraico, greco, antiqua e corsivo. È inoltre in grado di identificare immagini xilografiche e dati irrilevanti (coperte, pagine bianche ecc.). In una seconda fase abbiamo creato un'infrastruttura online (okralact) che facilita l'uso di vari motori OCR open source come Tesseract, OCRopus, Kraken e Calamari e che, allo stesso tempo, facilita l'apprendimento di modelli specifici per famiglie di caratteri. L'elevata precisione di questo software per il riconoscimento apre la strada all'opportunità senza precedenti di distinguere i caratteri utilizzati da ogni stampatore. Con una maggiore quantità di dati per il raffinamento e aggiustamenti successivi, questo strumento può rivelarsi utile nel colmare una lacuna considerevole nella ricerca storica.

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2020

ARIADNEplus: l'avventura continua

«DigItalia» 2-2020
DOI: 10.36181/digitalia-00016

Franco Niccolucci

*PIN srl – Servizi didattici e scientifici per l'Università di Firenze,
coordinatore scientifico del progetto*

L'articolo descrive il progetto europeo ARIADNEplus, continuazione del precedente progetto ARIADNE. Questo nuovo progetto ha, come il precedente, l'obiettivo di superare la frammentazione degli archivi digitali di dati archeologici europei disponibili online attraverso la creazione di un catalogo. Rispetto al precedente progetto, ARIADNEplus ne estende significativamente la copertura geografica, tematica e temporale. I dati sono forniti da autorevoli istituzioni nazionali che ne garantiscono la qualità. Il catalogo permette di effettuare ricerche sui metadati, che sono conformi a standard internazionali, con vari criteri, e quindi di accedere ai dati direttamente presso l'archivio del proprietario dei dati stessi. Questi possono poi essere ulteriormente rielaborati con vari strumenti software forniti nel sistema. Tutti gli strumenti informatici utilizzati si avvalgono delle tecnologie più recenti. Una parte importante delle risorse del progetto è dedicata alla formazione degli utenti sia all'uso del sistema ARIADNEplus sia, in generale, all'applicazione in campo archeologico delle politiche e delle strategie europee riguardo alla open science.

Nel 2013 la Commissione Europea finanziò il progetto¹ quadriennale ARIADNE (Advanced Research Infrastructure for Archaeological Dataset Networking in Europe) per la creazione di un'infrastruttura digitale per l'archeologia², con l'obiettivo di aggregare gli archivi digitali europei di dati archeologici in un unico sistema. Con il termine "aggregazione" si intende qui la creazione di un sistema che raccoglie e ordina i metadati relativi a tali archivi, consentendone l'accesso online, mentre gli archivi stessi restano collocati e gestiti dal proprietario dei

¹ Il sito web del progetto è <<https://ariadne-infrastructure.eu/>>. Tale sito (in inglese) riguarda attualmente il progetto ARIADNEplus, descritto di seguito, continuazione di ARIADNE. È inoltre disponibile il sito <<https://whatis.ariadne-infrastructure.eu/>> che illustra in termini non tecnici e per un pubblico di non addetti ai lavori il significato e l'importanza del progetto. Questo sito è disponibile in tutte le lingue europee compreso l'italiano.

² Franco Niccolucci, *Un'infrastruttura di ricerca per l'archeologia: il progetto ARIADNE*, «DigItalia. Rivista del digitale nei beni culturali», 8 (2013), n. 2, p. 154-161, <<http://digitalia.sbn.it/article/view/833/558>>. Ulteriori e più dettagliate informazioni si trovano anche in: Franco Niccolucci – Julian D. Richards, *ARIADNE: Advanced Research Infrastructures for Archaeological Dataset Networking in Europe*, «International Journal of Humanities and Arts Computing», 7 (2013), n. 1-2, p. 70-88, e in: Franco

dati. Il catalogo degli archivi risultante dopo l'aggregazione permette di effettuare delle ricerche sui metadati secondo vari criteri e fornisce una lista dei risultati con indicazioni sommarie dedotte dai metadati. A questo scopo, ARIADNE sviluppò un modello dati, chiamato ACDM (ARIADNE Catalogue Data Model), idoneo a registrare i metadati di una tipologia di archivi estremamente diversificata: report, immagini, disegni, database, risultati di analisi scientifiche e altro ancora. Per consentire l'interoperabilità di questi dati, ognuna delle istituzioni proprietarie degli archivi catalogati in ARIADNE sviluppò un proprio *mapping*, cioè una descrizione della corrispondenza fra i metadati originari usati per ciascun file e quelli richiesti dall'ACDM. In questo modo è stato possibile catalogare circa 1.700.000 archivi di dati archeologici provenienti da tutta Europa e forniti dai 16 partner di ARIADNE di tipo archeologico – gli altri hanno contribuito a predisporre la tecnologia necessaria – provenienti da altrettanti paesi europei. Si poteva, e si può tuttora, accedere al catalogo attraverso un portale dedicato e utilizzare strumenti di ricerca per selezionare gli elementi desiderati, impostando criteri relativi a parole chiave contenute nei metadati degli archivi, a tempo, luogo, tipologia di file e proprietario degli archivi. La stragrande maggioranza di questi archivi era costituita dalla cosiddetta *grey literature*, cioè rapporti non pubblicati di indagini archeologiche, spesso di archeologia preventiva; relativamente meno i database o gli archivi inerenti ad aspetti più specialistici come le indagini scientifiche, i GIS archeologici e così via.

Il catalogo è continuato a crescere anche dopo il periodo finanziato dalla Commissione Europea, che si è concluso il 31 gennaio 2017. Alla fine del progetto l'attività non è stata infatti interrotta, con l'impegno volontario di alcuni partner e la collaborazione di tutti, come dimostrano i 300.000 record di catalogo aggiunti nel frattempo. Alla data odierna il catalogo di ARIADNE contiene infatti circa 1.900.000 elementi, di cui 60.000 (pari al 3%) database o collezioni di immagini o disegni, e circa 4.800 (lo 0,25%) archivi di dati scientifici, mentre il resto sono appunto rapporti testuali.

ARIADNE venne valutato molto positivamente sia dall'autorità finanziatrice che dalla comunità scientifica, come testimoniano alcune citazioni da parte di ricercatori e gestori del patrimonio archeologico. Nell'ultimo periodo, quello di piena funzionalità del portale, vi sono stati 11.000 accessi distinti al portale, che

Niccolucci, *Digital archaeological archives of Europe: opportunities and challenges*, in: *Setting the Agenda: Giving New Meaning to the European Archaeological Heritage*, ed. by P.A.C. Schut, D. Scharff and L. C. de Wit, Budapest: Archaeolingua, 2015, p. 20-22. Una descrizione dell'intero progetto e dei suoi risultati è contenuta soprattutto nell'articolo riassuntivo: Carlo Meghini – Roberto Scopigno – Julian D. Richards – Holly Wright – Guntram Geser – Sebastian Cuy – Johan Fihn – Bruno Fanini – Hella Hollander – Franco Niccolucci – Achille Felicetti – Paola Ronzino – Federico Nurra – Christos Papatheodorou – Dimitris Gavrilis – Maria Theodoridou – Martin Doerr – Douglas Tudhope – Ceri Binding – Andreas Vlachidis, *ARIADNE: A Research Infrastructure for Archaeology*. «ACM Journal on Computing and Cultural Heritage (JOCCH)», 10 (2017), n. 3, Special Issue on Digital Infrastructure for Cultural Heritage Part 2, Article No. 18.

confermano l'interesse dei ricercatori. Le descrizioni dell'impatto positivo sulle diverse realtà dei partner sono state raccolte in un volume³ pubblicato nell'estate 2019.

Fra la fine del periodo finanziato per ARIADNE e la possibilità di accedere a un nuovo finanziamento europeo sono trascorsi quasi due anni, in conseguenza delle date di pubblicazione dei nuovi bandi europei, dei tempi necessari per la preparazione e l'invio delle proposte, la loro valutazione e approvazione, e infine per la stipula dei contratti di finanziamento (il cosiddetto *Grant Agreement*). Le attività sono però continuate, come già ricordato, anche in questo difficile periodo, descritto da alcuni dei partner come "l'attraversamento della *Death Valley*", il noto deserto americano famoso per la sua aridità e desolazione tanto da essere appunto chiamato la "Valle della Morte". Certamente, l'apertura del bando a fine 2017 ha contribuito a dare la speranza necessaria a continuare. Il nuovo progetto, chiamato ARIADNEplus per sottolinearne la continuità con il precedente, è stato approvato nell'agosto 2018 ed è iniziato il primo gennaio 2019. Non si tratta però di una semplice continuazione: ARIADNEplus infatti estende ARIADNE per quanto riguarda la copertura geografica, l'estensione tematica, la ricchezza dei metadati e la potenza degli strumenti di ricerca nel catalogo.

Per quanto riguarda la copertura geografica, il nuovo progetto copre tutta l'Europa attraverso i suoi partner, che ora sono 41 rispetto ai 24 di ARIADNE.

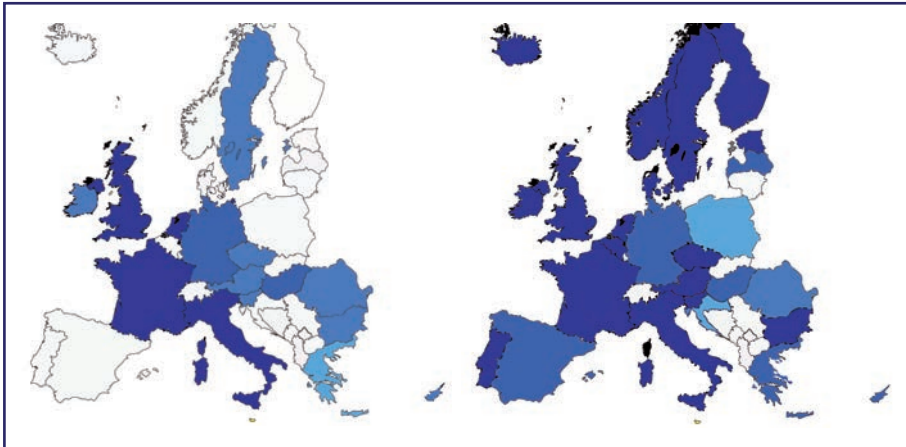


Figura 1. Copertura geografica di ARIADNE (a sinistra) e ARIADNEplus (a destra)

³ *The ARIADNE Impact*, ed. by Julian D. Richards. Franco Niccolucci, Budapest: Archaeolingua, 2019. Il volume è anche disponibile in formato PDF attraverso il sito web <<https://ariadne-infrastructure.eu>> nella sezione Resources.

La figura 1 mostra gli stati europei coperti, e le diverse tonalità di blu indicano la maggiore o minore penetrazione dell'aggregazione eseguita nel progetto rispetto ai dati disponibili in ciascun paese. Allo scopo di assicurare – almeno in linea di principio – la qualità dei dati catalogati, il progetto infatti richiede che questi vengano forniti da enti di livello nazionale, ad esempio nel Regno Unito dall'Archaeological Data Service dell'Università di York (ADS); in Olanda dal Data Archiving and Networked Services della Koninklijke Nederlandse Akademie van Wetenschappen, l'Accademia reale olandese delle scienze (DANS-KNAW); in Francia dall'Institut National de Recherches Archéologiques Préventives (INRAP); in Italia dal Ministero per i Beni e le attività culturali e per il turismo (MiBACT); e così via per gli altri paesi. Difficoltà sorgono per quei paesi come la Spagna o la Germania che decentrano la gestione dei dati archeologici a strutture regionali (Spagna) o alla struttura federale dei *Länder* (Germania). Per l'Italia, i dati sono forniti dall'Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU), che è attualmente l'unico organismo ministeriale con un'efficiente struttura digitale in grado di collaborare al progetto, viste le note criticità e inefficienza del Sistema informativo generale del catalogo (SIGECweb) e gli interventi che si sono succeduti, anche in tempi recenti, sulle attività digitali del MiBACT. Una collaborazione è in corso con l'Istituto centrale per l'archeologia (ICA) per la realizzazione di un geoportale italiano dell'archeologia, da integrare con sistemi simili a livello europeo. Contatti sono stati avviati

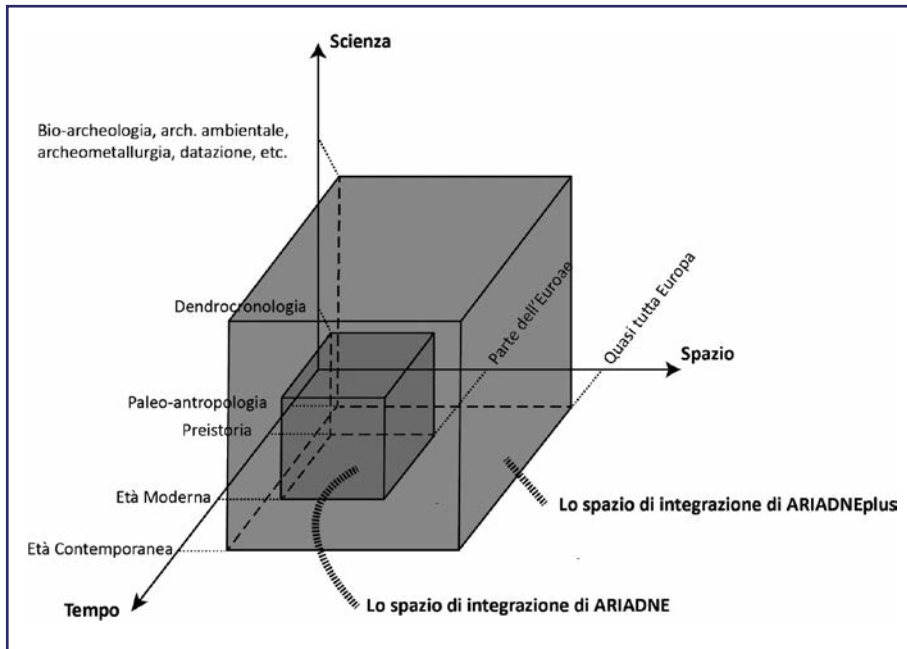


Figura 2. Estensione tematica del contenuto del catalogo di ARIADNEplus

con Istituto centrale per il catalogo e la documentazione (ICCD) per la realizzazione di standard condivisi. Altre informazioni provengono anche dal Sistema informativo territoriale archeologico di Roma (SITAR). Infine, molti dati sulle ricerche in Italia sono contenuti nel database Fasti online, gestito da un altro partner del progetto, l'Associazione Internazionale di Archeologia Classica (AIAC) che copre vari paesi del sud Europa.

Vale infine la pena di ricordare che la copertura di ARIADNEplus è ormai diventata mondiale. Fra i suoi partner infatti figurano la Israel Antiquity Authority (IAA, Israele), il National Institute for Cultural Heritage (NARA, Giappone), la rete archeologica del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, Argentina) e l'archivio americano denominato The Digital Archaeological Record (tDAR) dell'associazione Digital Antiquity, coordinata dall'Arizona State University.

All'estensione geografica si accompagna, come già ricordato, una significativa estensione tematica e temporale (Fig. 2).

La modesta copertura dei dati scientifici del catalogo di ARIADNE è ora in fase di significativo ampliamento. L'Istituto Nazionale di Fisica Nucleare (INFN), partner del progetto, attraverso la sua rete sui beni culturali CHnet ha già realizzato per l'Italia un sistema per la gestione integrata dei dati provenienti da analisi scientifiche, che è in corso di estensione all'Europa, attraverso la collaborazione con vari partner di altri paesi e la creazione di un modello dati standard valido per tutti.

In generale l'obiettivo di ARIADNEplus è quello di migliorare la qualità e la ricchezza dei metadati che utilizza. Per questo scopo è già stato implementato un nuovo schema generale di metadati per il Catalogo, denominato AO-Cat (ARIADNEplus Ontology for the Catalogue), che migliora il precedente ACDM ed è totalmente compatibile con lo standard internazionale CIDOC CRM⁴. È in corso o già programmata la definizione di *application profiles* cioè di estensioni per gestire i dati relativi a sotto-domini specifici quali, ad esempio, i già menzionati dati scientifici, quelli paleoantropologici, i dati relativi alle strutture in elevato, le iscrizioni, l'archeologia subacquea e altri ancora. Tutte queste estensioni saranno totalmente compatibili con lo standard CRM.

Il lavoro di ARIADNEplus si basa infatti sulla standardizzazione, senza la quale l'integrazione dei dati non sarebbe possibile. A questo riguardo, un aspetto importante è l'utilizzo di vocabolari controllati, adottando in linea generale un'estensione di AAT (Art & Architecture Thesaurus) del Getty Research Institute, opportunamente ampliato per la terminologia archeologica. Vi è poi una difficoltà ben nota rispetto all'uso corrente di indicare intervalli di tempo con un nome anziché come un intervallo di tempo, come ad esempio *Iron Age* oppure *Roman Period*, che corrispondono a intervalli temporali diversi a seconda del luogo cui sono riferiti. In questo caso la soluzione adottata è stata quella del progetto PeriodO⁵ adattata al-

⁴ <http://www.cidoc-crm.org/>.

⁵ <http://perio.do/en/>.

le nostre specifiche esigenze. Questa soluzione prevede che ogni fornitore di archivi indichi in una tabella l'intervallo temporale a cui si riferiscono i nomi utilizzati nei propri dati con i relativi riferimenti spaziali, in modo da collocare in modo inequivocabile i dati archeologici nella fascia temporale di loro effettiva pertinenza.

Dal punto di vista tecnologico, la soluzione progettuale per la gestione dei dati prevede la creazione di un *cloud* archeologico in cui collocare il catalogo degli archivi e in cui poter trasportare i dati da analizzare trovati attraverso l'uso del catalogo e rielaborarli mediante una serie di strumenti forniti dal progetto. Ad esempio, un ricercatore può cercare nel catalogo tutti i riferimenti ad una determinata località, a un determinato intervallo temporale o a una parola chiave e trasportare nel *cloud* di ARIADNEplus i risultati di questa selezione. Questi risultati, probabilmente archivi testuali per la maggior parte, possono essere qui sottoposti a ulteriori elaborazioni, ad esempio di tipo NLP (Natural Language Processing) e quindi ulteriormente raffinati oppure collegati fra di loro dai risultati dell'NLP. Si verrà così a creare un VRE (Virtual Research Environment), cioè un ambiente collaborativo virtuale di ricerca in cui sarà possibile trovare e accedere a dati interoperabili per analizzarli, confrontarli e riutilizzarli, realizzando i principi noti come FAIR (Findable-Accessible-Interoperable-Reusable)⁶, ormai di vastissima applicazione come realizzazione del concetto di open science.

ARIADNEplus, come del resto ARIADNE, non è però solo dati e tecnologia informatica. Una componente fondamentale delle sue attività è il *networking*, cioè l'insieme delle azioni dirette verso la propria comunità e in generale verso la comunità di ricerca per promuovere la metodologia digitale nella ricerca archeologica, fornire la formazione necessaria e comunicare le proprie attività sostenendo l'adozione dei propri strumenti operativi. Ad esempio, vengono documentate e illustrate in seminari e workshop le politiche europee relative ai dati e al loro riuso, come ad esempio quelle derivanti dall'applicazione dei principi FAIR sopra ricordati. Altre iniziative di formazione riguardano gli strumenti tecnologici forniti da ARIADNEplus, la definizione e l'uso delle ontologie sviluppate dal progetto e così via.

ARIADNEplus non è un'iniziativa ristretta ai partner; anzi, promuove la collaborazione con altri enti e progetti attraverso la sottoscrizione di un *Memorandum of Understanding* che definisce i rispettivi impegni e diritti. Lo scopo è infatti quello di allargare l'aggregazione a quanti più archivi possibili, studiando anche le modalità con cui i ricercatori possono depositare i propri dati di ricerca e questi possono essere quindi inseriti nel catalogo di ARIADNEplus.

ARIADNEplus partecipa attivamente a iniziative e a eventi internazionali, organizzando proprie sessioni all'interno delle più importanti conferenze internazionali quali la Computer Applications and Quantitative Methods in Archaeology (CAA) e la conferenza annuale della European Association of Archaeologists (EAA), oppure

⁶ <https://www.go-fair.org/fair-principles/>.



Figura 3. I partecipanti all'assemblea virtuale di ARIADNEplus del 15 aprile 2020

organizzando propri eventi. Quest'attività è stata momentaneamente sospesa a causa della recente epidemia del coronavirus e ai conseguenti provvedimenti restrittivi. Però, queste difficoltà sono state superate organizzando eventi online con larghissima partecipazione (Fig. 3).

Questa modalità, dettata da un'esigenza contingente, ha indicato la possibilità di adottare metodi di comunicazione finora meno utilizzati come i *webinar*, i corsi online e i workshop virtuali. Il successo di questi strumenti ha suggerito una modifica della prassi comunicativa e di formazione di ARIADNEplus che consentirà una più larga partecipazione di tutti i ricercatori e gli operatori del settore⁷.

The paper describes the European project ARIADNEplus, continuation of the previous ARIADNE project. This new project, as the previous one, has the objective of overcoming the fragmentation of European archaeological digital archives available online by creating a catalogue. Compared to the previous project, ARIADNEplus significantly extends the geographic, thematic and temporal coverage. Data are provided by authoritative national institutions that guarantee their quality. The catalogue allows searching the metadata, which comply with international standards, according to various criteria, and then directly access the data at the data owner's repository. The retrieved data may then be further processed availing of various software tools also provided in the system. All the computing tools provided rely on the most recent IT technology. A substantial part of the project resources addresses users' training both on the use of the ARIADNEplus system and, in general, on the application in the archaeological domain of EU policies and strategies concerning open science.

⁷ ARIADNEplus è un progetto finanziato dalla Commissione Europea nel programma Horizon 2020, contratto n. 823914.

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2020

L'adozione del IIF nell'ecosistema digitale della Biblioteca Apostolica Vaticana

«DigItalia» 2-2020
DOI: 10.36181/digitalia-00017

Paola Manoni

Responsabile del Coordinamento dei Servizi Informatici della BAV

L'implementazione del IIF nell'ecosistema digitale della Biblioteca Apostolica Vaticana si avvia nel 2015 con la sperimentazione di un progetto pilota, a cui segue l'allestimento della biblioteca digitale (DVL), basata sull'adozione di standard per la gestione dei metadati e dei protocolli che consentono alle collezioni digitali di essere interoperabili. L'ecosistema successivamente si incrementa con software open source per la gestione di gallerie virtuali (Thematic Pathways on the Web) in cui selezioni di manoscritti, annotati secondo tecnica IIF, offrono agli studiosi percorsi di ricerca tematici.

Introduzione

Nel 1953, in una prefazione a un catalogo di una mostra di manoscritti medievali e rinascimentali italiani presso la Morgan Library, il grande storico dell'arte Bernard Berenson scriveva:

«Illuminated manuscripts are not easily accessible to the public and for good reasons. Most of them are still in codices and can be shown only two pages at a time. There is no other way unless the leaves are extracted and exhibited separately. This is not recommendable as it takes away from their character as book illustration and besides makes them liable either to lose or change color or to fade away from permanent exposure to light. Moreover many of them are too fragile, indeed so fragile that most keepers of illuminated manuscripts would prefer to keep them like houris in a harem. Not infrequently they bar access to their treasures by exacting from the common art lover a written declaration of where and when he means to publish the manuscripts he wishes simply to look at. The only compromise is to show them as best one can under glass, and only for a short time»¹.

A distanza di più di mezzo secolo l'irruzione del digitale nella riproduzione delle collezioni librarie ha capovolto la prospettiva in un modo che Berenson non avreb-

¹ Meta P. Harrsen – George K. Boyce, *Italian manuscripts in the Pierpont Morgan library: descriptive survey of the principal illuminated manuscripts of the sixth to sixteenth centuries, with a selection of important letters and documents*, introduction by Bernard Berenson, New York: The Pierpont Morgan Library, 1953 (Mediaeval and renaissance manuscripts in the Pierpont Morgan Library), p. 2.

be potuto certamente prevedere. La disponibilità delle immagini e il loro accesso remoto sono di esperienza comune, dove la visione in digitale può sostituire un originale sotto teca e in taluni casi migliorare la presentazione di un foglio nei suoi dettagli. E altrettanto di normale prassi è oramai la consultazione e lo studio fuori dalle sale di lettura. A meno dell'ineguagliabile analisi autoptica degli originali, che resta riservata agli specialisti, il "pubblico del web" ha accesso diretto, nei termini di Berenson, a quei tesori lungamente "sequestrati".

Ma ancora, restando sul filo del discorso dello storico dell'arte:

«Students of manuscripts live in a world apart and have developed a vocabulary, a phraseology, and nomenclature, a mode of reference by number which outsiders cannot easily follow»².

Berenson si riferisce a una conoscenza della catalogazione così come della classificazione e dell'ordinamento delle collezioni, che sono del tutto peculiari della tradizione delle biblioteche, e che tali permangono anche nell'apertura al web, dove la "consegna" digitale del materiale manoscritto è generalmente organizzata in modo del tutto analogo alle collezioni degli originali. Al di là della considerazione relativa alle necessarie conoscenze richieste per il reperimento e l'individuazione delle informazioni, l'approccio alla ricerca via web si sta ulteriormente trasformando, non solamente per la compresenza di cataloghi elettronici e immagini digitali, bensì in virtù del cambiamento paradigmatico che ha interessato la fruizione e la diffusione nel web degli oggetti digitali, partendo proprio dal patrimonio manoscritto digitalizzato. Questo mutamento è prodotto dall'introduzione del protocollo d'interoperabilità, noto con l'acronimo IIIF³, ormai largamente adottato dalle più importanti biblioteche digitali che offrono liberamente nel web milioni di immagini digitalizzate.

I manoscritti, al momento della loro produzione, hanno idealmente tracciato la trama della cultura "prima dell'introduzione della stampa". Dietro le barriere che menziona Berenson è stata costruita la conoscenza profonda di questi testimoni del passato, intesi nella loro rarità, nelle loro complesse relazioni con le collezioni di appartenenza e nelle connessioni delle varianti di testo delle opere che essi tramandano. Con l'allestimento dei cataloghi elettronici questo sapere, in tutto o in parte, si è trasferito nel web, e con le biblioteche digitali si è consentito un accesso diretto ai manoscritti. Il salto significativo del IIIF sposta ancora in avanti il punto di vista che un "occhio tecnologico" può cogliere e consegnare al nostro sguardo. Si potrebbe ritenere che per portare definitivamente le "*huri* fuori dagli *harem*" (per dirla nei termini metaforici impiegati da Berenson), la posta in gioco sia tutta in termini tecnologici. Non solamente in un senso evolutivo, per il migliora-

² *Ibidem*.

³ International Image Interoperability Framework, cfr. <<https://iiif.io>>.

mento delle procedure di digitalizzazione, bensì per un “impiego aperto” delle immagini, al fine di favorire la conoscenza e la scoperta di nessi non ancora esplorati di quella trama culturale di cui siamo eredi. La tecnologia del IIIF è un’ipotesi avvincente soprattutto se si considera l’idea, semplice ed elegante, attorno a cui si sviluppa: l’accesso ai depositi degli oggetti digitali attraverso l’esposizione nel web semantico dei dati e degli indirizzi (URI) delle immagini mediante l’impiego di API⁴ definite in modalità standard.

Di per sé il IIIF mira alla definizione e all’impiego normato di alcune API, di cui menzioniamo solo le due principali.

La prima, la cosiddetta *Image API*, fornisce un modo standardizzato per richiedere e consegnare immagini attraverso il web.

L’URI può specificare la regione, le dimensioni, la rotazione, le caratteristiche di qualità e il formato dell’immagine richiesta. La struttura standard dell’URI abilita, da parte delle applicazioni *client*, la possibilità di richiedere informazioni sull’immagine. In sintesi, possiamo dire che tale API è stata concepita per facilitare la richiesta e il riutilizzo sistematico delle risorse d’immagini presenti negli archivi digitali degli enti conservatori del patrimonio culturale.

L’obiettivo di *Presentation API*, è invece quello di fornire le informazioni necessarie relative all’oggetto digitalizzato per arricchire l’esperienza della visualizzazione online di oggetti costituiti da immagini. In altre parole, questa API ci fornisce un complesso di funzionalità per “presentare” oggetti digitali unitamente ai dati descrittivi che ne specificano i contenuti.

Non è questa la sede per entrare in altri dettagli tecnici o menzionare interamente l’insieme di API finora identificate dallo standard IIIF. Menzioniamo tuttavia le due API IIIF per il nesso relativo a quello “sguardo” sui tesori del passato a cui stiamo mirando. La conformità del patrimonio digitale a queste API consente agli strumenti di visualizzazione, software anch’essi conformi alle regole dello standard, di interpretare le immagini e di fornire tutto l’apparato funzionale che tipicamente correda l’ecosistema del digitale compatibile col IIIF.

Un caratteristico strumento in grado di interpretare le API IIIF è l’assai diffuso software *Mirador*.

Mirador è un visualizzatore di immagini configurabile, estensibile e facile da integrare in altre applicazioni informatiche. Presenta due caratteristiche importanti: la possibilità di inserire annotazioni all’interno delle immagini, consentendo la ricerca e la classificazione dei contenuti annotati, così come la possibilità di confrontare due o più immagini tratte da biblioteche digitali compatibili col protocollo d’interoperabilità, per l’analisi comparativa di manoscritti appartenenti alle più diverse collezioni.

⁴ API (informatica): acronimo di *Application Programming Interface*: in linea generale una API è costituita da definizioni e protocolli attraverso cui è possibile richiamare determinate funzioni di un software applicativo, facilitando altresì l’interazione tra programmi e sistemi diversi.

In altre parole, Mirador è uno strumento in grado di gestire risorse multiple basate su immagini, consentirne una navigazione strutturale e la sincronizzazione dei contenuti visivi.

Mirador è un software sviluppato quale progetto open source che riceve regolarmente contributi da una comunità internazionale di sviluppatori per migliorare quella “finestra” che offre una vista panoramica sulle più importanti collezioni del patrimonio digitale.

L’implementazione del IIIF in Biblioteca Vaticana

La sperimentazione dello standard IIIF inizia nel 2015 con l’allestimento di un piccolo progetto pilota realizzato attraverso la selezione di un campione di 10 manoscritti. Il successo dell’esperimento ebbe una risonanza molto importante perché da esso scaturì, l’anno successivo, la piattaforma di ricerca per i contenuti digitali (interagente con gli OPAC della Biblioteca), nota con l’acronimo DVL (DigiVatLib⁵), la cui architettura di sistema si fonda sulla tecnologia del IIIF. Sicché, con l’adozione dello standard, le collezioni digitalizzate della Vaticana hanno libera circolazione nel web semantico, al fine di consentire agli studiosi analisi comparative di oggetti digitalizzati, con la possibilità di visualizzare contemporaneamente più esemplari, così come di facilitare l’esplorazione delle immagini mediante le funzionalità offerte dai visualizzatori. In senso generale, un visualizzatore confor-

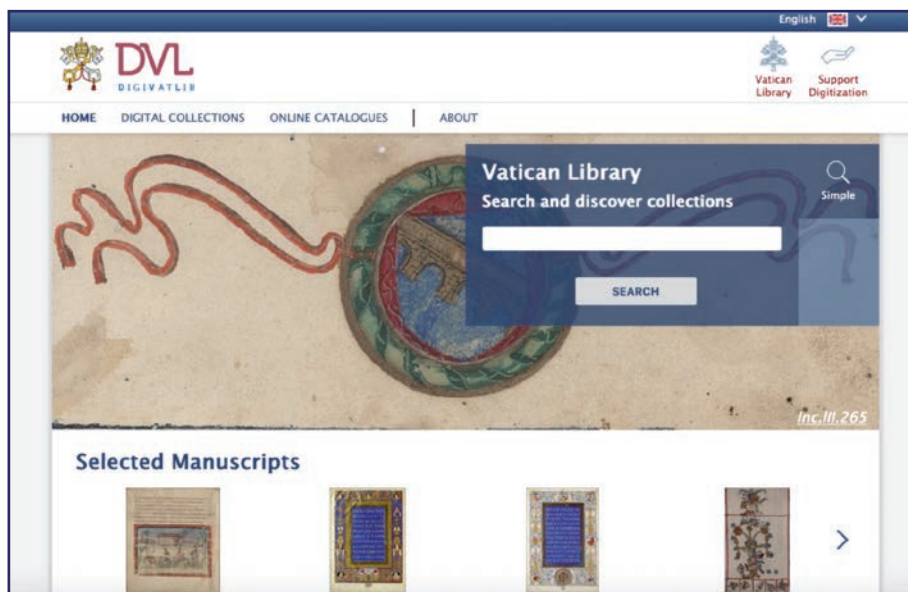


Figura 1.

⁵ <https://digi.vatlib.it>.

me al protocollo è al contempo uno strumento molto efficace per la fruizione delle immagini (profondità di zoom, rotazione/orientamento, gestione di contrasto/saturazione) e dei dati ad esse associati poiché consente anche la visualizzazione di tutti i contenuti (annotazioni e metadati) di cui si compone un oggetto digitale compatibile col IIIF.

Il progetto pilota aveva due obiettivi: non solo la capacità di esporre nel web semantico il cosiddetto *manifest*, in sintassi json, con cui si declina nel web tutta l'informazione disponibile relativa a immagini e dati del manoscritto digitalizzato, ma anche la possibilità di annotare, evidenziare i commenti, le trascrizioni, le note in relazione a ogni elemento di conoscenza presente nei fogli. Dall'esito positivo del progetto pilota, relativo alla gestione delle annotazioni secondo lo standard IIIF, scaturiva una ricerca triennale (2016-2019) dal titolo *Thematic Pathways on the Web: IIIF annotations of manuscripts from the Vatican collections*, che ha dato luogo all'implementazione di una piattaforma per la pubblicazione di "percorsi ragionati" attorno a selezioni di manoscritti annotati che si propongono nel web come una sorta di esposizioni virtuali⁶. Tale piattaforma, unitamente al DVL e agli OPAC, completa l'offerta informatica della BAV attualmente disponibile per lo studio dei manoscritti digitalizzati.

Il progetto pilota del 2015 e successivamente la ricerca triennale sono stati interamente sostenuti dalla fondazione statunitense Andrew W. Mellon che ha sovvenzionato le più importanti iniziative intraprese a livello internazionale per la promozione del IIIF, in particolare per quanto riguarda lo sviluppo della tecnologia per gli studi delle collezioni di libri rari e manoscritti.

La concezione della piattaforma dei *Thematic Pathways*, oggetto della ricerca, oltre alla produzione dei contenuti curatoriali, ha visto la collaborazione tecnica della Stanford University Libraries. La sfida posta dal progetto consisteva nella realizzazione di uno strumento in grado di gestire, al contempo, sia i contenuti di metadati e di annotazioni dei manoscritti su cui verte ciascun percorso a tema, sia la narrazione svolta in ogni percorso ed "esibita" in *exhibit* dello Spotlight, ovvero il software open source in cui sono gestiti i *Pathways*.

Nelle sue linee essenziali, ciascun percorso tematico offre tre diversi livelli di informazione:

- 1) una descrizione generale (introduzione, informazioni storiche ecc.) del tema scelto. Questo testo, arricchito con molti elementi visuali, è fornito in inglese e in italiano;
- 2) i metadati descrittivi e contenutistici corredati da note curatoriali, che illustrano ciascun manoscritto selezionato;
- 3) le annotazioni, i commenti, le analisi approfondite su parti dettagliate di un manoscritto e le trascrizioni di fogli.

⁶ <https://spotlight.vatlib.it>.

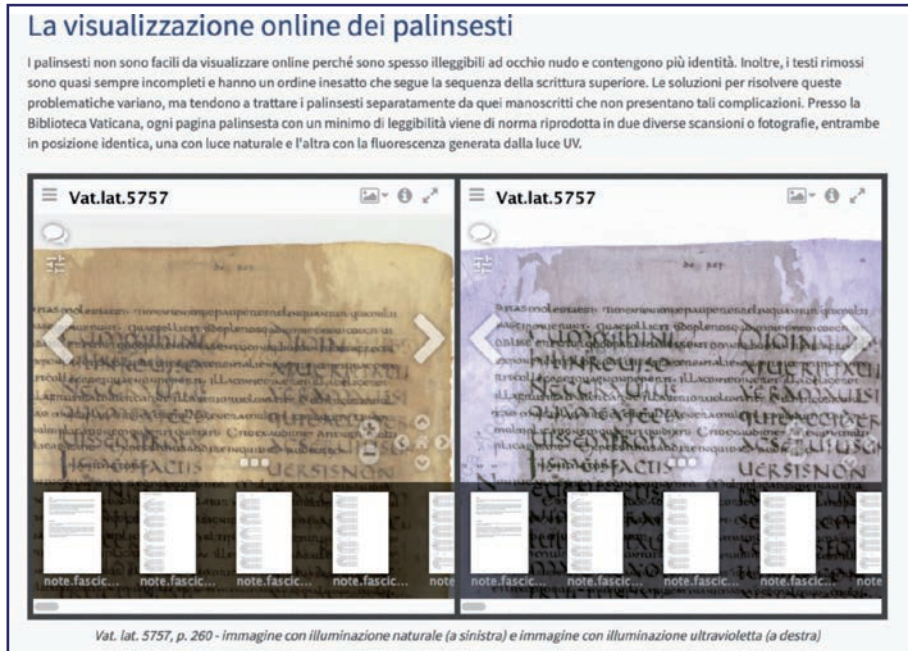


Figura 2. Esempio tratto dal percorso Palinsesti vaticani: comparazione di due immagini in Mirador incapsulato in una pagina web

Il software in cui sono state gestite le annotazioni è il Mirador, visibile come “oggetto incapsulato” nelle pagine web dei Percorsi.

Sin dalla prima sperimentazione del 2015 la Vaticana ha contribuito allo sviluppo della gestione delle annotazioni nell’ambito della comunità aperta. Il software consente la selezione di regioni di pagine/fogli secondo diverse forme, ma grazie all’esperienza maturata in Vaticana tali selezioni includono ora anche i ritagli di immagini dal perimetro irregolare e le annotazioni multiple, annidate l’una dentro l’altra all’interno di una medesima forma.

I manoscritti selezionati nei Percorsi, annotati via IIF, includono distinte tipologie di annotazioni, contrassegnate da colori diversi. Ad esempio, tutte le annotazioni il cui perimetro è di colore turchese trattano trascrizioni mentre il colore bianco denota le annotazioni di elementi iconografici.

L’esplorazione delle annotazioni, dei metadati e delle schede di approfondimento è gestita nel già menzionato software Spotlight attraverso un apparato di indici e una struttura di filtri (definiti, ad esempio, per provenienza, datazione, lingua, nomi occorrenti nelle descrizioni), al fine di consentire ulteriori approfondimenti di ricerca. In questa struttura sono inoltre elencati i termini (*tag*, in italiano e in in-

²⁷ Lorcan Dempsey, *Library Collections in the Life of the User: Two Directions*, «Liber Quarterly», 26 (2017) n. 4, p. 338 -359.

glese) con cui sono state categorizzate le annotazioni. Ad esempio, l'indicizzazione dei soggetti iconografici identificati oppure la presenza di capilettera o di molte altre tipiche caratteristiche rilevate, consente il recupero immediato dei fogli dei manoscritti in cui tali elementi categoriali compaiono associati alle annotazioni. L'obiettivo della ricerca triennale, oltre al complesso allestimento del sistema, si è concentrato sulla pubblicazione del primo nucleo di percorsi a tema che sono:

Corsi di paleografia: *Paleografia Latina (dall'Antichità al Rinascimento)*, a cura di A. M. Piazzoni, e *Paleografia Greca (dall'Antichità al Rinascimento)*, a cura T. Janz. Si tratta invero di due percorsi che scaturiscono dalla selezione di due distinte sezioni di manoscritti, ma entrambi hanno una medesima finalità: sperimentare un nuovo metodo per l'insegnamento della paleografia sostituendo il tradizionale repertorio iconografico, cartaceo o "statico", con cui si dimostra l'evoluzione delle scritture, con la visualizzazione "dinamica" dei manoscritti digitalizzati e annotati attraverso Mirador. In questi due percorsi è ovviamente molto consistente la presenza di trascrizioni relative a fogli specifici.

L'impiego del IIF in questi corsi di paleografia può essere considerato come un ausilio di auto-apprendimento per leggere e comprendere le antiche scritture negli alfabeti latino e greco.

Classici Latini (Evoluzione e trasmissione di opere classiche), a cura di M. Buonocore.

Il percorso intende far *vedere* i classici latini offrendo una sorta di *campionatura* dei principali manoscritti vaticani "illustrati", latori di testi classici latini, così da consentire un dialogo costruttivo e ricco di confronti con altri testimoni della cultura europea tra tarda antichità, medioevo e umanesimo.

Palinsesti Vaticani (Recupero digitale di identità cancellate), a cura di A. Németh. La Biblioteca Apostolica Vaticana ha individuato nelle proprie collezioni oltre 550 manoscritti con palinsesti, ossia fogli di pergamena cancellati e poi riutilizzati. Il percorso si snoda attorno a questo cospicuo materiale e si concentra sulla selezione di 24 manoscritti proponendone una ricerca approfondita, al fine di recuperare le identità cancellate di questi palinsesti. I codici scelti coprono la storia delle scritture latine e greche dal IV-V secolo in poi e abbracciano una vasta gamma di generi letterari e ambiti culturali precedenti al XII secolo, per lo più relativi al mondo bizantino.

La Biblioteca di un 'principe umanista' (Federico da Montefeltro e i suoi manoscritti), a cura di M. G. Critelli.

Il percorso, attraverso la descrizione di una selezione di manoscritti, intende illustrare le caratteristiche principali della collezione urbinata, mirabile esempio di biblioteca signorile quattrocentesca.

Infine l'*exhibit Overview* consente di svolgere ricerche incrociate su tutti i manoscritti: metadati, note curatoriali e annotazioni presenti nei diversi percorsi.

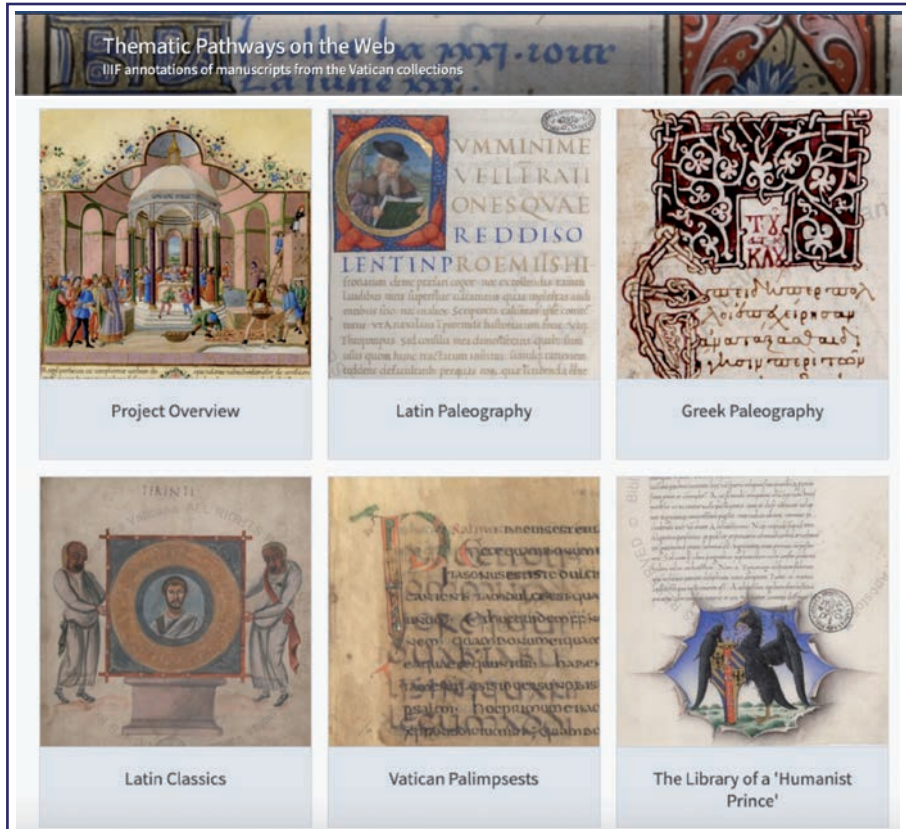


Figura 3.

Oltre agli strumenti di ricerca precedentemente menzionati, in ciascuna di queste strutture sono proposti (alla voce di menu "browse") delle esplorazioni dei materiali suddivisi per categorie al fine di evidenziare aspetti peculiari dei manoscritti selezionati quali, ad esempio, provenienze, datazioni, tipologie librarie.

Nei tre anni di lavoro sono state complessivamente prodotte, per 256 manoscritti selezionati, 26.233 annotazioni: un numero di dati ragguardevole, che pone questo progetto al primo posto tra le sperimentazioni sulla gestione delle annotazioni secondo lo standard IIIF.

La Biblioteca ha inteso coinvolgere i visitatori del proprio sito web sulla possibilità di utilizzare questi manoscritti annotati in IIIF, secondo specifici percorsi tematici: quelli finora realizzati e auspicabilmente future ricerche, fornendo strumenti per scoprire e confrontare i materiali digitalizzati. Oltre alla produzione dei percorsi espositivi, pubblicati come esito del progetto triennale, è attualmente disponibile

anche un'altra esposizione relativa a una selezione di materiali tratti da una precedente iniziativa di digitalizzazione, finanziata dalla Fondazione Polonsky. Grazie a questo sostegno, tra il 2012 e il 2017, le Bodleian Libraries dell'Università di Oxford e la Biblioteca Vaticana hanno svolto un progetto di digitalizzazione congiunto. Oltre un milione e mezzo di pagine tratte da importanti collezioni sono state rese disponibili online per gli studiosi e per il pubblico generale. Il progetto di digitalizzazione si è concentrato su tre gruppi principali di testi: Manoscritti ebraici, Manoscritti greci e Incunaboli o Cinquecentine, oltre a una significativa selezione di Manoscritti latini della Vaticana. In questa mostra viene presentata una selezione di materiali appartenenti al Progetto⁷, in alcuni casi arricchiti da annotazioni, scelti secondo le categorie mostrate nella sezione "Browse".

Il crescente numero di visitatori della piattaforma dello Spotlight della Vaticana è un indicatore favorevole per la prosecuzione di questo metodo di pubblicazione e di ricerca web, mediante l'impiego del IIF. L'auspicio è dunque arricchire tecnicamente quanto finora realizzato, gestire funzionalità per porre al servizio degli studi nuovi strumenti che consentano di diffondere la conoscenza del patrimonio culturale di cui siamo custodi.

Thematic Pathways on the Web

Crediti

Curatori: Marco Buonocore, Maria Gabriella Critelli, Timoty Janz, Andras Németh, Ambrogio M. Piazzoni

Collaboratori: Anna Berloco, Ilaria Maggiulli, Lola Massolo, Eva Ponzi, Domenico Surace

Gestione del Progetto: Benjamin Albritton, Paola Manoni

Gestione tecnica: Erin Fahy, Domenico Izzo, Sean Martin, Riccardo Moroni, Jack Reed

Team di sviluppo: Scott Bailey, Chris Beer, Javier de la Rosa, Jessie Keck, Mark Matienzo, Stu Snyderman, Camille Villa, Drew Winget

UX design: Gary Geisler

Coordinamento (collaboratori): Anna Berloco

Traduzioni in inglese: sr. Maria Panagia Miola

Architettura e Relazioni esterne: Tom Cramer, Michael A. Keller

Direzione: Tom Cramer, Michael A. Keller, Paola Manoni, Mons. Ceare Pasini, Ambrogio M. Piazzoni

⁷ <http://bav.bodleian.ox.ac.uk/>.

The implementation of IIF standards in the digital ecosystem of the Vatican Apostolic Library starts in 2015 with the experimentation of a pilot project followed by the setting up of the digital library (DVL), based on the adoption of standards for metadata and protocols that allow the digital collections to be interoperable. The ecosystem subsequently increases with open source software for the management of virtual exhibits (Thematic Pathways on the Web) in which selections of annotated manuscripts, according to the IIF technique, offer to scholars thematic research pathways.

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2020

I-Tal-Ya books: il censimento digitale dei libri ebraici d'Italia

Un progetto di cooperazione dell'Unione delle comunità ebraiche italiane, della Biblioteca nazionale centrale di Roma e della Biblioteca nazionale di Israele

«DigItalia» 2-2020
DOI: 10.36181/digitalia-00018

Andrea De Pasquale - Biblioteca nazionale centrale di Roma

Si presenta il progetto recentemente denominato I-Tal-Ya books, iniziato nel 2018, realizzato dall'Unione delle Comunità ebraiche italiane (UCEI) come capofila, con la partnership della Biblioteca nazionale centrale di Roma (BNCR), della Biblioteca nazionale di Israele (NLI) e della Rothschild Foundation Hanadiv Europe (RFHE) che ha garantito il supporto finanziario.

L'obiettivo è la realizzazione di un censimento, in un database bilingue, di tutti i volumi presenti nelle biblioteche italiane, dalle origini della stampa ebraica fino al 1960, utilizzando come riferimento il catalogo della NLI per la trascrizione ebraica e il sito Worldcat della Library of Congress, per la traslitterazione in caratteri latini. Ai dati bibliografici si aggiungono anche quelli di esemplare (segnalazioni di marginalia, censure, note di possesso, legature di pregio) e la localizzazione. I volumi inoltre sono digitalizzati nelle pagine principali e più significative. Tutte le informazioni sono allocate nella teca digitale della Biblioteca nazionale centrale di Roma, opportunamente adattata. Si prevede che l'intero progetto includerà un numero complessivo di 35.000 volumi di 14 comunità ebraiche e 25 istituzioni statali.

È da sempre auspicata e sollecitata la realizzazione di un censimento che permetta di avere contezza del patrimonio bibliografico ebraico italiano a stampa attualmente esistente presso le comunità e le istituzioni culturali, dopo le dispersioni e le distruzioni subite nell'ultimo conflitto bellico, ma fino ad oggi, data anche la grande mole di materiale, tutti i tentativi si sono arenati sul nascere. Anche nelle biblioteche statali, di cui molte detengono importanti fondi ebraici, dimostrando il forte interesse rivolto nel corso del tempo all'ebraismo in genere, agli studi biblici e alle tradizioni orientali, i progetti di censimento si sono limitati agli istituti tradizionalmente più noti e per lo più ai fondi manoscritti, tralasciando generalmente i libri posteriori al XVII secolo.

Ci si riferisce in particolare alla Biblioteca Palatina di Parma, che conserva il fondo

appartenuto all'abate Gian Bernardo De Rossi, comprendente anche, oltre a straordinari manoscritti, un'eccezionale raccolta di edizioni antiche, tra cui l'unica copia sopravvissuta del più antico testo stampato in caratteri ebraici con data certa (Reggio Calabria, 1475)¹ e la serie quasi completa di edizioni prodotte dalla famiglia Soncino, che è stata oggetto del progetto europeo *Judaica Europeana* ed ha riguardato la digitalizzazione degli incunaboli e di parte delle cinquecentine; o alla Biblioteca nazionale universitaria di Torino, dove si trova un fondo di opere a stampa - tra cui pezzi provenienti dalle collezioni di Casa Savoia e altri dalla donazione dell'abate Tomaso Valperga di Caluso - che costituisce una fonte unica per studiare la storia della tipografia ebraica, che è stato recentemente valorizzato con una mostra intitolata *Judaica Pedemontana*². Si segnalano anche le numerose cinquecentine ebraiche presenti nelle biblioteche Casanatense³ e Angelica⁴, che furono oggetto di cataloghi a stampa, o la raccolta dei fratelli Lattes della Biblioteca nazionale Braidense⁵, che ancora giace senza specifici strumenti descrittivi. Si possono citare numerosi altri esempi, tanto che dall'indagine effettuata, attraverso un questionario, preliminare al progetto, inviato a tutte le biblioteche statali, comunali, delle Comunità o delle istituzioni ebraiche, risultano essere presenti almeno 35.000 volumi ancora non censiti esistenti presso una quarantina di biblioteche italiane. L'idea di procedere alla realizzazione di un catalogo unico per il libro ebraico in Italia deriva non solo dalla constatazione della presenza presso le biblioteche di numerosi fondi ebraici non catalogati a causa della cronica carenza di fondi, ma anche dall'estrema scarsità di bibliotecari esperti di catalogazione in lingua ebraica che di fatto ha impedito negli anni la realizzazione di schedature per oggettiva incapacità. Il nuovo progetto, che ha recentemente assunto il nome di I-Tal-Ya books, ha

¹ Andrea De Pasquale, *I fondi ebraici e orientali della Biblioteca Palatina di Parma*, in *Exoticis linguis. Libri ebraici e orientali della Biblioteca Palatina di Parma*, Parma: MUP, 2009, p. 9-68; Id., *Il fondo ebraico della Biblioteca Palatina di Parma: catalogazione, digitalizzazione e valorizzazione*, in *Testimonianze della cultura ebraica: ricerca, valorizzazione, digitale. Il progetto Judaica Europeana. Atti del convegno internazionale. Fermo, 6-7 ottobre 2011*, a cura di G. Capriotti e P. Feliciati con la collaborazione di C. Ferrara, Macerata: Eum, 2011, p. 291-294.

² Andrea De Pasquale, *Le edizioni ebraiche della Biblioteca Nazionale Universitaria di Torino*, in *Judaica Paedemontana. Libri e argenti da collezioni piemontesi*, Cinisello Balsamo: Silvana Editoriale, 2015, p. 29-35. Inoltre: Id., *Le collezioni di libri a stampa ebraici della Biblioteca nazionale universitaria di Torino e della Biblioteca Palatina di Parma*, in *Il collezionismo di libri ebraici tra XVII e XIX secolo. Atti del convegno*, Torino, 27 marzo 2015, a cura di C. Pilocane e A. Spagnoletto, Supplemento al vol. 82, n. 2-3 (maggio-dicembre 2016) de «La Rassegna mensile di Israel», Firenze: Giuntina, 2017, p. 23-26.

³ Biblioteca Casanatense. Roma, *Le cinquecentine ebraiche. Catalogo*, a cura di A. Di Nola, Milano: Aisthesis, 2001; Margherita Palumbo, *Il fondo ebraico della Biblioteca Casanatense*, in *Il collezionismo di libri ebraici tra XVII e XIX secolo. Atti del convegno*, cit., p. 37-52.

⁴ Biblioteca Angelica. Roma, *I libri ebraici della biblioteca angelica. I. Incunaboli e cinquecentine*, a cura di E. Abate e S. De Gese, Roma: Istituto Italiano per l'Africa e l'Oriente, 2005.

⁵ *I fondi speciali delle biblioteche lombarde: censimento descrittivo, I: Milano e provincia*, a cura dell'Istituto Lombardo per la Storia della Resistenza e dell'Età contemporanea, Milano: Editrice Bibliografica, 1995.

preso forma nel 2018 e vede la partecipazione dell'Unione delle Comunità ebraiche italiane (UCEI), come capofila, e la partnership della Biblioteca nazionale centrale di Roma, della Biblioteca nazionale di Israele (NLI) e della Rothschild Foundation Hanadiv Europe (RFHE), che ha garantito il supporto finanziario.

Il suo obiettivo è quindi quello di censire, in un database bilingue, tutti i volumi presenti nelle istituzioni, estesi su un arco cronologico amplissimo, dalle origini della stampa ebraica fino al 1960, utilizzando come riferimento il catalogo della



Figura 1. Homepage della Teca della Biblioteca nazionale centrale di Roma

Elenco dei campi	Fonte	Commenti	MARC2
Autore in alfabeto ebraico	NLI	L'autore in ebraico è mutuato dal catalogo della NLI secondo lo standard israeliano di catalogazione ebraica..	1xx+7xx
Autore in alfabeto latino	OCLC	Il nome dell'autore è mutuato dalle informazioni di autorità gestite dalla NLI ed è conforme allo standard LCNA..	1xx+7xx
Titolo in alfabeto ebraico	NLI	Il titolo in ebraico è il titolo principale ed è mutuato dal catalogo della NLI.	245
Titolo traslitterato	OCLC	Il titolo in alfabeto latino è la traslitterazione del titolo ebraico ed è mutuato da un dato OCLC. La traslitterazione è conforme allo standard LC (campo 245).	245
Dichiarazione di pubblicazione	OCLC	La dichiarazione di pubblicazione comprende editore, luogo e anno ed è mutuata dai dati OCLC..	260/264
Descrizione fisica	OCLC	La descrizione fisica è mutuata dal dato OCLC, la lingua di questo campo è l'inglese.	300
Numero di scaffale	photo	Il numero del libro nella biblioteca originaria è registrato nell'immagine scattata. Potrebbero risultare casi molteplici di questo campo per ciascun volume, per ciascuna copia e per più biblioteche. Tuttavia, ogni riga si riferisce a un'unica copia..	
Note	OCLC	Le note sono mutate dal dato OCLC e sono in italiano	500
Numero NLI	NLI	Questo campo contiene il numero del sistema NLI.	1
Numero OCLC	OCLC	Questo campo comprende il numero di identificazione OCLC, nel caso di un dato esatto	035
Nome del file della fotografia	photo	Vi è una riga per ogni immagine	
Link alla digitalizzazione	NLI	Link a una copia digitalizzata del libro, se disponibile	856/ROS
Provenienza: (Y/N) S/N	identification	Nell'ambito dell'identificazione del libro, questo sarebbe contrassegnato Y per ogni pagina che include informazioni sulla provenienza..	
Frontespizio mancante: Y/N	identification	Nell'ambito dell'identificazione, questo campo è contrassegnato in caso di frontespizio mancante.	

NLI per la trascrizione ebraica e il sito Worldcat della Library of Congress, che detiene i record di oltre 72.000 biblioteche di 170 nazioni, per la traslitterazione in caratteri latini. Ai dati bibliografici si aggiungono anche quelli di esemplare (segnalazioni di *marginalia*, censure, note di possesso, legature di pregio) e la localizzazione.

I volumi vengono fotografati da professionisti nelle pagine principali e identificative (piatto anteriore della coperta; frontespizio; due pagine a metà volume; pagina finale numerata; colophon se esistenti; *marginalia*, note di possesso, censure; dorso, se reca informazioni) e le immagini vengono quindi caricate in Google Drive, il cui accesso è stato abilitato per tutte le persone che fanno parte del progetto; se-

gue quindi la fase di identificazione, curata da bibliotecari specializzati, individuati a seguito di una selezione pubblica, che si avvalgono della consulenza di NLI per i casi dubbi, la quale anche provvede a schedare in maniera esaustiva, incrementando i propri database, le edizioni ancora sconosciute; vengono quindi costruiti i metadati che consentono l'esatta individuazione delle edizioni, completati dai link ai cataloghi di NLI e della Library of Congress, e costruito l'abbinamento alle immagini digitalizzate. Ogni pubblicazione viene registrata una volta sola e, se più edizioni sono rilegate insieme, si schedano le singole unità.

I dati rilevati sono i seguenti:

- Autore traslitterato
- Autore in ebraico
- Titolo traslitterato
- Titolo in ebraico
- Altri Autori
- Dati di pubblicazione traslitterati
- Dati di pubblicazione in ebraico
- (Record number) nella NLI
- Numero di registrazione OCLC⁶
- Link al catalogo della NLI
- Link alla Biblioteca dove è posseduto il volume
- Provenienza (se esistono indicazioni, in qual caso si indica "sì")
- Missing (mancanza del frontespizio, in qual caso si indica "sì")
- Segnatura

Le informazioni, predisposte in tabelle excel (Metadata Spreadsheet), subiscono quindi l'*ingestion* nella teca della Biblioteca nazionale centrale, che effettua il controllo di congruità, l'abbinamento con le relative immagini digitali, e garantisce la loro visibilità attraverso l'interfaccia utente, opportunamente dotata di apposita sezione dedicata e specifica maschera di ricerca, in modo che studiosi o persone interessate possano esaminare il volume, interrogare per autore, città di stampa, editore, sapere se in quel volume ci siano note di censura o di proprietà e capire in quale biblioteca si conserva.

In particolare nella pagina dedicata al progetto si possono visualizzare tre box, dedicati rispettivamente a libri antichi (fino al 1830); libri moderni (1831-1960); lista⁷.

⁶ La ricerca viene effettuata con OCLC Connexion, uno strumento di catalogazione full-service che consente alle istituzioni membro di cercare, creare e modificare le voci bibliografiche e di autorità di alta qualità e quindi di condividerle con l'intera cooperativa OCLC. Nell'ambito di questo progetto, l'OCLC Connexion è utilizzato per cercare dati e copiare i campi "romanizzati" predefiniti (traslitterati) nel Metadata Spreadsheet. L'obiettivo della ricerca nell'OCLC Connexion è individuare titoli traslitterati e informazioni su impronte e descrizioni fisiche in caratteri latini. È consentito copiare un titolo, ipotizzando che sia il titolo corretto, sebbene tutti gli altri dettagli non siano esattamente i medesimi.

⁷ <http://digitale.bnc.roma.sbn.it/tecadigitale/progettoVolumiEbraici>.

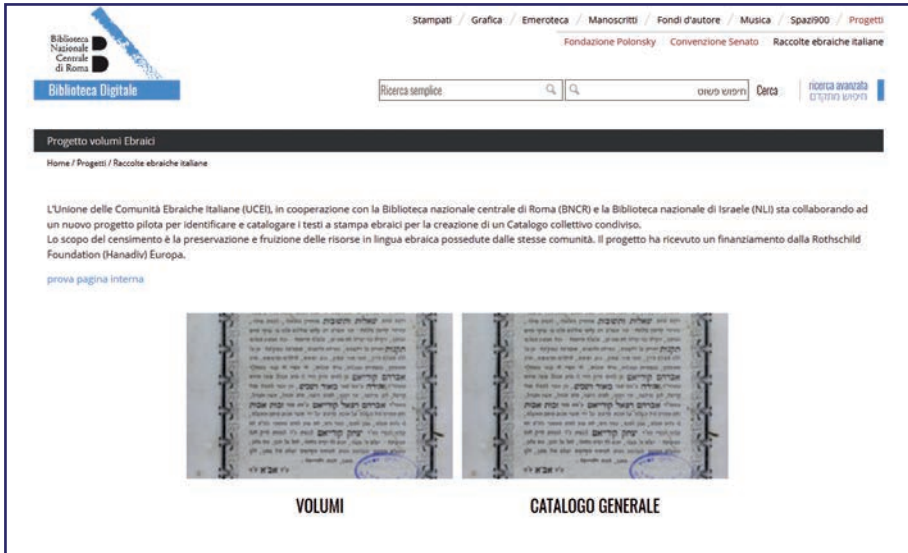


Figura 2. Raccolte ebraiche italiane - Teca della Biblioteca nazionale centrale di Roma

Per ogni box si apre una pagina di risultato della ricerca con tutti i volumi elencati in ordine alfabetico (in forma traslitterata); è stato scelto di fondere i volumi antichi e moderni nel box denominato Lista; in quest'ultima si visualizzano due colonne: quella di sinistra mostra tutti i volumi in ebraico traslitterato in ordine alfabetico per titolo; in quella di destra appaiono gli stessi titoli in alfabeto ebraico in ordine alfabetico per titolo.

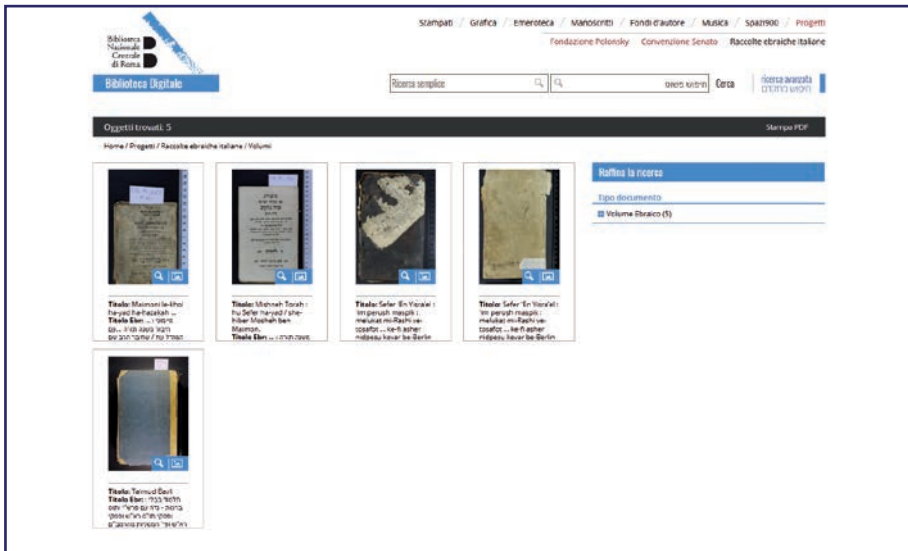


Figura 3. Raccolte ebraiche italiane - Teca della Biblioteca nazionale centrale di Roma

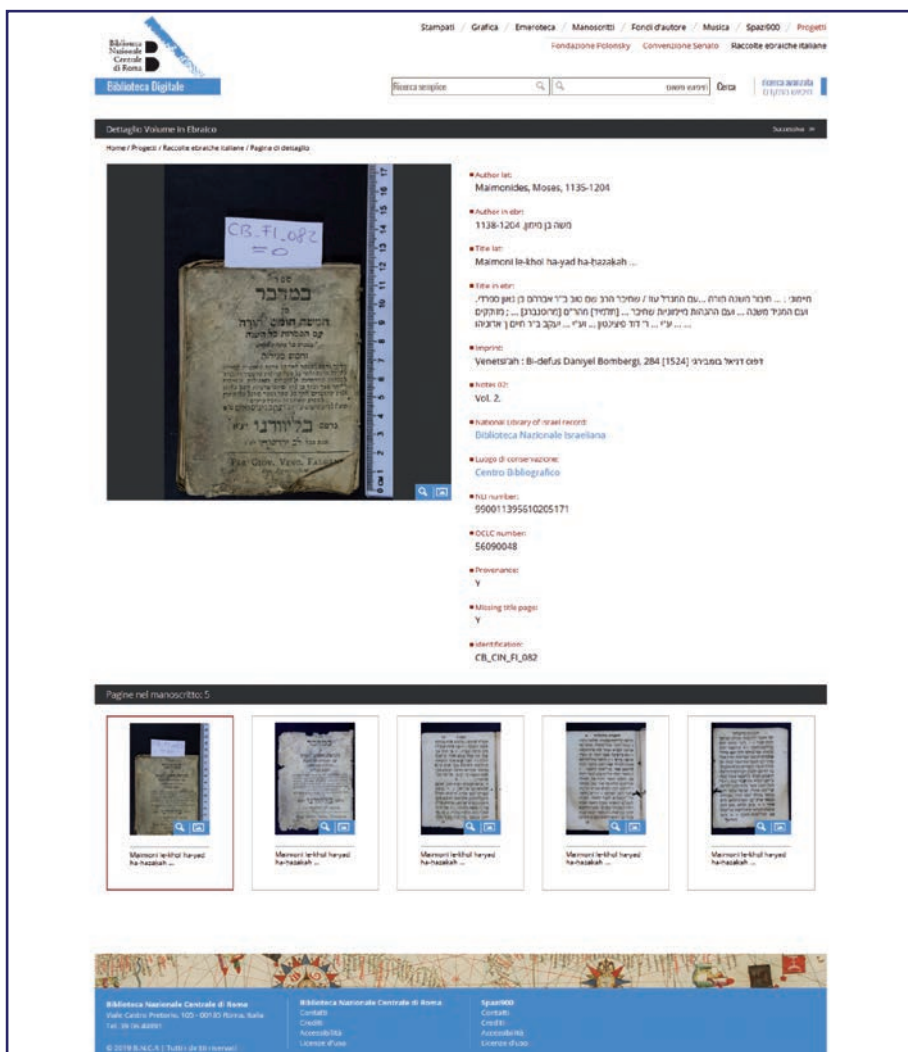


Figura 4. Raccolte ebraiche italiane - Teca della Biblioteca nazionale centrale di Roma

Per quanto riguarda le possibilità di ricerca, nella home del progetto è presente uno specifico box attraverso cui è possibile ritrovare i record per autore e titolo solo in caratteri latini. Invece, nella pagina della sottosezione Lista, ci sono due diversi box: uno per la ricerca in caratteri latini e un altro in caratteri ebraici, dove, in entrambi i casi, è possibile fare interrogazioni per autore e titolo. Quando si opera nel box in ebraico, si apre una tastiera virtuale con caratteri ebraici per poter scrivere in tale lingua. Nell'aprile 2018 il gruppo di lavoro si è riunito a Gerusalemme presso la National Library e ha posto le basi organizzative per iniziare il progetto di catalogazione. Inizialmente si è deciso di svolgere una fase pilota concentrandosi su duemila volumi presenti in alcune biblioteche a Roma, Milano, Torino e Genova.

Nel progetto pilota del censimento dei volumi ebraici in Italia hanno partecipato le seguenti comunità ebraiche:

- UCEI – Centro bibliografico “Tullia Zevi” di Roma
- UCEI – Pitigliano (Firenze)
- Genova - Comunità ebraica
- Milano - Fondazione Centro di Documentazione Ebraica Contemporanea (CE.D.E.C.)
- Torino - Comunità ebraica

Sono state lavorate tre tipologie di materiale:

- 149 cinquecentine;
- 1.283 volumi antichi (fino al 1830);
- 817 volumi moderni (dal 1931 al 1960);

Per un totale di 2.249 volumi con circa 15.000 immagini.

Stampati / Grafica / Ebraica / Manoscritti / Fondi Craione / Musica / S042900 / Progetti

Fondazione Polonsky / Convenzione Senato / Raccolte ebraiche italiane

Biblioteca Nazionale Centrale di Roma

Biblioteca Digitale

Titolo Cerca [ricerca semplice](#) 0/0/0 12/1/1

Titolo (EBF)

Autore

Autore (EBF)

Autori aggiuntivi

Autori aggiuntivi (EBF)

Città

Edizione

Anno da - a

Filtro Tessuto

Biblioteca

N° scaffale

Progetto volumi ebraici

Home / Progetti / Raccolte ebraiche italiane

L'Unione delle Comunità Ebraiche Italiane (UCEI), in cooperazione con la Biblioteca nazionale centrale di Roma (BNCR) e la Biblioteca nazionale di Israele (BNI) sta collaborando ad un nuovo progetto pilota per identificare e catalogare i testi a stampa ebraici per la creazione di un Catalogo collettivo condiviso. Lo scopo del censimento è la preservazione e fruizione delle risorse in lingue ebraiche possedute dalle stesse comunità. Il progetto ha ricevuto un finanziamento della Rothschild Foundation (Hananah) Europa.

[prova pagina interna](#)

VOLUMI

CATALOGO GENERALE

Biblioteca Nazionale Centrale di Roma
Viale Cassini 156/a - 00157 Roma, Italia
Tel. 06 4981991

Biblioteca Nazionale Centrale di Roma
Contatti
Orario
Accessibilità
Licenze/Più

Spazio999
Contatti
Orario
Accessibilità
Edizione Offerta

© 2019 B.N.C.E. | Tutti i diritti riservati

Figura 5. Raccolte ebraiche italiane - Teca della Biblioteca nazionale centrale di Roma

Il lavoro è stato effettuato grazie anche alla collaborazione di volontari che hanno proceduto ad individuare i pezzi, ordinarli, spolverarli, numerarli con un numero identificativo, che serve per tracciare lo specifico volume in tutto il processo, l'apposizione di etichette con codice a barre contenente il numero identificativo, e predisporli per la campagna fotografica.

Il personale coinvolto, compresi i volontari e i fotografi, è stato formato con specifiche giornate mirate, inerenti la tipografia ebraica e le operazioni pratiche del progetto, da esperti dell'UCEI, della Biblioteca nazionale centrale di Roma e soprattutto della NLI, che ha sviluppato un programma di formazione e sviluppo "At the Source" (Alla Fonte) per aiutare i bibliotecari e gli archivisti che trattano le collezioni ebraiche a preservare e a conservare il patrimonio ebraico in Europa.

Al termine della fase pilota, ottobre 2018-marzo 2019, il Comitato scientifico, opportunamente individuato, è stato invitato a collegarsi alla Teca per analizzare i primi risultati e consentire al gruppo di lavoro di raccogliere feedback e suggerimenti. Grazie all'analisi del Comitato scientifico è stato possibile migliorare la funzione di ricerca nella Teca in modo da poter selezionare per autore, titolo, luogo di edizione e tipografia/editore, sia in caratteri latini che in ebraico. Cercare un nome di autore o di luogo che, nel corso dei secoli e a causa della traslitterazione può essere stato espresso in forme diverse, costituisce un'evidente difficoltà, ma attraverso l'uso di authority file la funzione di ricerca è in grado di individuare un nome anche se scritto in modi differenti (Abramo, Avraam, Abraham; Venezia, Venice, Vendig, Venetia, etc).

Autore	Titolo	Lingua	Edizione	Anno
Talmud Bavli		עברית		
Ben Meir, Isaac ben Solomon, 1445-1513 or 1516	Sefer 'in yotzer: 'in penush mishpat: mekullat me-Bavli ve-tsofat ... ha-R' Asher nedava le-aver be-Bavli ve-atah me-hachshu be-kamati tsofat ...	עברית	Frankfurt am Main, 1807	[1722]
Ben Meir, Isaac ben Solomon, 1445-1513 or 1516	Sefer 'in yotzer: 'in penush mishpat: mekullat me-Bavli ve-tsofat ... ha-R' Asher nedava le-aver be-Bavli ve-atah me-hachshu be-kamati tsofat ...	עברית	Frankfurt am Main, 1807	[1722]
Maimonides, Moses, 1135-1204	Maimoni le-hal hayad ha-bechukim ...	עברית	Frankfurt am Main, 1807	[1524]
Maimonides, Moses, 1135-1204	Mishneh Torah: hu Sefer hayad / she-hiber Moshen ben Maimon.	עברית	Frankfurt am Main, 1807	[1555-1551]

Figura 6. Raccolte ebraiche italiane - Teca della Biblioteca nazionale centrale di Roma

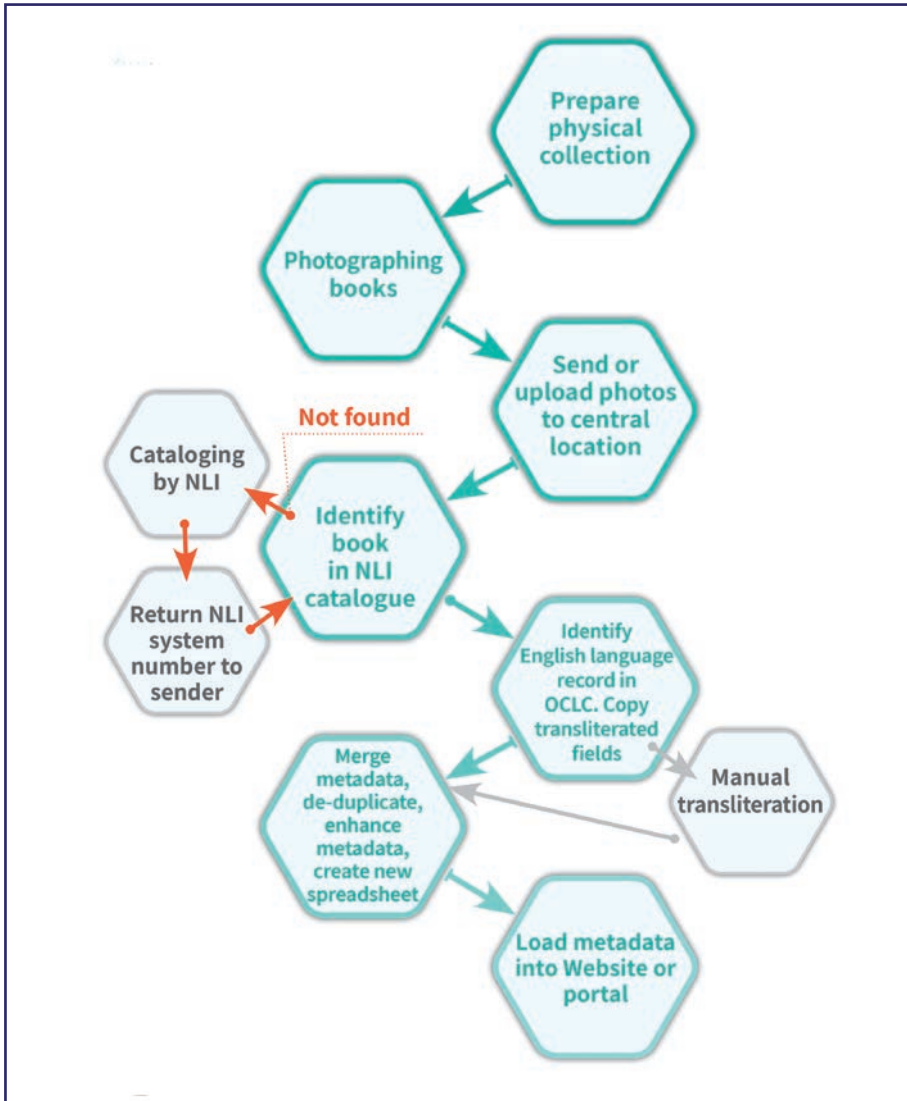


Figura 7.

La Rothschild Foundation Hanadiv Europe nell'aprile 2019 ha chiesto di presentare la richiesta di contributo finanziario per completare la catalogazione. Per gestire al meglio il progetto, RFHE ha fornito il sistema di gestione per la catalogazione e digitalizzazione GOOBI della Scandata, che consente l'interazione tra tutti gli attori del processo in oggetto: i fotografi caricano sul sistema le foto, effettuano il controllo qualità, verificano la correttezza del procedimento e il collegamento immediato sulle foto da parte dei catalogatori del progetto e della NLI. Questo sistema permette quindi al gruppo di lavoro di operare sulla stessa

piattaforma informatica in tutte le fasi progettuali avendo a disposizione le informazioni necessarie per lavorare in modo sinergico.

A settembre 2019 la RFHE ha approvato la seconda fase del progetto che prevede la catalogazione di altri 15.000 volumi circa, da realizzare nel corso del 2020 e 2021. A questa il team di progetto spera che segua la terza e ultima fase di completamento che consentirà la registrazione di tutto il patrimonio librario in ebraico custodito nelle biblioteche italiane: si prevede che l'intero progetto includerà un numero complessivo di 35.000 volumi di 14 comunità ebraiche e 25 istituzioni statali.

Sono in corso anche modifiche alla struttura della Teca, per agevolare la consultazione degli utenti: le distinte basi dati per volumi antichi e moderni sono state fuse in un solo catalogo e la sezione Lista sarà sostituita con un Catalogo generale che visualizzerà le informazioni sia in ebraico che traslitterate. L'aggiornamento della Teca della Biblioteca nazionale centrale di Roma sarà programmato periodicamente in modo da rendere disponibili da subito i primi volumi che verranno catalogati senza dover attendere la fine del progetto.

Infine, se la finalità del progetto è certamente l'inventario e il catalogo di tutti i volumi ebraici affinché siano patrimonio di studiosi, accademici e studenti, non si può non menzionare l'importanza che ha l'emersione pubblica di questi materiali, fra cui rarità e pezzi unici, consentendo di assicurare questo patrimonio da furti e vendite sul mercato illegale.

The project recently called I-Tal-Ya books, started in 2018, carried out by the Union of Italian Jewish Communities (UCEI) as leader, with the partnership of the National Central Library of Rome and the National Library of Israel (NLI) is presented. The Rothschild Foundation Hanadiv Europe (RFHE) provided financial support.

The goal is to create a census, in a bilingual database, of all the volumes in Italian libraries, from the origins of the Jewish press up to 1960, using the NLI catalog for the Hebrew transcription and the Worldcat website of the Library of Congress for transliteration into Latin characters. In addition to the bibliographic data, there are also those of the items (marginalia, censorship, notes of ownership, valuable bindings) and the location. The volumes are also digitized in the main and most significant pages. All information is located in the digital repository of the National Central Library of Rome, suitably adapted. The entire project is expected to include a total of 35,000 volumes from 14 Jewish communities and 25 state institutions.

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2020

La tecnologia 5G e i beni culturali: percorsi di storie e architetture all'Aquila

«DigItalia» 2-2020
DOI: 10.36181/digitalia-00019

Silvia Mantini - Fabio Graziosi - Fabio Franchi - Stefano Boero¹
Università degli Studi dell'Aquila

All'indomani del terremoto del 2009, l'Università dell'Aquila ha realizzato progetti di tutela e valorizzazione dei beni culturali che rispondono a esigenze di comunicazione dell'invisibile, di fruizione del visibile differentemente collocato e del recupero di facies sparite.

Con particolare riferimento al patrimonio storico-artistico della città, la tecnologia 5G ha permesso la sperimentazione di soluzioni di realtà aumentata e virtuale che hanno riguardato chiese e palazzi ricostruiti.

La traduzione delle ricerche d'archivio in approcci storici digitali, in pratiche di public history e in esperienze di editoria aumentata ha consentito, attraverso le ICT, una più ampia accessibilità ai contenuti multimediali da parte di diversi pubblici di fruitori.

Narrazioni digitali di Storia attraverso i beni culturali

Le nuove tecnologie dell'informazione e della comunicazione (ICT) hanno consentito, in particolare in questo ultimo decennio, lo sviluppo di molteplici e ricche modalità di trasmissione del patrimonio culturale di realtà urbane e aree di valore storico-artistico, con metodologie più diffuse per i saperi e le coscienze di territori meno noti². Queste pratiche hanno creato percorsi partecipati di conoscenza della storia e dell'identità dell'Italia e della sua immensa eredità culturale. In questo contesto l'Università degli Studi dell'Aquila, anche con riferimento alle applicazioni della futura tecnologia 5G in corso sul territorio, sta realizzando una serie di servizi per la tutela, la valorizzazione e la fruizione del patrimonio materiale e immateriale (storico, architettonico, artistico e naturalistico) finalizzati alla promozione di prodotti culturali e di politiche innovative integrate, di supporto al turismo.

¹ Silvia Mantini è autrice del paragrafo "Narrazioni digitali di Storia attraverso i beni culturali"; Fabio Graziosi e Fabio Franchi sono autori del paragrafo "Applicazioni 5G per la valorizzazione dei beni culturali: un caso di studio per la città dell'Aquila"; Stefano Boero è autore del paragrafo "Totem multimediali e comunicazione della Storia".

² Cfr. Elisa Bonacini, *Nuove tecnologie per la fruizione e valorizzazione del patrimonio culturale*, Roma: Aracne, 2011; Maria Teresa Natale – Michela Grimaldi, *Quando le app sono al servizio del patrimonio culturale*, «Digitalia», 11 (2016), n. 1-2, p. 70-90.

In queste riflessioni, la sperimentazione che si presenta riguarda la valorizzazione di beni culturali, attualmente non visibili nelle loro sedi per il sisma o per restauri, attraverso esperienze di visite immersive di siti di particolare interesse storico/artistico, con visori di realtà aumentata e virtuale. Sono modelli applicabili alla comunicazione dei beni culturali non solo nelle realtà post-sisma, ma anche di degrado. Queste sperimentazioni sono agli esordi e presentano anche primi prodotti di editoria aumentata, fruibile da remoto, come ad esempio modellazioni virtuali di progetti seicenteschi mai realizzati.



All'Aquila, dopo il sisma del 2009, si è costituita in Ateneo un'esperienza di comunicazione e valorizzazione dei beni culturali all'interno del progetto INCIPICT (Innovating City Planning Thought Information and Communication Technologies), condotto da docenti e da ricercatori dell'Università, in particolare ingegneri delle telecomunicazioni che si occupano di 5G, ma anche da architetti e da storici³.

Dopo il terremoto del 2009, il centro della città è stato completamente chiuso e ancora oggi sono pochi gli edifici pubblici e le chiese tornate ad avere le originali sedi parrocchiali e, comunque, riaperte al culto. Partendo dall'esigenza di voler comunicare ciò che per anni non è stato visibile, ma anche ciò che, con queste modalità digitali, può essere "visto", il progetto INCIPICT si è dedicato ad alcuni aspetti di scoperta e di condivisione di beni culturali progettati e mai realizzati, di architetture riemerse dopo il sisma con stratificazioni anche precedenti, di storie che abbinavano le scoperte derivanti dai manoscritti d'archivio con le ri-emergenze prima sconosciute.

Il centro storico oggi sta rinascendo, ma ancora molto c'è da fare per il patrimonio artistico, architettonico e documentario, depositario di memorie secolari che rappresentano tracce di identificazione per la comunità e per le nuove generazioni che da dieci anni, ad esempio, non sono mai rientrate nel duomo e nella maggior parte delle chiese.

Cosa fare quando luoghi culturali (teatro, auditorium, biblioteche) e di culto (duomo, parrocchie, monasteri) artisticamente e storicamente di grande pregio restano chiusi alla fruizione della comunità? Trovare soluzioni non è un tema che riguarda solo le realtà terremotate o colpite da catastrofi e guerre, ma può e deve costituire un tema legato a degradi e a mancanze di restauri, che confinano spesso il patrimonio culturale ad un oblio che ne disperde la memoria e la coscienza⁴.

All'Aquila, dove la grande "fame di storia" che proviene dalla comunità ha attivato molte proposte di comunicazione e di pratiche di *Public History*, il progetto INCIPICT, mettendo a disposizione un anello ottico sperimentale che connette i principali punti culturali della città, sta consentendo ad architetti, storici e ingegneri di "riportare" la Storia in palazzi, nuovi e quasi senza segno del tempo, in chiese che mostrano le stratificazioni di fasi artistiche e ricostruttive, che restituiscono tracce che nessuno conosceva neanche prima del 2009, in musei divenuti sedi anche di postazioni di visioni da remoto⁵.

³ Fabio Graziosi, *Nuove tecnologie per la città smart*, in: *L'Aquila, la città e il nuovo millennio*, a cura di R. Aloisio [et al.], Pescara: Carsa, 2018, p. 77-84.

⁴ Silvia Mantini, *Coltivare memorie: la Storia dopo il terremoto*, in *Individui, comunità e istituzioni in emergenza*, a cura di S. Marantoni – A. Vaccarelli, Milano: FrancoAngeli, p. 249-259.

⁵ Silvia Mantini, *Experiences of Public History and ICT for the representation of cultural heritage*, «Disegnarecon», 12 (2019), n. 23, p. 161-164; Serge Noiret, *Making Public History*, in: *Italy in Making Histories*, ed. by Paul Ashton – Tanya Evans – Paula Hamilton, Berlin: Walter de Gruyter GmbH, 2020; Silvia Mantini, *L'Aquila e la riscoperta del Barocco con realtà aumentata e 5G*, in *Summer school «Ripensare il barocco (secoli XVII e XVIII)»: nuove prospettive storico-critiche*, Torino: Fondazione 1563 per l'Arte e la Cultura della Compagnia di San Paolo, 2019, p. 15-18.

Quattro sono i punti alla base degli interventi delle tecnologie della comunicazione e dell'informazione applicate ai beni culturali che emergono:

- 1) la comunicazione dell'invisibile, ad esempio, di ciò che per sisma o degradi è chiuso e non più fruibile o che è riemerso (come il caso della chiesa di Santa Maria Paganica e le chiese del territorio, a lungo chiuse, come Santa Maria ad Cryptas a Fossa);
- 2) la fruizione del visibile diversamente collocato, come i dipinti e le sculture nei luoghi di conservazione, che potrebbero essere, con il 5G, "ricollocati" nelle loro sedi e osservati nel luogo che li ha concepiti, come la *Visitazione* di Raffaello Sanzio, prima nella chiesa di S. Silvestro dell'Aquila e oggi esposta al Prado a Madrid;
- 3) il recupero di *facies* sparite per danni o per scelte storiografiche di restauri progressi, come l'area del San Basilio e del San Salvatore, ma anche edifici di culto soggetti alla negazione del Barocco, che, con una azione di *damnatio memoriae* condotta dalla Soprintendenza alla fine degli anni Sessanta del Novecento, furono riportate ad architetture dichiarate più coerenti (come nel caso della Basilica di S. Maria di Collemaggio sventrata dal Barocco in "coerenza" allo stile medievale della facciata).
- 4) la proposta di comunicazione di ricostruzioni storiche, con ricerche d'archivio, attraverso la realtà aumentata per l'editoria. Questo è il caso del Palazzo Camponeschi, sede per circa tre secoli dei Gesuiti, che sta consentendo la comunicazione della storia del Palazzo da sede di governo, a quella dei Padri Ignaziani e del Collegio, fino a Palazzo privato e pubblico, valorizzando tutte le fasi con i ritrovamenti derivati dai restauri post-sisma e da ricerche d'archivio.

Gli obiettivi di queste nuove pratiche di uso delle ICT per la comunicazione della storia, attraverso i modelli virtuali e le riproposizioni di contenuti culturali fruibili con le visualizzazioni delle aree urbane e delle architetture riemerse, sono quelli di:

- recuperare la conoscenza di luoghi pubblici e religiosi, negati da catastrofi o degradi, per la ricostruzione di percorsi storici che hanno caratterizzato la storia della città e del territorio;
- mostrare le evoluzioni degli edifici e dei loro usi modificati in seguito a momenti storici e normativi;
- ricostruire la rete culturale e sociale nei/dei luoghi del patrimonio pubblico ed ecclesiastico attraverso i secoli, con strumenti che consentano la conoscenza dei siti, anche archivistici, per la realizzazione di percorsi storici, religiosi, didattici e turistici volti a una narrazione collettiva che recuperi l'identità e la cultura nel tempo.

È un campo molto complesso, quello tra uso delle ICT e comunicazione di saperi per la collettività, ma è un'opportunità importante per dare voce a beni culturali spesso rimasti in oblio per catastrofi o degradi, che devono tornare a darci nuove narrazioni.



Applicazioni 5G per la valorizzazione dei beni culturali: un caso di studio per la città dell'Aquila

Il terremoto del 6 aprile 2009 ha colpito duramente la città dell'Aquila e un numero rilevante di centri minori dell'Abruzzo interno, provocando 309 vittime e danni gravi al patrimonio urbano e alle attività economiche e sociali. Altri terremoti hanno colpito più recentemente l'Italia centrale, causando perdite molto pesanti di vite umane e notevoli danni.

Il processo di ricostruzione, faticosamente iniziato, richiede ancora tempi lunghi e risorse molto ingenti, sia per il recupero e la trasformazione dei centri urbani, sia per il rilancio dello sviluppo, in un contesto reso più difficile dalla crisi economica degli ultimi anni. Al tempo stesso, i cantieri della ricostruzione materiale e immateriale del "cratere sismico" costituiscono un'opportunità straordinaria di valorizzazione delle energie imprenditoriali, culturali e sociali presenti nella comunità locale, o attratte dall'esterno.

Con questa premessa, la città dell'Aquila è stata scelta⁶ come area per la realizzazione di sperimentazioni pre-commerciali 5G nella porzione di spettro 3.6-3.8 GHz. I progetti innovativi e le attività in corso all'Aquila offrono alla sperimentazione un contesto particolarmente favorevole nel quale sarà possibile disporre di servizi che vanno dal monitoraggio ambientale e delle infrastrutture ai trasporti e alla valorizzazione dei beni culturali. Proprio con riferimento ai beni culturali la nuova tecnologia 5G vede nei servizi di realtà aumentata e virtuale la cosiddetta *killer application*. La sperimentazione di soluzioni di realtà virtuale ed aumentata applicata ai beni

⁶ Gabriella Coluccelli et al., *5G Italian MISE Trial: Synergies Among Different Actors to Create a "5G Road"*, in: *2018 IEEE 4th International Forum on Research and Technology for Society and Industry (RTSI)*. IEEE, 2018.



culturali permette di offrire nuove ed arricchite esperienze di visita di musei, zone archeologiche, edifici storici e fruizione di contenuti multimediali in grado di rendere estremamente coinvolgente e ricca di contenuti l'interazione con la cultura⁷.

Con particolare riferimento alla città dell'Aquila e dei centri storici del proprio comprensorio, queste soluzioni innovative rappresentano lo strumento per connettere la realtà con una dimensione virtuale in modalità contestuale ed interattiva. Un'immersione a 360 gradi nelle opere e nella storia con la possibilità di "viaggiare nel tempo", per raccontare i luoghi come erano in passato, mostrando le trasformazioni subite nel tempo

con un'esperienza sensoriale unica, e come saranno in futuro soprattutto in relazione alle opere di ricostruzione del patrimonio artistico abruzzese gravemente danneggiato dagli eventi sismici recenti.

La valorizzazione dei beni culturali con le tecnologie di realtà virtuale ed aumentata ha l'immediato vantaggio di attrarre e avvicinare i cittadini e i turisti al patrimonio culturale ed artistico nazionale, quindi con ritorni in termini sociali, culturali ed economici. Queste nuove tecnologie permetteranno ai musei di offrire nuovi servizi, di essere utilizzate come efficace veicolo di comunicazione, di incrementare le visite con ritorno immediato e diretto per il contesto dei beni culturali e per tutto l'indotto del turismo.

La sperimentazione fa riferimento ad una serie di casi d'uso tra i quali le applicazioni di realtà aumentata e virtuale risultano particolarmente competitive⁸. Queste applicazioni offrono la possibilità di dimostrare le capacità della futura rete 5G di supportare il trasferimento di un'enorme quantità di dati in tempi estremamente

⁷ Carlo Battini, *Realtà virtuale, aumentata e immersiva per la rappresentazione del costruito*, Firenze: Altralinea, 2017; Tommaso Empler, *APP design con uso della realtà aumentata per la divulgazione dei beni culturali*, «Disegnare idee, immagini», L (2015), p. 2-11; Elisa Bonacini, *La realtà aumentata e le app culturali in Italia: storie da un matrimonio in mobilità*, «Il Capitale culturale», IX (2014), p. 89-121.

⁸ ETRI, "ITU-R IMT 2020 requirements", COM (2016) 588.

ridotti. Un primo caso di studio individuato per la sperimentazione in corso a L'Aquila fa riferimento alla chiesa di Santa Maria ad Cryptas, nel comune di Fossa, che ospitava la Madonna realizzata da Gentile da Rocca, raro e finora unico esempio di tabernacolo duecentesco abruzzese datato 1283, oggi esposto al Museo Nazionale d'Abruzzo.

Attraverso l'utilizzo di visori di realtà aumentata, l'obiettivo è quello di proporre la ricostruzione tridimensionale ad altissima risoluzione della chiesa di Santa Maria ad Cryptas a complemento dell'opera originale esposta nel museo. Inoltre, la stessa tecnologia potrà essere in futuro utilizzata per proporre una ricollocazione virtuale dell'opera stessa presso la chiesa lì dov'era originariamente collocata.

Totem multimediali e comunicazione della Storia

All'indomani del terremoto del 2009, il progetto INCIPICT si è proposto di valorizzare i beni culturali dell'Aquila attraverso le ICT, avviando sperimentazioni in alcuni luoghi-simbolo della città⁹. Nell'ambito di queste iniziative è stato avviato un "progetto-pilota" riguardante il Palazzo Camponeschi. Attuale rettorato dell'Università, il palazzo è stato storicamente sede dell'*Aquilanum Collegium* della Compagnia di Gesù, ma anche dimora di alcune delle principali famiglie aristocratiche dell'Aquila. L'intento è di restituire memorie sparite in seguito al sisma, tenendo conto delle stratificazioni riemerse nell'edificio dopo i lavori di restauro.

Sono stati realizzati dei totem multimediali con display touch-screen che hanno consentito esperienze di fruizione del patrimonio storico-culturale di Palazzo Camponeschi mediante percorsi di visita mirati. L'obiettivo è, in questo caso, accrescere il coinvolgimento degli utenti che possono distinguersi per diversi profili di interessi. All'Aquila, dopo il terremoto, si è registrata una forte richiesta di storia della città; l'utilizzo del digitale contribuisce a recepire istanze di nuove comunità di fruitori dei contenuti culturali. L'installazione e l'inserimento dei materiali multimediali all'interno dei dispositivi è il risultato di un impegno sinergico, portato avanti da un gruppo di lavoro composto da storici, ingegneri informatici, ingegneri delle telecomunicazioni e architetti.

In seguito alla ricerca archivistica e bibliografica sono riemerse le tappe della storia dell'edificio tra i secoli XV e XXI. La ricostruzione della vita del palazzo e dell'area circostante è stata possibile attraverso una rigorosa ricerca che ha condotto al rinvenimento di un'ampia documentazione manoscritta presso l'Archivum Historicum Societatis Iesu, l'Archivio Storico della Provincia Euromediterranea, l'Archivio Segreto Vaticano, l'Archivio di Stato dell'Aquila e la Biblioteca "Salvatore Tommasi".

⁹ Cristian Antonelli – Dajana Cassioli – Fabio Franchi – Fabio Graziosi – Andrea Marotta – Marco Pratesi – Claudia Rinaldi – Fortunato Santucci, *The city of L'Aquila as a living lab: the Incipict project and the 5g trial*, in: *IEEE 5G World Forum (5GWF) 2018*, p. 410-415.

I totem narrano un percorso virtuale, *Palazzo Camponeschi. Famiglie e trasformazioni tra i secoli XV e XXI*; i testi, a cura di Silvia Mantini e Stefano Boero, ripercorrono le tappe salienti della storia dell'edificio. La navigazione si svolge a partire da una schermata iniziale con cui prende avvio il tour virtuale. Il menu di avvio mostra una linea del tempo che propone cinque sezioni in ordine cronologico: le origini del palazzo (sec. XV-XVI), il primo ingresso della Compagnia di Gesù (1596-1767), l'arrivo di nuovi attori pubblici e privati (1767-1926), il secondo ingresso dei padri ignaziani (1926-1979) e, infine, la storia dell'Università dell'Aquila (1979-2020).

A ciascuna sezione corrispondono schede con contenuti testuali, iconografici e multimediali, corredati da didascalie.

L'obiettivo è di fare di Palazzo Camponeschi un luogo di comunicazione culturale e scientifica, attraverso linguaggi che consentano l'avvicinamento di diversi tipi di pubblico e una più ampia accessibilità agli esiti delle ricerche.

Per la realizzazione di queste installazioni interattive è stato acquisito un repertorio di cartoline, provenienti da collezioni private, dei decenni tra Otto e Novecento, anni in cui il palazzo fu sede dell'amministrazione postale. Numerose foto scattate durante il ventennio fascista e nell'immediato dopoguerra sono state rinvenute attraverso ricerche condotte presso l'Archivio della Provincia Euromediterranea a Roma. Proprio nell'ex-Archivio della Provincia Napoletana è confluito il complesso documentario del Collegio aquilano all'indomani dell'abbandono della città da parte dei gesuiti. Si tratta di un materiale inedito, che permette di fare luce sulle trasformazioni spaziali avvenute nel palazzo e sul ruolo che ha avuto da un punto di vista culturale a metà Novecento. Nell'Archivio Storico dell'Università dell'Aquila sono stati individuati filmati del XX



secolo e dell'inizio del XIX, contenenti riprese di eventi significativi nel Palazzo Camponeschi. Importante è stata l'acquisizione del materiale fotografico messo a disposizione dai restauratori intervenuti dopo il sisma.

Dall'esperienza dei totem di Palazzo Camponeschi nasce così un prodotto in cui le informazioni provenienti dall'indagine storica sono trasformate in ricerca comunicata attraverso le ICT. L'allestimento rientra in un programma di promozione e valorizzazione dell'offerta turistico-culturale, oltre che di implementazione delle strategie di comunicazione di siti al pubblico¹⁰.

In the aftermath of the 2009 earthquake, the University of L'Aquila carried out projects for the protection and enhancement of cultural heritage, which responded to the need for communication of the invisible, for the use of the differently located visible and for the recovery of vanished facies.

With particular reference to the historical-artistic heritage of the city, the 5G technology has allowed the experimentation of augmented and virtual reality solutions involving churches and reconstructed buildings.

The translation of archival research into digital historical approaches, public history practices as well as augmented publishing experiences has allowed, through ICT, a wider accessibility to multimedia contents by different audiences.

¹⁰ Stefano Boero, *Editoria aumentata e Storia: il Palazzo Camponeschi all'Aquila*, in *Summer school «Ripensare il barocco (secoli XVII e XVIII)»*, cit., p. 21-22.

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2020

EAGLE: storia di un'idea dalle origini all'ingresso di EDF*

«Digitalia» 2-2020
DOI: 10.36181/digitalia-00020

Maria Letizia Caldelli – Silvia Orlandi

Sapienza Università di Roma

Il contributo fornisce una breve storia di EAGLE: nasce nel 2003 come Electronic Archive of Greek and Latin Epigraphy, una federazione di banche dati che si riconoscono nello stesso modo di concepire l'epigrafia digitale; si evolve tra il 2013 e il 2016 come aggregatore e content provider di Europeaana (Europeana network of Ancient Greek and Latin Epigraphy); dal 2020 accoglie EDF (Epigraphic Database Falsae), la prima banca dati dedicata al fenomeno della falsificazione epigrafica.

EAGLE: dalle origini al 2020

Da quando epigrafia e informatica si sono incontrate, nella seconda metà del secolo scorso, molti progetti sono stati ideati e realizzati per archiviare, ordinare e consultare testi e immagini relativi alle iscrizioni latine e greche, prima sotto forma di archivi consultabili solo sul posto o su supporti mobili come CD-ROM, poi accessibili attraverso il world wide web. La molteplicità di iniziative ha generato un gran numero di potenziali strumenti di lavoro, non tutti, però, in grado di dialogare tra loro e non tutti accomunati dallo stesso destino: alcuni, dopo un primo annuncio, non sono di fatto mai entrati in funzione, altri hanno avuto vita breve, altri ancora non hanno saputo o potuto adeguarsi al vertiginoso ritmo di evoluzione della tecnologia e fanno ormai parte della storia dell'informatica umanistica. Una situazione del genere rischiava di recare più danno che beneficio alla disciplina per il progresso della quale tali iniziative erano state concepite. È per questo che, in occasione dell'XI "Congresso Internazionale di Epigrafia Greca e Latina", tenutosi a Roma nel 1997, fu fondata, in seno all'Association Internationale d'Épigraphie Grecque et Latine (AIEGL), una commissione di "Epigrafia e Informatica", alla testa della quale fu posto Silvio Panciera. L'obiettivo era quello di coordinare le numerose iniziative esistenti in questo campo, in modo da indirizzarle verso il raggiungimento di un fine comune. Dopo un primo censimento dei progetti avviati e in corso d'opera, i rappresentanti delle principali istituzioni coinvolte in queste iniziative furono invitati a riunirsi a Roma per fare il pun-

* Il capitolo *EAGLE: dalle origini al 2020* è stato scritto da Silvia Orlandi, il capitolo *L'ingresso in EAGLE di EDF* è stato scritto da Maria Letizia Caldelli.

to della situazione e dare vita ad un'impresa comune, che teoricamente avrebbe dovuto consistere nella creazione di un'unica banca dati epigrafica destinata a raccogliere non solo i testi delle iscrizioni, ma anche altre informazioni essenziali per lo studio e la completa comprensione di un documento epigrafico: il contesto di rinvenimento, il luogo di conservazione, la tipologia e il materiale del supporto, la tecnica scrittoria e così via. Alla luce dei fatti, l'impresa, in questa forma, si rivelò irrealizzabile per motivi sia tecnici che diplomatici e, ferma restando l'esigenza di creare uno strumento di lavoro con queste caratteristiche, si decise allora, nel corso di un incontro tenutosi a Trieste e Aquileia nel 2003, di dare vita ad una federazione di banche dati epigrafiche che si riconoscessero in uno stesso modo di concepire l'epigrafe – non un semplice testo, ma un monumento iscritto inserito in un contesto storico e topografico – e, di conseguenza, la forma della sua digitalizzazione¹. Nacque così EAGLE (Fig. 1), ovvero l'Electronic Archive of Greek and Latin Epigraphy, cui aderirono in un primo momento 3 banche dati: Epigraphic Database Bari (EDB)², destinata ad occuparsi delle iscrizioni cristiane di Roma; Epigraphic Database Heidelberg (EDH)³ che, nata come schedatura informatizzata dei volumi dell'*Année Épigraphique*, ha concentrato, da allora, la propria sfera di azione sulle province europee dell'Impero Romano; Epigraphic Database Roma (EDR)⁴, che si propone la digitalizzazione delle iscrizioni latine e greche di Roma (tranne le cristiane) e dell'Italia antica, comprese le province di Sicilia e Sardinia.



Figura 1. Logo della Federazione EAGLE (Electronic Archive of Greek and Latin Epigraphy)

Nel 2009, ai "membri fondatori" della Federazione è andata ad aggiungersi una nuova banca dati, Hispania Epigraphica Online (HEpO)⁵, che si occupa delle iscrizioni della penisola iberica. Scopo della federazione è essenzialmente quello di darsi dei criteri comuni, nell'ottica di una costante collaborazione internazionale, utili a coordinare l'opera di digitalizzazione del materiale epigrafico che ogni banca dati continua a perseguire autonomamente⁶. Tali criteri corrispondono non tanto a standard descrittivi internazionali,

¹ Su queste vicende si veda: Silvio Panciera, *EAGLE: cronistoria di un problema e di un progetto*, in *Epigrafi, epigrafia, epigrafisti. Scritti vari editi e inediti con note complementari e indici*, Roma: Edizioni Quasar, 2006, p. 1913-1917.

² <http://www.edb.uniba.it/>.

³ <https://edh-www.adw.uni-heidelberg.de/home>.

⁴ www.edr-edr.it.

⁵ <http://eda-bea.es/>.

⁶ www.eagle-eagle.it.

quanto all'esigenza di strutturare i dati in modo da renderli almeno in parte consultabili anche attraverso un unico portale. Di qui, ad esempio, la scelta di impostare il luogo di provenienza separando i campi *Regio* e *Urbs* e distinguendo tra denominazioni antiche e moderne, oppure la decisione di non includere nelle banche dati aderenti alla federazione i materiali definibili come *instrumentum* "ripetibile", costituiti da bolli impressi su prodotti di varia natura (anfere, laterizi, ceramica da tavola e da cucina, lucerne...) noti in migliaia di esemplari, ma riconducibili a un certo numero di matrici, che richiedono un trattamento informatico a sé. Attualmente il portale della Federazione dà accesso al materiale archiviato nelle banche dati EDB, EDH e EDR attraverso un numero relativamente limitato di campi – gli unici i cui dati sono tecnicamente compatibili tra loro – costituiti dal luogo di provenienza (diviso, appunto in Provincia o Regione antica, Stato moderno, Città antica, Città moderna e Sito di rinvenimento), dalla bibliografia e dal testo (in cui la ricerca sulle parole latine può essere combinata attraverso gli operatori booleani AND e OR)⁷.

L'impronta interdisciplinare che caratterizza i progetti aderenti a questa federazione, la cui realizzazione richiede interesse e attenzione per i dati topografici, storici, archeologici e filologici che accompagnano lo studio e l'edizione dei testi epigrafici, ha indotto il suo primo presidente, Silvio Panciera, a individuare nel Centro Linceo Interdisciplinare Beniamino Segre l'istituzione ideale da scegliere come sede ufficiale di EAGLE. E il Centro Segre ha effettivamente supportato per anni la Federazione sia attraverso una serie di borse di studio e di scambi culturali che hanno favorito il processo di digitalizzazione dei materiali, sia mettendo a disposizione la sede dell'Accademia dei Lincei per alcuni incontri, tra cui si segnala il convegno "Epigrafia, informatica e ricerca storica", tenutosi nel novembre 2008⁸. Sulla stessa pagina web⁹ si trova anche il documento relativo ad un'altra importante tappa della storia di EAGLE: l'accordo stipulato nel 2005 con la Direzione generale per l'innovazione tecnologica e la promozione dell'allora Ministero per i beni e le attività culturali che, in considerazione della rilevanza internazionale del progetto, concede alle banche dati afferenti alla Federazione la possibilità di utilizzare gratuitamente, a determinate condizioni, tutte le immagini delle iscrizioni edite di proprietà dello Stato italiano. L'accordo precede di molti anni il decreto legge Franceschini del marzo 2014, che consente la realizzazione, senza preventiva autorizzazione, di foto nei musei a scopo di studio, e, da un lato, segna un importante riconoscimento del valore scientifico e didattico che sta alla base dell'idea di

⁷ Sugli oggettivi limiti di questo portale si vedano le osservazioni di Antonio Enrico Felle, *Esperienze diverse e complementari nel trattamento digitale delle fonti epigrafiche: il caso di EAGLE ed EpiDoc*, in *Diritto romano e scienze antichistiche nell'era digitale. Convegno di studio (Firenze, 12-13 settembre 2011)*, a cura di N. Palazzolo, Torino: G. Giappichelli Editore, 2012, p. 117-124.

⁸ I documenti relativi sono disponibili sulla pagina <<http://www.edr-edr.it/Download/EAGLE-Testi.pdf>>.

⁹ <http://www.edr-edr.it/Documenti/Accordo%20MIBAC-EAGLE.pdf>.

EAGLE, dall'altro, ribadisce l'attenzione e il rispetto delle persone e degli enti che afferiscono al progetto per le norme vigenti in materia di copyright. Nello stesso spirito va letta anche la scelta di immettere nei database solo materiali epigrafici già editi in pubblicazioni precedenti, per evitare la circolazione in rete, senza la dovuta autorizzazione da parte degli enti preposti alla loro tutela, di reperti ancora in corso di studio. Una scelta che il progetto EAGLE ha operato sin dall'inizio e che porta avanti con convinzione, benché sia invalsa, in alcuni casi, la pratica di considerare lecito tutto ciò che è reso tecnicamente possibile dalla facilità con cui quotidianamente accediamo a dispositivi come fotocamere digitali, siti internet e social network.

Un autentico punto di svolta nella storia di EAGLE è stato segnato dalla possibilità di accedere per il triennio 2013-2016 a un consistente finanziamento della Commissione Europea, erogato nell'ambito del programma ICT-PSP del 7° Programma Quadro. Individuando la presenza di un "gap" nei contenuti presenti in Europeana, infatti, un consorzio di cui hanno fatto parte, oltre ai "soci fondatori" della Federazione, altri partner che disponevano di ricche collezioni epigrafiche digitali (come la British School at Rome, il progetto Last Statues of Antiquity della Oxford University, l'archivio Arachne del Deutsches Archäologisches Institut e così via) si è proposto di aggregare tali contenuti per renderli disponibili attraverso un unico portale e, al tempo stesso, conferirli ad Europeana. EAGLE è diventato, così, l'acronimo del progetto European network of Ancient Greek and Latin Epigraphy¹⁰, che, nel giro di 3 anni, ha cambiato il volto dell'epigrafia digitale in Italia e in Europa (Fig. 2). Obiettivo del progetto, infatti, era rendere accessibili un gran numero di testi e immagini relativi a iscrizioni antiche non solo agli addetti ai lavori, attraverso un comune portale di ricerca, ma anche ad un pubblico più vasto, attraverso una serie di strumenti di mediazione culturale come le traduzioni in lingue moderne, la tecnica dello storytelling, il riconoscimento attraverso le immagini. Grazie ad una imponente

opera di armonizzazione e disambiguazione dei contenuti, è stata innanzi tutto disegnata una nuova maschera di ricerca, che, nel formato "advanced search", consente l'accesso ai materiali aggregati attraverso un maggior numero di campi, espandibile secondo le esigenze dell'utente (non solo provenienza, bibliografia e testo, ma anche luogo di conservazione, tipolo-



Figura 2. Logo del progetto EAGLE (*European network of Ancient Greek and Latin Epigraphy*)

¹⁰ www.eagle-network.eu.

gia e materiale del supporto, tipologia del testo, tecnica di scrittura, stato di conservazione). Fondamentale, per ottenere questo risultato, è stata la creazione di una serie di “vocabolari controllati” per armonizzare le definizioni presenti nei campi oggetto di ricerca, vocabolari che poi sono stati messi liberamente a disposizione di tutti coloro che desiderano utilizzarli, adattandoli alle proprie esigenze, per altri progetti simili¹¹. Sono stati inoltre realizzati, sotto l’egida di EAGLE, una serie di “servizi aggiuntivi” destinati ad arricchire il panorama dell’offerta didattica nel campo dell’epigrafia digitale: una piattaforma EAGLE Mediawiki per l’inserimento delle traduzioni (o delle parafrasi) delle epigrafi, che ospita ad oggi oltre 18.000 testi in inglese, francese, tedesco e italiano¹²; una storytelling application destinata all’inserimento dei testi epigrafici e degli altri materiali ad essi correlati in uno stile di presentazione “narrativo”, con delle notevoli e ancora non pienamente esplorate potenzialità nel campo della divulgazione storica¹³. La ricerca delle iscrizioni mediante l’uso di immagini, invece, è stata resa fruibile sia attraverso un’interfaccia web¹⁴ che tramite un’applicazione per smartphone¹⁵. Per effettuare una ricerca per immagini attraverso il portale EAGLE è sufficiente caricare dal proprio computer un’immagine di un’iscrizione di proprio interesse e cliccare sul tasto “Search”: una lista di risultati verranno proposti all’utente che, mediante un semplice click, potrà accedere alla scheda contenente le informazioni relative all’oggetto selezionato. Per effettuare una ricerca tramite l’App, invece, basterà puntare la fotocamera del proprio smartphone verso un’iscrizione che si vuole riconoscere e scattare una foto; il sistema invierà la foto al server di EAGLE¹⁶, che effettuerà il riconoscimento e, in tempo reale, fornirà all’utente una scheda dettagliata dell’oggetto fotografato. La ricerca delle iscrizioni tramite l’uso di immagini è realizzata tramite sistemi di Content Based Image Retrieval (CBIR), capaci di archiviare e reperire le immagini in base al loro “contenuto visivo” e senza l’uso di metadati come il testo. L’idea che sta alla base di tali sistemi è quella di trasformare ciascuna immagine in un oggetto matematico, che viva in uno “spazio metrico”, ovvero uno spazio ove sia possibile calcolare una funzione di distanza o similarità tra gli oggetti¹⁷.

¹¹ <https://www.eagle-network.eu/resources/vocabularies/>.

¹² <https://wiki.eagle-network.eu/wiki/>.

¹³ <https://www.eagle-network.eu/stories/>.

¹⁴ <http://www.eagle-network.eu/image-search/>.

¹⁵ <https://www.eagle-network.eu/resources/flagship-mobile-app/>.

¹⁶ Attualmente conservato e amministrato dall’Istituto di Scienza e Tecnologie dell’Informazione (ISTI) del CNR di Pisa.

¹⁷ Per maggiori dettagli sul funzionamento del sistema di riconoscimento per immagini si veda: Giuseppe Amato – Fabrizio Falchi – Fausto Rabitti – Lucia Vadicamo, *Inscriptions Visual Recognition. A comparison of state-of-the-art object recognition approaches*, in *Information Technologies for Epigraphy and Cultural Heritage. Proceedings of the First EAGLE International Conference*, a cura di S. Orlandi, R. Santucci, V. Casarosa, P. M. Liuzzo, Roma: Sapienza Università Editrice, 2014, p. 117-131.

Nel complesso, dunque, un progetto molto innovativo, ma non lontano dall'idea da cui la Federazione EAGLE era nata, tanti anni prima, e cioè che per comprendere adeguatamente il patrimonio epigrafico occorre un insieme complesso di informazioni, che vanno archiviate con cura e attenzione, armonizzate e rese disponibili gratuitamente all'intera comunità, accademica e non.

Dopo la fine ufficiale del progetto europeo tutti i servizi creati nel triennio 2013-2016 (motore di ricerca, piattaforma per l'inserimento delle traduzioni ecc.) continuano a funzionare e ad accogliere i nuovi contenuti che periodicamente vengono inviati al server comune da vecchi e nuovi partner grazie all'impegno dell'associazione International Digital Epigraphy Association (IDEA)¹⁸, che si è assunta l'onere di assicurare la manutenzione e l'aggiornamento dei risultati conseguiti grazie al finanziamento europeo. Al tempo stesso, il ruolo di raccordo e di coordinamento tra i partner "storici" del progetto continua ad essere svolto dalla Federazione EAGLE che, a partire dal 2020, sta vivendo una stagione di nuova vitalità, che l'ha portata a trovare una nuova sede e un nuovo sostegno nell'Istituto Italiano per la Storia Antica, e ad accogliere tra i suoi membri una nuova banca dati, destinata ad occuparsi, per la prima volta nell'era digitale, delle iscrizioni false: EDF.

EDF e l'ingresso in EAGLE

Epigraphic Database Falsae (EDF), finanziato dal Progetto PRIN 2015 - False testimonianze. Copie, contraffazioni, manipolazioni e abusi del documento epigrafico, vede riuniti ricercatori afferenti alle maggiori università pubbliche italiane, quali l'Università Ca' Foscari di Venezia, cui appartiene il Principal Investigator (Lorenzo Calvelli), Sapienza - Università di Roma, cui appartengono le autrici di questo contributo, l'Università degli studi di Bari Aldo Moro, l'Università degli Studi di Genova, Macerata, Milano, Pisa e Trieste. Il progetto triennale (5 febbraio 2017- 5 febbraio 2020), appena concluso, ha ricevuto un finanziamento di 178,000 Euro: è previsto un convegno a Roma nell'aprile 2021 destinato a disseminare i risultati ottenuti e a tirare le fila della esperienza soprattutto nel quadro delle Digital Humanities (Fig. 3).

L'idea a monte del progetto scientifico è quella di gettare nuove basi per lo studio delle iscrizioni genericamente etichettate come *falsae* dagli editori dell'ottocentesco *Corpus Inscriptionum Latinarum* (CIL), tra le quali sono comprese, senza essere distinte, realtà molto differenti tra loro: epigrafi d'invenzione che imitano dolosamente l'antico, copie di iscrizioni di età greco-romana, iscrizioni originali di età post classica fatte sul modello dell'antico, trascrizioni o rielaborazioni di passi letterari o di monete, alterazioni di iscrizioni antiche.

Gli editori della monumentale opera, mossi dal proposito di creare una raccolta

¹⁸ Su cui si veda: Antonio Enrico Felle, *Oltre EAGLE: l'International Digital Epigraphy Association (IDEA). Una presentazione in anteprima*, «Archeologia e calcolatori», 27 (2016), p. 353-359.



Figura 3. Home page del portale EDF (Epigraphic Database Falsae)

senza precedenti per quantità e qualità di iscrizioni genuine, decisero di relegare tra le *falsae* sia tutte quelle iscrizioni che, previa autopsia, si fossero rivelate non antiche, sia quelle, note solo da tradizione manoscritta o a stampa, il cui primo editore (talora unico) fosse noto come falsario o anche solo sospettato tale. Questa operazione di epurazione, peraltro condotta senza sistematicità, ebbe come prima conseguenza quella di creare una epigrafia di serie B, come effetto collaterale quello di scoraggiare fino a tempi recenti¹⁹ sia una rimozione dell'anatema mommseniano sia un approccio al problema con nuovi strumenti di indagine.

Forte della convinzione che i falsi non siano un sottoprodotto dell'epigrafia ma che siano invece una importante testimonianza della cultura che li ha prodotti e un fondamentale veicolo per lo studio della ricezione dell'antico, il gruppo di ricerca ha posto tra le sue priorità la creazione sia di una base documentaria vasta, solida e da poter condividere, sia di uno strumento valido per poterla rendere fruibile.

Da qui è nato il progetto di ideare un database per testi ed immagini specifico per

¹⁹ Anthony Grafton, *Forgers and Critics, Creativity and Duplicity in Western Scholarship*, Princeton: Princeton University Press, 1990; Marc Mayer Olivé, *L'art de la falsificació. Falsae inscriptiones a l'epigrafia romana de Catalunya*, Barcelona, 1998; *El monumento epigráfico en contextos secundarios. Procesos de reutilización, interpretación y falsificación*, a cura di J. Carbonell Manils, H. Gimeno Pascual, J. L. Moralejo Álvarez, Bellaterra: Universitat Autònoma de Barcelona, 2011, p. 15-38; Ilaria Andreoli, *Pensare il falso: un percorso critico-bibliografico*, «Studiolo», 11 (2014), p. 16-39; Jacqueline Lichtenstein, *Penser le faux*, *Ibidem*, p. 11-15.

le iscrizioni *falsae*, che tenesse presenti gli standard degli altri database epigrafici compresi nella Federazione EAGLE e che anzi fosse in grado di dialogare con questi. Dallo sviluppo di una versione beta, ospitata dal server del DigiLab dell'Università di Roma La Sapienza, è nato così EDF, una risorsa digitale open-access, che si propone di riunire i "falsi" epigrafici prodotti in Italia dal medioevo a oggi. È permanentemente ospitato su un server fornito dall'Università Ca' Foscari di Venezia²⁰ ed è raggiungibile anche attraverso il portale condiviso EAGLE (anche se al momento non è ancora consultabile dal portale)²¹. È inoltre in dialogo con le principali risorse online esistenti per le iscrizioni genuine, in particolare EDR²², EDB²³, e Epigraphische Datenbank Claus Slaby (EDCS)²⁴. Un manuale per lo schedatore, utile però anche per il fruitore, è stato redatto da Maria Letizia Caldelli e Carlo Slavich e sarà a breve anch'esso consultabile online.

Il progetto risponde allo spirito delle politiche inclusive promosse dall'UE e definite nel programma Horizon 2020 - Inclusive Societies che, tra gli altri obiettivi, mira a sostenere i meccanismi per promuovere una crescita intelligente, sostenibile e inclusiva delle società: ciò si esprime sia attraverso la possibilità di accesso al database gratuito e ad un pubblico vasto, non limitato ai soli addetti ai lavori, sia stimolando la riflessione sul fenomeno della falsificazione nelle sue diverse declinazioni (contraffazioni, fake news, alterazioni e simili).

La struttura del database permette di fare sia ricerche semplici che ricerche avanzate.

Figura 4. EDF (Epigraphic Database Falsae): ricerche semplici

²⁰ www.edf.unive.it.

²¹ www.eagle-eagle.it.

²² www.edr-edr.it.

²³ www.edb.uniba.it.

²⁴ www.manfredclaus.de.

Figura 5. EDF (Epigraphic Database Falsae): ricerca avanzata

Nel primo caso il materiale archiviato è accessibile attraverso un numero di campi limitato (Fig. 4), quali la città antica attribuita (dal falsario: dobbiamo ricordare che si tratta di iscrizioni non genuine), le parole nel testo (“Trascrizione interpretativa”), la bibliografia, oltre ovviamente il numero identificativo (ID).

Nel caso delle “Ricerche avanzate” il numero dei campi si moltiplica ed è possibile effettuare ricerche (Fig. 5) che consentono di isolare o combinare insieme sia campi già previsti per le iscrizioni genuine ma mai utilizzati per le iscrizioni “false”, quali quelli relativi al materiale, al tipo di supporto, alla classe epigrafica attribuita, sia campi appositamente ideati, quali ad esempio:

- 1) la tipologia, distinguendo tra copie – integrali, parziali o interpolate – di iscrizioni antiche (o di iscrizioni false) e testi d’invenzione; nel caso delle iscrizioni cristiane, riservando uno spazio anche alle iscrizioni che sono il risultato di un’alterazione degli originali o di una imitazione; considerando cosa a sé le iscrizioni moderne; prevedendo la possibilità di iscrizioni riabilitate.
- 2) l’epoca della prima attestazione, che è in molti casi quella di realizzazione;
- 3) la località di produzione;
- 4) il “falsario”, laddove sia possibile identificare un autore riconosciuto per certo o solo sospettato;
- 5) le vicende collezionistiche (“luogo di conservazione”);
- 6) la categoria sociale degli individui presenti nel testo.

I punti 3 e 5 sono stati pensati allo scopo di tentare di localizzare le “officine” della falsificazione epigrafica e i canali di distribuzione dei falsi, con la possibilità di individuare le tappe intermedie e finali del percorso collezionistico. I punti 1 e 6,

Scheda: EDF000955 (Falso materiale)

Completare questa scheda: Epigraphic Database Falsae, scheda EDF000955, c. stavitch (02-11-2016)

Regione antica attribuita:	Roma	Paese moderno:	Italia
Città antica attribuita:	Roma	Città moderna attribuita:	Roma
Epoca prima attestazione:	dal 1544 al 1547		
Luogo di conservazione:	• Casacoli (Poli) e Cesana, collezione Di Bagno del 1544 al 1570 (Dezidi 1973) • Palazzo Tomasin (Rimini), collezione Di Bagno del 1970		
Materiale:	metallo	Supporto:	metallo
Misure (A x L x P) (cm):	15.6 x 23.4 x 2.8	Altezza lettere:	1,7-1,7
Stato di conservazione:	iscrizione integra		
Scrittura:	incisa a scabello (scritture a scabello)		
Lingua:	latino		
Tipologia attribuita:	accusatorie		
Categoria sociale attribuita:	tutti i casi di persone che non rientrano in una delle categorie sopra indicate		
Trascrizione diplomatica:	ILABERIVS ELPHLEPROS LABERIA L LI VCIVNDA D CASCLIVS D J L PRIMVS		
Trascrizione interpretativa:	Iliberia (Luchina) (Laberia) Epiphlepro (Luc) (Liberta) (Philepro). Labera (Luc) (Iberia) Iuranda D(ecimus) Casclius D(ecimus) (et) (militaria) I(beria) Primus		
Tipologia:	copia integrale di iscrizione parvula		
Modello:	materiale (EDR)		
Bibliografia moderna:	• Corpus Inscriptionum Latinarum, Berlin 1902-06, 2597f • G. Dezidi, Le collezioni Di Bagno: le iscrizioni greche e latine, Faenza 1976, nr. 150, col. 100		



Creata da Carlo Slavich il 24-06-2019
 Supervisionata da Maria Letizia Colabelli il 18-08-2019
 Approvata da Letizia Colabelli il 10-11-2019
 Ultima modifica di Carlo Slavich del 11-07-2019



Finanziato dal Progetto PRIN 2015 - Falso Testimonianze, Copie, contraffazioni, manipolazioni e abusi del documento epigrafico

Figura 6. EDF (Epigraphic Database Falsae): la scheda

eventualmente incrociati tra loro e/o con voci relative ad indicazioni geografiche e/o cronologiche e/o al materiale, al tipo di supporto, alla classe epigrafica attribuita, possono permettere di individuare linee di tendenza nel gusto delle diverse epoche e a seconda dei luoghi (Fig. 6).

INDEX PRINCIPIORVM **FALSAE**

INDEX PRINCIPIORVM.

Formulo d. m. d. m. s., memoria, Mo facit, domino nostro similesque vulgares praetermissae sunt; vocabula imperfecta restitutionis incertae, a quibus tituli incipiunt, item praenomina si qua inscriptiones ordiuntur, aliaque eiusdem generis adscripta quidem sunt neque tamen relata ad ordinem litterarum; vocabula significata per compendium certo et facile solvendum intelleguntur tanquam plene scripta essent. Titulorum vere antiquorum, qui perperam in hoc volumine recepti sunt, numeros uncis inclusimus, addito aut nota gen., aut ubi fieri potuit numero, quo insigniti sunt in serie genitorum.

l. ababascantius l. f. 980*. abasianor. famil. m. abasianus 979*. abbia c. l. sasinus 981*. abtinneus aug. caes. lib. 982*. l. abortennius 983*. abupola ti. claudi aug. liberta 483*. l. acato aed. speculatori 3501*. scipias donagi 3809*. m. accio aurtico 985*. m. accio m. f. ouf. primigenio 3374*. c. accio c. f. quir. priaco 986*. m. accio m. f. caq. victorino 987*. acciae paulinae 989*. l. accolei m. f. pal. primigenio 989*. c. accurtius c. l. iulianus 990*. m. accurtius m. l. plocamus 991*. c. schrysius aug. ser. 992*. scidinius faustae 993*. scilianae t. f. uenustae 994*. t. seli arelii uenusti 1025*. l. selius l. f. pal. bolonius 1026*. m. selio commodo 2290*. p. selius continthus 3210*. f. selius aug. lib. cynipa 1027*. t. selio t. f. daimacho 1028*. p. selio epaphroditio 1029*. l. selio l. l. spicichido 1030*. t. selius t. l. halenius 1031*. t. selius t. f. pop. heliatius 1032*. d. selius d. f. pal. ianuaris 1033*. l. selio l. f. latino 1036*. l. selio l. l. muschillo 1036*. l. selius l. lib. nouethus 1036*. l. selio l. f. primigenio 1037*. t. selio t. f. palatina proculiano 1038*. selio prozagone 1039*. selii tamyr 3164*. d. selius d. f. agrn. trebonius 1041*. t. aterio ualentio 1077*. aeterinis sac. soli et lunae 106*. q. aetherius q. f. fuscus 1078*. c. alfeius c. l. berillus 1079*. q. afranio q. f. iuperco 1080*. m. afranius m. l. promaceus 1081*. agnatus casaria ang. l. 828*. a. gasticus 3319*. agatobulus f. l. pyrrus 3484*. agilia cleuberis [1083 = 10518]. agrippinali germanici caesaris filii 712* b. m. agris m. f. quinzia agrillanus 1084*. m. agrus m. f. quir. salenus 1085*. agrine tryphosae 983* b. a. amilius sex. f. trom. caluus 1055*. m. amilio m. lib. eubaec 1087*. l. amilius l. f. k. nepos fabianus 1061*. m'. amillius m'. f. tromentina maxumus 1060*.

Figura 7. Corpus Inscriptionum Latinarum (CIL), vol. VI, fasc. V (falsae): indice per incipit

Se si considera che fino a questo momento nella migliore delle ipotesi era possibile fare solo ricerche sul testo e per *incipit* (Fig. 7) si tratta davvero di un radicale avanzamento. Un problema è ovviamente quello della implementazione del database. Al momento sono disponibili online circa 2.000 iscrizioni, ma presto altre, i cui contenuti sono stati già preparati, saranno riversate in EDF.

In linea con la politica di EAGLE, anche in questo caso si è scelto di inserire nel database solo materiali già editi, sebbene nel caso dei “falsi” sia materiali sia cartacei non sempre si sia potuto lavorare con edizioni aggiornate e in più di un caso per edizione si sia dovuto intendere la sola trascrizione del testo.

La ricerca del materiale non è stata e non è compito semplice. La asistematicità con cui le iscrizioni *falsae* sono state trattate dagli editori del CIL ha costretto a ricercare queste non solo nelle apposite sezioni del *Corpus*, ma anche tra le genuine, costringendo ad una nuova schedatura dell’intera raccolta. A ciò si aggiunga il fatto che assai spesso le iscrizioni *falsae* presenti nelle fonti manoscritte e a stampa, fonti pur schedate dal CIL, sono state intenzionalmente tralasciate, rendendo necessaria una nuova schedatura di queste. Ci sono poi le iscrizioni *falsae* rese note dopo il CIL, pubblicate nelle sedi più disparate e ancora una volta senza alcuna organicità.

Nel caso dei “falsi” materiali, dopo una fase di censimento tramite la bibliografia, si è cercato di rintracciare le iscrizioni, procedendo per lo più per luoghi di conservazione e confidando nella collaborazione delle Soprintendenze archeologia, belle arti e paesaggio e degli allora Poli museali: non sempre la ricerca ha avuto successo.

Ma come si è detto EDF è un database per testi ed immagini. Queste ultime rive-



Figura 8. EDF (*Epigraphic Database Falsae*): le immagini

stono una particolare importanza non solo perché permettono di apprezzare il falso in tutte le sue componenti (come per le iscrizioni genuine un falso epigrafico non è solo testo), ma anche perché consentono di conoscere o riconoscere le mani dei falsari e per questa via di accostare tra loro prodotti che talora le vie del collezionismo hanno allontanato (Fig. 8).

Per l'inserimento delle immagini digitali nella banca dati è stato fondamentale poter usufruire della convenzione con la Direzione Generale del Ministero dei Beni e delle Attività Culturali e del Turismo per il loro libero uso, a scopi didattici e scientifici.

In assenza di specifica convenzione, tuttavia, attraverso il campo "URL immagine", è stato possibile agganciare foto che si trovano in fototeche esterne: questo è al momento possibile grazie a specifici accordi con EDH – Photographic Database²⁵, Corpus Inscriptionum Latinarum²⁶, Ubi erat Lupa²⁷, The British School at Rome Photographic Archive²⁸, Epicum²⁹.

I lavori sono stati svolti dai singoli ricercatori nelle loro sedi di appartenenza, ma è evidente come un lavoro di questo tipo non avrebbe potuto essere condotto senza un rapporto di continua collaborazione. Incontri specifici sono stati comunque calendarizzati e si sono regolarmente svolti: a Roma ad inizio progetto (workshop metodologico del 22 marzo 2017), a Venezia a metà percorso (*La falsificazione epigrafica in Italia. Questioni di metodo e casi di studio*, Università Ca' Foscari, 10-11 ottobre 2018)³⁰, a Gargnano quando il progetto si avviava alla conclusione (*False notizie...fake news e Storia Romana. Falsificazioni antiche, falsificazioni moderne*, Fondazione Feltrinelli, 3-5 giugno 2019)³¹.

EDF non è lo scopo finale del progetto, ma è senza alcun dubbio uno strumento senza il quale sarà più difficile ripensare il concetto di falso epigrafico.

The present paper provides a short history of EAGLE: it was born in 2003 as Electronic Archive of Greek and Latin Epigraphy, a federation of databases that share the same concept of digital epigraphy; in the period 2013-2016 it becomes Europeana network of Ancient Greek and Latin Epigraphy, aggregator and content provider of Europeana; since 2020 it includes also EDF (Epigraphic Database Falsae), the first database devoted to different kinds of epigraphic forgeries.

²⁵ <https://edh-www.adw.uni-heidelberg.de/foto/suche>.

²⁶ <https://cil.bbaw.de/index.php?id=17>.

²⁷ <http://www.ubi-erat-lupa.org/>.

²⁸ <http://www.bsrdigitalcollections.it/photos.aspx>.

²⁹ <http://epicum.istc.cnr.it/EPICUM/partners>.

³⁰ I risultati dell'incontro sono ora editi in *La falsificazione epigrafica. Questioni di metodo e casi di studio*, a cura di L. Calvelli (Antichistica 25, Storia ed epigrafia 8), Venezia: ECF - Edizioni Ca' Foscari, 2019, consultabile anche all'indirizzo <<https://edizionicafoscariv.unive.it/media/pdf/books/978-88-6969-387-8/978-88-6969-387-8.pdf>>.

³¹ I contributi presentati sono ora pubblicati in *False notizie...fake news e storia romana. Falsificazioni antiche, falsificazioni moderne*, a cura di S. Segenni (Studi sul Mondo Antico 13), Milano: Mondadori, 2019.

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2020

Iste liber est... on-line. Una rassegna di strumenti in rete per lo studio e il censimento dei *provenance marks**

«DigItalia» 2-2020
DOI: 10.36181/digitalia-00021

Lucrezia Signorello

Ph.D. Student. Sapienza Università di Roma

Negli ultimi decenni si è riscontrato un crescente interesse per lo studio dei provenance marks quali fonte privilegiata per l'analisi e la ricostruzione delle collezioni librarie, nonché delle pratiche di lettura e della circolazione di manoscritti e libri a stampa. Numerosi sono, pertanto, i progetti di censimento e descrizione di tali segni di provenienza realizzati da varie istituzioni, sia a livello nazionale che internazionale. Il contributo vuole richiamare brevemente i presupposti teorici e metodologici alla base di tali iniziative e offrire una panoramica sommaria ma esemplificativa di alcuni strumenti resi disponibili in rete in tale ambito di ricerca.

*How is provenance materialized,
made visible, reified, expressed?
What triggered Walter Benjamin's reveries
about past histories of each volume
while he was unpacking his library?*¹

Già negli anni Settanta del secolo scorso, José Ruyschaert evidenziava come – anche nello studio della decorazione libraria –

* La presente trattazione non si propone, né potrebbe essere, esaustiva riguardo a tale tematica. Essa si limita quindi a delineare per sommi capi l'argomento, e a fornire contestualmente alcuni spunti e indicazioni sugli strumenti e le banche dati attualmente disponibili, con la consapevolezza che l'elencazione potrebbe essere esponenzialmente ampliata. A tal proposito, si segnala l'utile rassegna – in continuo aggiornamento – dedicata alle risorse disponibili a livello internazionale fornita dal Consortium of European Research Libraries (CERL): <<https://www.cerl.org/resources/provenance/geographical>>. Sempre su questo tema si può consultare anche la linkografia curata dal Gruppo di ricerca sulle Biblioteche filosofiche private in età moderna della Scuola Normale Superiore di Pisa: <<http://picus.unica.it/index.php?page=Linkografia>>. Per quanto riguarda le immagini a corredo del testo: fig. 1 su concessione dell'Accademia dell'Arcadia, fig. 2-3 tratte dalle collezioni in public domain di Europeana (<<https://www.europeana.eu/it>>), fig. 4-5 screen dei siti web del Provenance Digital Archive (<https://www.arkyves.org/r/section/him_CERLPDA/>) e di Manus OnLine (<https://manus.iccu.sbn.it/indici_detail.php?id=3>).

¹ Cfr. Gail Feigenbaum, *Manifest Provenance*, in: *Provenance: An Alternate History of Art*, edited by G. Feigenbaum and I. Reist, Los Angeles: Getty Research Institute, 2012, p. 6-28: 6. Feigenbaum si riferisce qui al breve ma affascinante scritto di Walter Benjamin (1892-1940) dal titolo *Ich packe meine Bibliothek aus. Eine Rede über das Sammeln* (1931), nel quale il filosofo tedesco coglie l'occasione di

«si impongono sempre di più le tortuose ricerche, che si sforzano di stabilire dove e quando quel tale bel libro è stato prodotto: libri di conti, registri comunali o notarili, ma anche ricostruzioni di biblioteche antiche, partendo sia dai vecchi inventari, sia dall'esame diretto dei volumi, che permette di circoscrivere in maniera più precisa le circostanze di tempo e di luogo della creazione»².

In effetti, sebbene l'attenzione, o comunque una certa sensibilità, per le provenienze librarie³ e per la contestualizzazione dei volumi, manoscritti o a stampa, all'interno della rispettive collezioni di appartenenza – presenti o passate – sia di lunga data negli studi dedicati al libro, si è affermata sempre con maggior forza la consapevolezza che tali informazioni non riguardano solo la storia del collezionismo e delle biblioteche, ma che esse possono essere altresì poste al servizio di percorsi di ricerca in aree limitrofe, arricchendone il contesto di riferimento e conferendo loro profondità storica⁴. Se inizialmente hanno stimolato l'attenzione degli studiosi principalmente le

un trasloco per riflettere sul collezionismo librario; per un'edizione italiana si veda: Walter Benjamin, *Disfo la mia biblioteca. Discorso sul collezionismo*, in: Walter Benjamin, *La mia biblioteca*, introduzione di P. Mauri, traduzione di C. Guarnieri, Roma: Elliot, 2016, p. 19-33; per un approfondimento su questo saggio di Benjamin si può partire da: Joseph D. Lewandowski, *Unpacking: Walter Benjamin and His Library*, «Libraries & Culture», 34 (1999), n. 2, p. 151-157.

² Cfr. José Ruyschaert, *La miniatura italiana del Rinascimento*, in: *Libri, scrittura e pubblico nel Rinascimento. Guida storica e critica*, a cura di A. Petrucci, Roma – Bari: Laterza, 1979, p. 59-78: 67-68.

³ A livello terminologico è necessario ricordare la differenziazione semantica prevista tra “provenienza” e “possessore”, per la quale mi rifaccio alle indicazioni fornite dall'Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU): «Si considerano provenienze persone ed enti che hanno avuto la proprietà o il possesso del singolo esemplare. La loro gestione prevede un legame tra l'inventario e i nomi di persone e/o enti che hanno posseduto la copia. Il legame può essere di due tipi: Possessore e Provenienza. Ciascun esemplare può avere più legami Possessore, ma un solo legame Provenienza da intendersi come l'ultima persona o ente che ha posseduto la copia», cfr. *Guida alla catalogazione in SBN – Materiale antico* <https://norme.iccu.sbn.it/index.php?title=Guida_antico/Descrizione/Informazioni_relative_all%27esemplare/Provenienze&oldid=4320>. Tale differenziazione non viene applicata, tuttavia, nella presente trattazione, nella quale i due termini vengono utilizzati in maniera sinonimica.

⁴ L'attenzione riservata alle collezioni riconducibili a singoli possessori (istituti o persone fisiche) – come è noto – si è concretizzata nel tempo in un'ampia varietà di trattazioni (cataloghi, saggi su tematiche specifiche, mostre ecc.) sulle quali è, pertanto, certamente superfluo soffermarsi. Si vuole rimarcare, invece, come in tale ambito di ricerca ci si concentri sempre di più anche su una certa tipologia di documentazione, per così dire “esterna”, quale fonte per l'indagine della produzione libraria e per l'identificazione e ricostruzione di antiche collezioni, tra cui epistolari e antichi inventari (spesso conservati separatamente dalle raccolte alle quali fanno riferimento). Anche in questo caso un'elencazione completa o comunque ampiamente rappresentativa sarebbe impossibile, ci si limita quindi a citare solo alcuni esempi. Per quanto riguarda gli epistolari e le informazioni che si possono desumere dalle corrispondenze, sembra opportuno citare il progetto Vespasiano da Bisticci, Lettere. A semantic digital edition, curato dal Centro di Ricerche per la Ricerca-Multimedia (CRR-MM) dell'Alma Mater Studiorum – Università di Bologna: esso si propone come l'«Edizione digitale delle Lettere di Vespasiano da Bisticci, copista fiorentino vissuto nell'arco del XV secolo. Le lettere, inviate e ricevute e fino ad oggi rintracciate, sono navigabili attraverso un sistema a faccette (corrispondente, data, luogo, segnatura). L'edizione è accompagnata da strumenti filologici di orientamento e guida (indice delle parole sottoforma di authorities, tavola sinottica, nota filologica, descrizione dei testimoni). Le lettere sono accompagnate da informazioni contestuali (la raccolta delle lettere, i corrispondenti, i manoscritti realizzati dalla scuola, i copisti di Vespasiano, le biblioteche

collezioni riconducibili a figure eminenti del panorama storico-culturale mondiale, si assiste ormai da alcuni decenni a un crescente interesse anche per le pratiche di lettura e quindi di possesso dei libri da parte di figure, per così dire, “secondarie”⁵:

«Interest in provenance information goes in and out of fashion. Once it was a bibliophilic interest concerned with authors’ association copies and books belonging to great men. Following the rise of the history of the book in the 1980s and 1990s, provenance studies have become an important ingredient in the work of social and cultural historians dealing with questions of readership and literacy. The ownership of books by craftsmen and women is now as significant as that of kings and archbishops»⁶.

prodotte) necessarie ad inserire i documenti in una prospettiva metatestuale» (cfr. <<http://vespasiano-dabisticileters.unibo.it/index.html>>). In riferimento, invece, agli studi dedicati agli antichi inventari ancora oggi conservati, è necessario sicuramente menzionare il progetto della Biblioteca Apostolica Vaticana RIC (Ricerca sull’Inchiesta della Congregazione dell’Indice): esso è dedicato all’analisi dei codici Vaticani Latini 11266-11326, nei quali «sono conservate le liste dei titoli dei libri posseduti dai religiosi dei conventi e dei monasteri italiani, acquisite dalla S. Congregazione dell’Indice dei libri proibiti dopo la pubblicazione dell’*Index librorum prohibitorum* da parte di papa Clemente VIII nel 1596, in un periodo che arriva sino al 1603. Esse riguardano 31 ordini regolari maschili e contengono elenchi corrispondenti a biblioteche monastiche e conventuali ovvero in uso di singoli monaci e frati. Occasionalmente sono stati inclusi nei codici anche liste di libri di monache, di ecclesiastici e di laici» (cfr. <<http://rici.vatlib.it/>>). La necessità di avere contezza dell’esistenza di queste fonti nello studio delle collezioni librarie è, inoltre, alla base della redazione del repertorio RICABIM (Repertorio di Inventari e Cataloghi di Biblioteche Medievali) – curato dalla Società Internazionale per lo Studio del Medioevo Latino (SISMEL) e pubblicato sia a stampa che online – il cui obiettivo primario è «la realizzazione di un repertorio di cataloghi, di inventari e di qualsiasi altra attestazione libraria presente in fonti archivistiche (come testamenti, donazioni, transazioni ed atti di varia natura) o in registri preposti a scopi precisi (note di prestito, note di spesa, note da obituari) contenenti riferimento immediato a manoscritti ed incunabili, anticamente posseduti, o sotto forma di catalogo dettagliato o di semplice lista o di enumerazione sommaria degli stessi, purché antecedenti al 1520 e senza limiti geografici in seno all’Occidente latino» (cfr. <<https://www.sismelfirenze.it/index.php/banche-dati/biblioteche-medievali-ricabim>>).

⁵ Significativo è stato, in tal senso, il 15cBOOKTRADE Project della University of Oxford diretto da Cristina Dondi (figura di spicco del CERL, di cui è il segretario, nonché docente presso l’ateneo oxfordiano), i cui risultati sono stati resi disponibili, oltre che nel sito web dedicato (<<http://15cbooktrade.ox.ac.uk/>>), anche in una mostra pensata per avvicinare il grande pubblico a queste tematiche, allestita al Museo Correr e alla Biblioteca Marciana di Venezia (1 settembre 2018 – 30 aprile 2019) dal titolo “PRINTING R-EVOLUTION AND SOCIETY, 1450-1500. Fifty Years that changed Europe” (su cui si veda: *Printing R-Evolution and Society 1450-1500. Fifty Years that Changed Europe*, edited by C. Dondi, Venezia: Edizioni Ca’ Foscari – Digital Publishing, 2020).

⁶ È quanto evidenziato in relazione alla storia degli studi di settore nella sezione dedicata alle provenienze del sito web del CERL (cfr. <<https://www.cerl.org/resources/provenance/main>>). Nell’ambito di questa sempre più marcata tendenza degli studi e nel solco della conseguente attenzione per le provenienze si colloca, peraltro, l’azione di sostegno all’attività di catalogazione e di promozione alla pubblicazione di strumenti catalografici aggiornati promossa dalla Commissione Nazionale “Indici e Cataloghi delle Biblioteche Italiane” con riguardo ai fondi provenienti dalle biblioteche claustrali, azione per la quale si rimanda a: Angela Adriana Cavarra, *Premessa*, in: Lucrezia Signorello, *Catalogo dei manoscritti di Santa Maria del Popolo della Biblioteca Angelica*, Roma: Istituto Poligrafico e Zecca dello Stato, 2019 (Indici e Cataloghi, 28), p. 7-10: 8-9. Se la presente trattazione si vuole concentrare sulle risorse interrogabili online, sembra tuttavia opportuno ricordare, avendo richiamato l’interesse per le provenienze connesse alla catalogazione dei fondi ecclesiastici incamerati dal Demanio all’indomani dell’Unità nazionale (tale passaggio era stato regolato dal r.d. 3036 del 7 luglio 1866 per la soppressione degli Ordini e delle

In tale contesto, lo studio e il censimento dei *provenance marks*⁷, quali fonte primaria per la ricostruzione dei contesti di committenza, possesso, utilizzo e circolazione dei volumi ottiene sempre maggiore attenzione all'interno delle azioni di valorizzazione delle collezioni librerie; ciò tanto nell'ambito delle campagne catalografiche⁸ – attività d'altronde preliminare a qualunque tipo di studio si voglia intraprendere a partire da o sul libro – quanto nei progetti dedicati al patrimonio bibliografico secondo le diverse e specifiche prospettive⁹.

Ad oggi il panorama internazionale delle risorse in rete disponibili per lo studio dei *provenance marks* è quanto mai ricco ed estremamente variegato¹⁰, offrendo agli

Corporazioni religiose, in esecuzione della l. 2987 del 28 giugno 1866, che aveva stabilito lo scioglimento delle congregazioni religiose, mentre la l. 3848 del 15 agosto 1867 per la liquidazione dell'Asse ecclesiastico aveva previsto l'incameramento dei loro beni da parte del Demanio pubblico; gli effetti di tali provvedimenti erano stati successivamente estesi a Roma con la Legge di soppressione delle Corporazioni religiose di Roma del 19 giugno 1873), l'importante contributo – corredato da riproduzioni di timbri, note di possesso, ex-libris e antiche segnature – fornito in: Marina Venier, *Librerie dei conventi riunite nella Vittorio Emanuele*, disponibile nella pagina del sito web del CERL dedicata alle *Online provenance resources s.v. Rome, Biblioteca Nazionale Centrale*: <<https://www.cerl.org/resources/provenance/geographical>> (link diretto al pdf: <https://www.cerl.org/_media/resources/provenance/librerie_dei_conventi_riuinite_nella_vittorio_emanuele_11_.pdf>); la studiosa è peraltro tornata a più riprese sull'argomento in varie pubblicazioni, che ne hanno approfondito, aggiornato e ampliato la trattazione; si consiglia dunque la consultazione della sua ricca bibliografia per maggiori informazioni sulla costituzione del patrimonio della Biblioteca Nazionale Centrale "Vittorio Emanuele II" di Roma (BNCR) a partire dalle collezioni conventuali confiscate.

⁷ Attribuendo qui a tale categoria l'accezione semantica più ampia possibile, includente non solo i marcatori di appartenenza più ovvi e comuni (note di possesso, stemmi, ex-libris, timbri, etichette), ma anche quanto desumibile dalla decorazione e dalle legature, dai segni di uso e lettura, di censura e del trattamento biblioteconomico; si veda, per delle primissime indicazioni sulla classificazione di tali segni e per delle più ampie indicazioni bibliografiche, le pagine dedicate a questa tematica in: Andrea De Pasquale, *Che cos'è la bibliologia?*, Roma: Carocci, 2018. Tale significato inclusivo viene preso in considerazione, d'altronde, in alcuni dei progetti di censimento delle provenienze di cui si tratterà più avanti (vedi infra).

⁸ «È solo la sistematica opera di catalogazione dei fondi antichi che consente di raccogliere, ordinare, classificare e mettere in sistema ampie masse di dati» ricorda Luca Rivali «Da questo punto di vista diventa allora indispensabile la formazione del bibliotecario e la sua applicazione a progetti di sistematica catalogazione del materiale librario antico al fine di mettere a disposizione degli studiosi queste preziose informazioni storiche», cfr. Luca Rivali, *Storia del libro e provenienze*. Introduzione al dossier, «La Bibliofila», 117 (2015), n. 3, p. 309–317: 311.

⁹ Le campagne di censimento sono state accompagnate, evidentemente, anche da una riflessione teorica e metodologica, che, se come spesso accade, è sollecitata da esigenze pratiche e talvolta prettamente contingenti, risulta peraltro necessariamente preliminare a qualunque azione concreta che miri a ottenere risultati strutturati, coerenti e fruibili. Solo per citare un esempio: *Provenienze. Metodologia di rilevamento, descrizione e indicizzazione per il materiale bibliografico*, documento elaborato dal Gruppo di lavoro sulle provenienze coordinato dalla Regione Toscana e dalla Provincia Autonoma di Trento, Trento: Provincia autonoma di Trento, Soprintendenza per i beni librari e archivistici, Firenze: Regione Toscana, Giunta regionale, 2009.

¹⁰ Tale abbondanza, se da una parte testimonia la vivacità degli studi in questo campo e la sensibilità per la rilevazione e descrizione dei *provenance marks*, è segno d'altronde della frammentazione delle iniziative in tale ambito, con progetti spesso facenti campo ai singoli istituti di conservazione o ricerca. Ciò impone all'utenza interessata non solo la consultazione di innumerevoli banche dati ma anche il rischio di non riuscire a reperire le informazioni di cui si è alla ricerca per la non conoscenza di tutti gli strumenti disponibili.

studiosi numerosi strumenti per ricerche sulla provenienza dei volumi, sia manoscritti che a stampa, e sulla loro contestualizzazione storico-culturale. I progetti di censimento e di messa a disposizione delle informazioni per l'utenza (specializzata o meno) hanno seguito e seguono criteri diversi, non solo per quanto riguarda l'architettura dei database, ma anche per la terminologia impiegata, per i dati che si è scelto di volta in volta di rilevare e per la disponibilità o meno di riproduzioni digitali dei segni di provenienza collegate alle relative descrizioni¹¹.

In questo ambito, il livello di indagine sulle provenienze più prossimo – o meglio contestuale – alle comuni azioni di catalogazione del patrimonio bibliografico è certamente quello costituito dalla rilevazione di segni di possesso sull'esemplare oggetto di schedatura e quindi della resa di tali dati negli OPAC tramite il legame della singola copia con il suo precedente possessore. L'utenza è così in grado di rintracciare attraverso la ricerca nel catalogo (preferibilmente effettuata secondo la modalità "avanzata") tutti i volumi riconducibili a un medesimo possessore all'interno del patrimonio generale delle biblioteche.

Derivano direttamente da tali operazioni catalografiche una serie di progetti sia in ambito nazionale che internazionale, circoscritti a singoli istituti o trasversali a più enti di conservazione, che dall'analisi di questi dati hanno tratto non solo la redazione di authority file ma anche creato delle vere e proprie banche dati.

La presenza di una lista controllata relativa alle provenienze è, ad esempio, riscontrabile nel progetto Manus Online – censimento dei manoscritti conservati nelle biblioteche italiane pubbliche, ecclesiastiche e private – curato dall'Area di attività per la bibliografia, la catalogazione e il censimento dei manoscritti dell'ICCU¹², rimarchevole innanzitutto per il suo respiro nazionale e quindi programmaticamente unitario. Il progetto, nato nel 1988, dedica oggi particolare attenzione alla redazione di record di autorità riguardanti le varie categorie di persone ed enti che hanno giocato un ruolo nella produzione o nella storia dei manoscritti descritti¹³.

¹¹ Si avverte qui, in effetti, come in molti altri contesti, la necessità di una maggiore normalizzazione per quanto riguarda gli standard descrittivi, di un vocabolario controllato e di un più forte coordinamento a livello nazionale e internazionale.

¹² Manus Online è raggiungibile all'indirizzo: <<https://manus.iccu.sbn.it/>>. Il sito web del progetto offre inoltre la possibilità di consultare direttamente online alcune pubblicazioni di carattere normativo per la redazione delle descrizioni e segnala anche l'organizzazione, da parte del personale ICCU, dei corsi di formazione dedicati ai bibliotecari e agli studiosi di settore per l'implementazione della base dati. Su questo progetto curato dall'ICCU, tra i vari contributi, si veda: Roberto Marcuccio, *Catalogare e fare ricerca con Manus online*, «Biblioteche oggi», 28 (2010), n. 6, p. 33-49; Claudia D'Arcamo, *I manoscritti in rete. Specimina del software Manus Online*, «Mediaeval Sophia», 14 (2013), p. 133-145; Giliola Barbero – Francesca Trasselli, *Manus OnLine and the Text Encoding Initiative Schema*, «Journal of the Text Encoding Initiative», 8 (2014-2015): <<https://journals.openedition.org/jtei/1054>>.

¹³ Sulla redazione di tali record di autorità e sulle questioni metodologiche ad essa connesse si veda: Valentina Atturo, *Problematiche e sviluppi recenti a proposito del controllo di autorità nella catalogazione dei manoscritti*, «AIB Studi», 59 (2019), n. 1/2, p. 251-270: 253.

Le liste prese in considerazione, e incluse nella sezione “Indici” del database, sono dedicate ai copisti, agli illustratori, alle provenienze e alle origini – tutte categorie che, come si è detto, rientrano a vario ruolo nell’ampio spettro dei produttori di *provenance marks* – ricercabili secondo il duplice criterio alfabetico e cronologico. La realtà italiana – costellata di numerosissimi istituti di conservazione di prestigio diffusi in maniera capillare sul territorio – ha favorito, d’altronde, la proliferazione di progetti che si interessano di provenienze e che fanno capo ai singoli enti. Tra di essi, e solo per richiamare all’attenzione alcuni dei più noti, sembra opportuno ricordare le risorse rispettivamente gestite da due tra le più rimarchevoli istituzioni bibliotecarie della Penisola: la Biblioteca Nazionale Marciana di Venezia (BNM) e la Biblioteca Nazionale “Vittorio Emanuele III” di Napoli (BNN). Dal 2014, spinta dall’«esigenza particolarmente sentita sia dai catalogatori che dagli utenti della Biblioteca [...] di poter repertoriare e ordinare nomi e immagini di contrassegni di possesso, identificati e non, per poter giungere a un definitivo riconoscimento o almeno per poter raggruppare le segnature che presentassero gli stessi contrassegni di possesso»¹⁴, la Marciana implementa il suo Archivio dei possessori, dedicato alla rilevazione, al censimento e alla documentazione fotografica delle provenienze



Figura 1. *Accademia dell'Arcadia, Atti arcadici 1, p. 100 (sigillo)*

di stampati e manoscritti della biblioteca¹⁵. L’Archivio possessori della Nazionale di Napoli¹⁶, nato anch’esso nel 2014, correda similmente le schede relative ai possessori con «le immagini dei segni di possesso reperiti su volumi», ciò nella consapevolezza che «Ex libris, timbri, annotazioni manoscritte, legature forniscono preziose informazioni sulla storia di ciascun esemplare e consentono, in prospettiva, di ricostruire virtualmente la composizione di antiche raccolte librerie disperse»¹⁷. I due progetti uniscono,

¹⁴ Cfr. Elisabetta Sciarra – Orsola Braides, *L’Archivio dei possessori in Biblioteca Nazionale Marciana*, «Gazette du livre médiéval», 61 (2014), p. 86-88: 86.

¹⁵ Il sito è raggiungibile all’indirizzo: <<https://marciana.venezia.sbn.it/la-biblioteca/cataloghi/archivio-possessori>>.

¹⁶ Raggiungibile all’indirizzo: <<http://www.bnnonline.it/index.php?it/330/archivio-possessori>>.

¹⁷ *Ibidem*. Per un necessario approfondimento su questo progetto si faccia innanzitutto riferimento a: Simona Pignalosa, *L’Archivio Possessori della Biblioteca Nazionale di Napoli*, «La Bibliofila», 117 (2015), n. 3, p. 344-350.

pertanto, la rilevazione e strutturazione dei dati alle riproduzioni dei segni di provenienza e permettono una ricerca strutturata e guidata grazie alla possibilità di combinare vari campi al momento della consultazione delle banche dati. Certo questa casistica non è esauriente e si potrebbero elencare altre risorse, come l'Archivio dei possessori della Biblioteca dell'Archiginnasio di Bologna¹⁸ o la sezione dedicata ai possessori del portale della Biblioteca Statale di Montevergine¹⁹.

Rimarchevoli sono anche i blasonari – disponibili online e interrogabili secondo diverse chiavi di ricerca – curati da istituzioni di primo piano quali l'Archivio di Stato di Firenze²⁰, la Biblioteca Casanatense di Roma²¹ e la Biblioteca del Kunsthistorisches Institut in Florenz – Max-Planck-Institut²².

Tali iniziative si sono sviluppate all'interno del più ampio e complesso contesto internazionale. Si contano, difatti, numerose campagne di indagine e relativi strumenti di accesso ai risultati ottenuti, sviluppati dalle varie nazioni o da istituti dedicati a un particolare ambito disciplinare²³.

Fra le innumerevoli risorse a disposizione degli studiosi, in ambito europeo si può citare, ancora una volta a solo titolo

esemplificativo, la sezione dedicata alle provenienze nel database Numelyo, della Bibliothèque municipale de Lyon. Il progetto è nato nell'ambito della digitalizzazione del patrimonio bibliografico di questa biblioteca e di altre istituzioni affiliate; in tale circostanza, una particolare attenzione per i *provenance marks* è stata sollecitata dalla consapevolezza che «Les marques de provenance des livres anciens permettent d'établir l'itinéraire géographique et intellectuel des livres en identifiant leurs anciens possesseurs et lecteurs»²⁴. Anche il



Figura 2. Treviso, biblioteca comunale, ms. 416, c. 1r (stemma famiglia Trevisan)

¹⁸ La risorsa, inserita nel ricco portale della Biblioteca Digitale dell'Archiginnasio – Archiweb, è raggiungibile all'indirizzo: <<http://badigit.comune.bologna.it/possessori/>>.

¹⁹ Si veda il seguente link: <<http://bibliotecastataledimontevergine.benculturali.it/index.php?it/309/possessori>>.

²⁰ I blasoni qui descritti, a cura dell'Archivio di Stato di Firenze, provengono dalla Raccolta Ceramelli Papiani: <<http://www.archiviodistato.firenze.it/ceramellipapiani/>>.

²¹ Per maggiori informazioni si veda: <<https://casanatense.benculturali.it/?p=208>>.

²² La risorsa è disponibile al seguente indirizzo: <<http://wappen.khi.fi.it/Plone>>.

²³ Per una loro elencazione più estesa e sistematica si faccia riferimento alle indicazioni fornite nella nota n. 1 del presente contributo.

²⁴ Si veda la sezione relativa alla *provenance des livres anciens* di tale progetto al seguente indirizzo: <https://numelyo.bm-lyon.fr/collection/BML:BML_06PRV01000COL0001>.



Figura 3. Treviso, biblioteca comunale, ms. 16, c. 2r (nota di possesso di Agostino da Treviso)

portale della Real Biblioteca di Madrid²⁵ offre molteplici risorse – ovviamente corredate da digitalizzazioni – dedicate agli ex-libris, alle legature e alla decorazione delle stesse, che si configurano come uno strumento pensato «para describir, clasificar, estudiar, y, por supuesto, divulgar esta colección» prestando «especial atención al tratamiento de los hierros, ya que es el componente de la encuadernación que, además de su interés artístico, permite la adscripción de las piezas al taller o encuadernador responsable de su producción»²⁶. Attraversando l’Atlantico, si può ricordare, invece, il Provenance Online Project della Penn University of Pennsylvania²⁷, nel quale sono particolarmente efficaci le strategie di comunicazione e di coinvolgimento degli utenti nel lavoro di ricerca e identificazione dei *provenance marks* messe in campo da parte del personale di questa biblioteca universitaria statunitense.

Come si è visto, abbondanza di risorse e frammentazione dei progetti sono le caratteristiche che connotano il panorama non solo italiano ma mondiale nel campo degli strumenti disponibili per lo studio delle provenienze. In un contesto così sfaccettato è essenziale, pertanto, sottolineare come esistano anche iniziative inclusive e coordinate, ispirate a principi di partecipazione e collaborazione internazionale. Tra queste, è opportuno ricordare, anche per la specificità dell’ambito disciplinare, il progetto Footprints: Jewish Books Through Time and Place, promosso dal Center for Jewish History di New York insieme al Columbia Center for Teaching and Learning (CTL): esso si configura come un database mirato a tracciare «the circulation of printed “Jewish books”», grazie anche alla collaborazione di varie istituzioni e al coordinamento con altre iniziative simili; esso si basa sulla consapevolezza che «The history of the book is an important part of humani-

²⁵ Dalla homepage del sito della Real Biblioteca del Palacio Real de Madrid (cfr. <<https://www.realbiblioteca.es/>>) è possibile accedere alle risorse dedicate ai *provenance marks*.

²⁶ Così sono delineate le finalità che sottendono al progetto, cfr. <<https://realbiblioteca.es/es/recursos/>>.

²⁷ Il portale è raggiungibile all’indirizzo: <<https://www.flickr.com/photos/58558794@N07/>>.

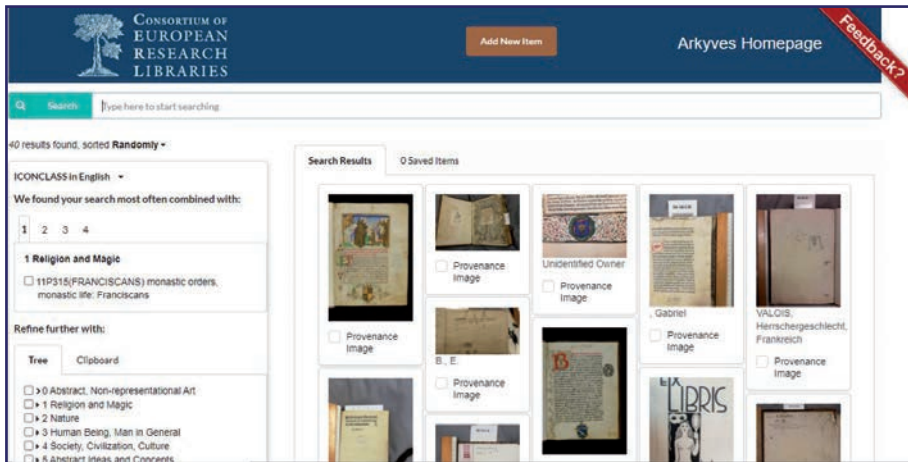


Figura 4. CERL Provenance Digital Archive (screen sito web)

ties scholarship. Especially as more books are digitized, scholars, librarians, collectors, and others have become increasingly attuned to the significance of individual books as objects with their own unique story»²⁸.

In ambito europeo, lo spirito consorziale si esprime certamente al meglio nei progetti promossi dal CERL²⁹. Tali iniziative, difatti, non solo rappresentano un punto di riferimento a livello internazionale³⁰, ma costituiscono anche uno stimolo co-

²⁸ Il progetto è raggiungibile all'indirizzo: <<https://footprintsctl.columbia.edu/>>. Come ricordato nella homepage del portale: «Jewish books in particular tell a fascinating story about the spread of knowledge and faith in a global Diaspora. Every literary work represents a moment in time and space where an idea was conceived and documented. But the history of a book continues long after composition as it is bought, sold, shared, read, confiscated, stored, or even discarded. This history is the essence of Footprints».

²⁹ Obiettivo delle attività del CERL è il miglioramento e la facilitazione degli studi riguardanti il patrimonio bibliografico, sia manoscritto che a stampa; tale finalità è perseguita tramite la creazione di risorse e strumenti digitali di alta qualità, unitamente all'organizzazione di seminari e workshop, e sostenendo la cooperazione tra le varie istituzioni interessate. Per maggiori informazioni si veda: <<https://www.cerl.org/main>>. Degli strumenti digitali curati dal CERL nel campo dello studio delle provenienze e della ricostruzione delle antiche biblioteche ha recentemente trattato Cristina Dondi nel seminario "Strumenti digitali e metodi innovativi per la storia delle biblioteche", terzo incontro del ciclo di videoconferenze "Culture del libro e del documento. Temi, questioni, linee di evoluzione" – organizzato dal Dipartimento di Studi Storici dell'Università di Torino – del 23 giugno 2020 (cfr. <https://www.unito.it/eventi/culture-del-libro-e-del-documento-ciclo-di-videoconferenze-cura-del-dip-di-studi-storici?fbclid=IwAR0IkloCHlqGXeKuvJhV8mzGi2MuBVaYxBAjv5u8_59DCEqDjMhqFN_9e10>).

³⁰ A tal proposito, solo per fare un esempio in cui le indicazioni proposte dal CERL sono state prese a modello – con alcuni adattamenti – per una campagna catalografica indipendente, si desiderano citare le parole di Francesca Aiello e Silvia Tripodi in relazione a un progetto della Biblioteca della Società di storia patria per la Sicilia orientale: «Nel nostro studio è stato un punto di riferimento il MEI (Material Evidence in Incunabula) [...]. L'approccio che appartiene alla filosofia del MEI è stato qui applicato alla produzione editoriale del XVI secolo», cfr. Francesca Aiello – Silvia Tripodi, *Segni d'uso nei libri del XVI secolo della Biblioteca della Società di storia patria per la Sicilia orientale*, in:

stantemente foriero di nuove scoperte e quindi di valorizzazione del patrimonio bibliotecario, e non solo, delle singole nazioni³¹. Tra i progetti messi in campo dal CERL sulle provenienze³² occupa un ruolo di primo piano il MEI (Material Evidence in Incunabula)³³, concepito con la finalità di ricostruire la storia del singolo esemplare e, attraverso ciò, quella delle antiche collezioni librarie nonché della circolazione degli incunaboli³⁴. L'approccio proposto dal MEI – che nel tempo è divenuto un canone catalografico – è stato applicato «to answer the need to use not only conventional provenance to retrace the history of books (such as former owners' various marks) but physical evidence too»³⁵. L'ampliamento del ventaglio dei *provenance marks* presi in considerazione ha, evidentemente, arricchito e approfondito notevolmente la comprensione dei percorsi seguiti dai volumi nei vari passaggi di mano di cui sono stati oggetto e che non sempre, o meglio solo fortunosamente, sono documentati o resi espliciti sui libri stessi. Al MEI sono poi collegati altri database satelliti: l'Owners of Incunabula³⁶, il CERL Thesaurus³⁷, l'Holding Institutions³⁸, come anche l'ISTC (Incunabula Short Title Catalogue)³⁹. Il sistema

Per libri e per scritture. Contributi alla storia del libro e delle biblioteche nell'Italia meridionale tra XVI e XVIII secolo, a cura di S. Insera, Milano: Ledizioni, 2018, p. 60-77: 62.

³¹ Anche in questo caso si ritiene opportuno riportare direttamente alcune frasi particolarmente significative: «The involvement of the main Lithuanian libraries, from early 2017, in the creation of the international database of incunabula provenance, Material Evidence in Incunabula (MEI), gave a stimulus to a more attentive examination of incunabula and their ownership marks and thus led to a number of important discoveries», cfr. Viktorija Vaitkevičiūtė – Agnė Zemkajūtė, *How Provenance Marks from Lithuanian Incunabula are Contributing to Historical Narrative*, in: *Printing R-Evolution and Society 1450-1500. Fifty Years that Changed Europe*, edited by C. Dondi, Venezia: Edizioni Ca' Foscari – Digital Publishing, 2020, p. 431-462: 433.

³² L'attenzione dedicata a tale aspetto dello studio dei libri dal CERL nasce «In recognition of the fact that our libraries are made of parts of libraries, private and institutional, dispersed over the course of hundreds of years, due to historical, political, administrative, economic, as well as casual reasons», cfr. Cristina Dondi, *CERL's work and vision for provenance research I: CERL Thesaurus, Material Evidence in Incunabula, and the 15cBOOKTRADE Project*, «La Bibliofila», 117 (2015), n. 3, p. 317-321: 318; pertanto solo attraverso la comprensione dei processi e dei percorsi che hanno portato alla costituzione delle attuali collezioni librarie è possibile comprenderne appieno il valore e il significato storico-culturale.

³³ Il MEI - base dati creata da Cristina Dondi e progettata per la registrazione e la ricerca dei dati materiali dei libri stampati nel XV secolo - è raggiungibile all'indirizzo: <https://data.cerl.org/mei/_search>.

³⁴ Cfr. L. Rivali, *Storia del libro e provenienze*, cit., p. 313.

³⁵ Cfr. C. Dondi, *CERL's work and vision*, cit., p. 319.

³⁶ L'Owners of Incunabula raccoglie informazioni biografiche sui possessori precedenti e attuali, sia che si tratti di persone fisiche che di enti (compresi legatori, miniatori, librai ecc.). Per tale database dedicato ai possessori degli incunaboli si veda: <https://data.cerl.org/owners/_search>.

³⁷ Il CERL Thesaurus, che contiene denominazioni di luoghi di stampa, nomi di tipografi, nomi di persona e di enti ricavati dagli Authority File delle biblioteche aderenti al CERL e da vari progetti di ricerca sulla storia del libro, è disponibile all'indirizzo: <https://data.cerl.org/thesaurus/_search>.

³⁸ Per la base dati del CERL dedicata alle istituzioni di conservazione presenti nel mondo si veda: <https://data.cerl.org/holdinst/_search>.

³⁹ Catalogo, curato dalla British Library, delle edizioni comprese fino al 1500 con segnalazione delle relative copie conservate nei vari istituti a livello internazionale: <https://data.cerl.org/istc/_search>.



Figura 5. Manus Online - Indici provenienze

costruito dal CERL risulta così ricco e approfondito, ma anche di semplice consultazione, grazie alla facile navigabilità dei singoli database e del passaggio guidato da un progetto all'altro. Recentemente, infine, anche questa costellazione di iniziative si è dotata di uno strumento che mette a disposizione degli utenti le riproduzioni digitali dei segni di provenienza: il PDA (Provenance Digital Archive)⁴⁰. In conclusione, sulla scorta di quanto affermato da Alfredo Serrai in riferimento alla Storia delle biblioteche, alla quale «competono, anzitutto, la definizione circostanziata di Biblioteca, la individuazione delle sue origini, della sua personalità libraria e della sua identità bibliografica, e, quindi, la ricostruzione dei suoi accrescimenti e del suo sviluppo, insieme alle eventuali modifiche nella sua costituzione, nella sua destinazione, e nelle sue funzioni»⁴¹, si deve rimarcare come tale indagine debba partire necessariamente dallo studio degli esemplari, i primi a poterci raccontare “di prima mano” la loro storia⁴².

⁴⁰ Annunciato nel 2015 e concretizzatosi nel 2018 in un database pubblicato online, il PDA è raggiungibile all'indirizzo: <http://arkyves.org/r/section/him_CERLPDA/>. Per degli approfondimenti su di esso si veda: Marieke van Delft, *CERL's Work and Vision for Provenance Research II: The Provenance Digital Archive in CERL*, «La Bibliofila», 117 (2015), n. 3, p. 321-324; Marieke van Delft, *A New Tool for Describing Provenance Images CERL's Provenance Digital Archive*, in: *Printing R-Evolution and Society 1450-1500. Fifty Years that Changed Europe*, edited by C. Dondi, Venezia: Edizioni Ca' Foscari - Digital Publishing, 2020, p. 909-920 e Lucrezia Signorello, *Il Provenance Digital Archive del CERL: il nuovo censimento online delle provenienze librarie*, «DigItalia. Rivista del digitale nei beni culturali», 1 (2020), p. 133-134.

⁴¹ Cfr. Alfredo Serrai, *Storia delle biblioteche. Parte I*, «Bibliothecae.it», 3 (2014), n. 1, p. 19-35: 19.

⁴² Sebbene, naturalmente, non ci si possa fermare a quanto desumibile da tali rilevazioni, ma i dati emersi vadano arricchiti e incrociati con le informazioni desunte da un'indagine della biblioteca condotta “dietro le quinte” attraverso lo studio della documentazione archivistica relativa [...] e poi intrecciata ad altri documenti, per esempio i cataloghi, gli inventari, i carteggi», cfr. Simona Insera, *Prefazione*, in: *Per libri e per scritture. Contributi alla storia del libro e delle biblioteche nell'Italia meridionale tra XVI e XVIII secolo*, a cura di S. Insera, Milano: Ledizioni, 2018, p. 7-12: 11.

Si chiude questa breve e sommaria rassegna di strumenti in rete dedicati all'identificazione e al censimento dei segni di provenienza con alcune parole – che si vorrebbero qui di stimolo alla ricerca – tratte dal portale, al quale si è accennato poc'anzi, del Provenance Online Project della Penn University: «Bookplates, inscriptions, stamps, bindings, binding waste and shelfmarks... they're all here, and they all need to be identified!»⁴³.

In recent decades there has been growing interest in the analysis of provenance marks as a privileged source for the study and reconstruction of book collections, as well as reading practices and the circulation of manuscripts and printed books. There are therefore numerous census projects and descriptions of these signs of provenance carried out by various institutions, both nationally and internationally. The contribution aims to briefly recall the theoretical and methodological assumptions underlying these initiatives and to offer a summary but illustrative overview of some tools made available on the net in this area of research.

⁴³ Così inizia, infatti, il testo di presentazione del progetto curato dalla Penn University of Pennsylvania: <<https://www.flickr.com/photos/58558794@N07/>>.

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2020

IDP - Illuminated Dante Project: un archivio e database per la più antica iconografia dantesca (secc. XIV-XV)

«DigItalia» 2-2020
DOI: 10.36181/digitalia-00022

Ciro Perna

Università degli Studi della Campania "Luigi Vanvitelli"

Nato nel 2015 e legato al Dipartimento di Studi Umanistici dell'Università di Napoli "Federico II" e al Dipartimento di Lettere e Beni Culturali dell'Università della Campania "Luigi Vanvitelli", nonché a diversi Enti e Istituzioni nazionali e internazionali, IDP-Illuminated Dante Project mira alla creazione di un archivio di codici danteschi miniati tra Trecento e Quattrocento, con relativo database open access, che potrà costituire un utile strumento di comprensione dei complessi meccanismi che regolano il rapporto tra versi danteschi e iconografia. Il contributo presenta la struttura del database IDP a partire dalle articolazioni delle griglie in ambiente back-end.

1. Il progetto IDP: un excursus su fase 1 (2015-2018) e fase 2 (2019-2021)

«La pratica del tradurre in immagini vive la Commedia è di lunga, lunghissima durata [...]. Di questa tradizione, documentata in vario modo [...], non esiste un censimento sistematico: dato deprecabile, perché il censimento di questi materiali potrebbe concorrere a ricostruire la storia della ricezione della Commedia, visto che essi testimoniano non solo il successo dell'opera nel tempo e nello spazio, ma anche le pressoché infinite modalità di lettura e di utilizzazione che di essa si sono date nel corso dei secoli e nelle varie culture [...]»¹.

È dalla piena consapevolezza di questa «deprecabile» mancanza, già denunciata alla comunità di studiosi e appassionati della *Commedia* da Lucia Battaglia Ricci oramai più di un decennio fa, che nel 2015, presso il Dipartimento di Studi Umanistici dell'Università di Napoli "Federico II", è nato, con un importante finanziamento erogato dalla Compagnia di San Paolo (Programma STAR-linea 1 2014), IDP-Illuminated Dante Project, con l'obiettivo di allestire, in prospettiva del settecentenario della morte di Dante (2021), un archivio online, con relativo database open access codicologico e iconografico, di tutti i testimoni della *Commedia*, cor-

¹ Lucia Battaglia Ricci, *Ai margini del testo: considerazioni sulla tradizione del Dante illustrato*, «Italianistica», 38 (2009), n. 2, p. 39-58: 39-40..

redati da un apparato illustrativo che intrattenga relazioni con il testo del poema². Il lavoro dell'équipe IDP nella fase 1 del progetto (2015-2018) è stato articolato su 4 fronti:

- 1) censimento e descrizione degli items;
- 2) avvio della campagna di digitalizzazione;
- 3) relazioni con Enti terzi;
- 4) strutturazione e creazione del database.

Dal 2019 è stata avviata la fase 2, a seguito di un altro importante finanziamento, ottenuto presso il Dipartimento di Lettere e Beni Culturali dell'Università degli Studi della Campania "Luigi Vanvitelli" (Progetto V:ALERE 2019-2021), che, oltre all'implementazione e al pieno raggiungimento dei desiderata di fase 1, è e sarà focalizzata sullo sviluppo web del "prodotto" IDP, che mira a ottenere la piena interoperabilità delle immagini.

Il progetto ha costituito un corpus di 283 manoscritti datati o databili tra XIV e XV secolo e conservati in Biblioteche, Musei, Archivi o Enti nazionali e internazionali, pubblici e privati, di cui ad oggi più della metà sono stati oggetto di visione autotopica da parte di membri del team e di analitica descrizione codicologica, paleografica, storico-artistica. La fase 1 del progetto ha visto altresì concludersi una prima considerevole campagna di digitalizzazione e relativa metadattazione, condotta dalla Space SpA, che ha riguardato circa 100 manoscritti, resa possibile grazie ad un'importante convenzione tra IDP e la Direzione Generale Biblioteche e Istituti Culturali del Ministero dei Beni e delle Attività Culturali e del Turismo (2016), recentemente rinnovata per il prossimo triennio, che concede il permesso di digitalizzare e riprodurre online in alta definizione e in conformità con i protocolli dell'International Image Interoperability Framework (IIIF)³, in cui IDP è coinvolto, tutti i codici conservati nelle Biblioteche Statali italiane (circa la metà del corpus). Ulteriori concessioni e accordi di partenariato scientifico con i maggiori enti conservatori internazionali legati al progetto (Bibliothèque nationale de France; Oxford Bodleian Library; British Library) e con altre istituzioni nazionali non statali (Archivio storico civico e Biblioteca Trivulziana; Biblioteca dell'Accademia dei Lincei e Corsiniana) contribuiscono a fare di IDP non solo il più grande archivio digitale di codici miniati della *Commedia*, ma – allo stato attuale – il più grande archivio in assoluto di codici danteschi, da offrire in libero accesso a specialisti e lettori appassionati.

La Convenzione con la Direzione Generale del MiBACT ha aperto la strada a uno scambio privilegiato con l'Istituto centrale per il catalogo unico delle biblioteche

² Si tratta di quei manoscritti che presentano un livello iconografico «medio» ed «elevato», secondo le categorie stabilite da Marisa Boschi Rotiroti, *Codicologia trecentesca della Commedia. Entro e oltre l'antica vulgata*, Roma: Viella, 2004, p. 24.

³ Per informazioni si veda: <<https://iiif.io/community/consortium/>>.

italiane e per le informazioni bibliografiche (ICCU) e con i suoi laboratori: la digitalizzazione e metadatezione dei manoscritti è infatti gestita in collaborazione con Internet Culturale (IC), mentre le descrizioni codicologiche confluiranno nella piattaforma di Manus OnLine (MOL), all'interno della quale saranno ricercabili attraverso una sezione speciale dedicata al progetto. Le schede di descrizione IDP verranno di seguito esportate – via XML TEI-P5 – da MOL sul portale www.dante.unina.it e integrate in un database in grado di elaborare tutti i metadati concernenti lo stile e l'attribuzione delle miniature, la complessa iconografia delle illustrazioni e il rapporto tra le immagini e il testo sulle carte del codice: database che sarà qui di seguito presentato nelle sue articolazioni. Il portale IDP accoglierà, dunque, le immagini in alta definizione di tutti i manoscritti del corpus di cui verrà concesso il copyright, rendendole interoperabili nella comunità web IIF con uno specifico web app manifest e attraverso l'utilizzo del viewer Mirador.

2. Articolazioni del database: presentazione del back-end

L'inserimento di una nuova unità codicologica sul portale IDP dovrebbe, a rigore, seguire quello della stessa in MOL, alla luce dell'esportazione dei dati sopra evidenziata, che permetterebbe così allo schedatore di procedere alla compilazione delle griglie iniziali in campi prestabiliti o liberi (categoria «Generale», Fig. 1-2),

The image shows a web form for manuscript identification and description, divided into four main sections:

- IDENTIFICAZIONE:** Contains a search bar for 'Dati dell'unità codicologica in Manus On Line' with a 'Cerca in MOL' button. Below it is an information icon and a note: 'Se l'unità codicologica non è ancora presente in Manus on line, si consiglia di inserire in modo provvisorio nel campo sottostante gli elementi che permettono di identificarla.' There is a large empty text field and a row of dropdown menus for 'Rodoevino 1984', 'Bosche Rotinoni 2004', 'Censimento I', and 'Berrelli 2011'.
- DESCRIZIONE SINTETICA:** Includes a text area for 'Compari nella descrizione breve del manoscritto (nel caso di composto compilare questo campo in una sola delle unità)' with a 'OK' button.
- AREA DI PRODUZIONE:** Features dropdown menus for 'Nazione' (set to 'Italia'), 'Area' (set to '-- scegli l'area --'), and 'Regione'. Below is a 'Città' field with 'nessuna' entered and an 'Aggiungi' button. A 'Note' field is also present.
- TIPOLOGIA GRAFICA:** Contains an 'In MOL:' section with a 'Mani' field set to 'Nessuna' and an 'Aggiungi' button. A 'Note' field is also included.

Figura 1.

The image shows a web form with four main sections:

- TERZINE PER CARTA:** A text input field labeled 'Note'.
- PAGINE INCIPITARIE:** A sub-section with the instruction 'Inserire nei campi la carta della pagina incipitaria'. It contains four text input fields labeled 'INFERNO', 'PURGATORIO', 'PARADISO', and 'COMMENTO'.
- INTITOLAZIONE (RUBRICHE):** A section with two dropdown menus: '--scegli la lingua--' and '--scegli la lunghezza--'. Below are three rows:
 - Intitolazioni:** A text input field with 'Nessuna' and an 'Aggiungi' button with edit and delete icons.
 - Finì:** A text input field with 'Nessuna' and an 'Aggiungi' button with edit and delete icons.
 - Note:** A text input field with an 'Aggiungi' button.
- STEMMA:** A text input field with 'Nessuno' and an 'Aggiungi' button with edit and delete icons.

Figura 2.

integrazione, cioè, di dati sostanzialmente codicologici alla scheda MOL (segnalazione dei corrispettivi items nei repertori di riferimento⁴, area di produzione, tipologia grafica, numerazione carte incipitarie, intitolazioni, identificazione e descrizione eventuale stemma).

Compilate le griglie integrative alla descrizione del manoscritto offerta in MOL ed esportata in IDP, si configura, a partire dalla seconda sezione, la parte propriamente iconografica del database, che dalla superficie del piano illustrativo arriva in profondità sino all'esame delle singole immagini. Ancora una volta, tra griglie pre-stabilite e campi liberi, si offre allo schedatore la possibilità di descrivere il «Progetto Decorativo» inteso nella sua globalità (Fig. 3).

Di nuovo in relazione alla totalità del programma iconografico sono strutturate le sezioni della «Bibliografia», esportabile da MOL, in cui si segnaleranno tutti gli studi relativi al manufatto in oggetto, del «Facsimile/Link», ove sarà possibile fornire un collegamento ipertestuale ad un eventuale facsimile del codice, e delle «Impostazioni Globali» (Fig. 4), i cui dati vengono poi automaticamente attribuiti a tutte le singole schede di descrizione delle immagini, ovvero, nel caso non infrequente di progetti complessi in diacronia e sincronia (diverse mani, interventi gra-

⁴ Si tratta di Marcella Roddewig, *Dante Alighieri, Die gottliche Komodie: vergleichende Bestandsaufnahme der Commedia-Handschriften*, Stuttgart: Anton Hiersemann, 1984; Marisa Boschi Rotiroli, *Codicologia trecentesca della Commedia*. Cit.; *Censimento dei Commenti Danteschi. 1. I commenti di tradizione manoscritta (fino al 1480)*, a cura di E. Malato e A. Mazzucchi, Roma: Salerno Editrice, 2011; Sandro Bertelli, *La tradizione della Commedia. Dai manoscritti al testo. I codici trecenteschi entro l'antica vulgata conservati a Firenze*, Firenze: Olschki, 2011.

GENERALE
 PROGETTO DECORATIVO
 IMMAGINI
 BIBLIOGRAFIA
 FACSIMILE / LINK
 IMPOSTAZIONI GLOBALI
 CARICA IMMAGINI

Progetto decorativo unitario
 Progetto decorativo completo
 Presenza di aggiunte seriori

DESCRIZIONE GENERALE

INIZIALI ORNAMENTALI

- **calligrafiche**: a penna, eseguite dal rubricatore o dal copista
- **filigranate**: indicarne il colore e l'estensibile estensione ai margini
- **ornata foliata**: decorazione nel campo della lettera o che costituisce il corpo della lettera stessa, ricavata da elementi vegetali, anche stilizzati

FREGIO ORNAMENTALE

Presente
 Fregio foliato: composto esclusivamente da elementi vegetali

Figura 3.

ILLUSTRATORE

Anonimo

Attributo

Autore

Note

DATAZIONE

Inizio intervallo:

 Post

Fine intervallo:

 Ante

Note

ORIGINE

Nazione
 Area
 Regione

Città

Note

TECNICA ESECUTIVA

Non finito

puro contorno senza blocchi di colore

Presenza di oro

Note

Figura 4.

fici seriori)⁵, modificabili di volta in volta nelle singole schede stesse. Si tratta di informazioni fondamentali che vanno dall'attribuzione alle diverse modalità possibili di tecnica esecutiva.

Completata la descrizione del progetto iconografico nella sua globalità, lo schedatore potrà, come accennato, procedere in profondità a scandagliare le singole immagini che arricchiscono il codice, con l'obiettivo di fornire in front-end un'analitica e uniforme descrizione delle stesse. Si parta innanzitutto dalla collocazione nelle carte del manoscritto e dall'inserimento del *canvas uri* per rendere il file/immagine interoperabile in ambiente IIIF⁶.

Le due sezioni successive, «Attribuzione» e «Tecnica Esecutiva», rappresentano la sede, già anticipata, ove eventualmente modificare le informazioni generali acquisite in automatico dalla sezione delle «Impostazioni Globali», relative a illu-

Figura 5.

Figura 6.

⁵ Si può ricordare, solo a titolo esemplificativo e in ragione della sua autorità codicologica, iconografica e testuale, il manoscritto Firenze, Biblioteca Nazionale Centrale, Palatino 313, cosiddetto "Dante Poggiali", corredato da un ricco e variegato corredo illustrativo, per cui si veda almeno Alvaro Spagnesi, *Le miniature del «Dante Poggiali»*, in: *Chiose Palatine*, a cura di R. Abardo, Roma: Salerno Editrice, 2005, p. 30-55.

⁶ Percorso in back-end: «Immagini» > «Nuova» > «Generale» (Fig. 5).

stratore, datazione, ambiente di produzione e tecnica di esecuzione. La parte dedicata alla «Tipologia Decorativa» (Fig. 6), invece, è volta a identificare le caratteristiche propriamente “formali” della singola immagine, a partire dalla tipologia (iniziale, miniatura, disegno, fregio, diagramma, immagine estemporanea, richiamo figurato), a cui corrispondono sottosezioni in griglie prestabilite pure corredate da possibili note in campo libero, che attengono alle caratteristiche “fisiche” dell’immagine o alla sua dislocazione spaziale. Per l’iniziale si potrà ad esempio spuntare tra una tipologia istoriata o animata (a seconda che intrattenga rapporti specifici o meno con il testo della *Commedia*); per la miniatura, tra modalità a piena pagina, incorniciata o non incorniciata (con possibilità a latere di selezionare il tipo «Antiporta»); per il disegno sarà possibile selezionare una delle opzioni di impaginazione a scelta tra intercolumnio e margini; per il fregio, tra tipo istoriato o animato, con specificazione della dislocazione spaziale sulla carta; per il diagramma, per cui si è scelto di intendere tutto ciò che illustra la *Commedia* o l’eventuale commento in forma schematica (ovvero rappresentazione dei tre regni ultraterreni con relative ripartizioni, dei pianeti e dei cieli, visione dei personaggi in maniera puntiforme, o ruote cronologiche), tra schema *Inferno*, schema *Purgatorio*, schema *Paradiso* o altro; per le ultime due tipologie (immagine figurata e richiamo estemporaneo) si offre, infine, la possibilità di libere descrizioni nel campo «Note».

L’ultima sezione di descrizione dell’immagine, «Iconografia» (Fig. 7), si concentra sulle relazioni tra apparato illustrativo e testo della *Commedia* (o anche dell’even-

The image shows a web-based form titled "RAPPORTO IMMAGINE-TESTO". It is organized into several sections:

- Cantica:** Includes three dropdown menus: "Cantica" (with "-- scegli la cantica --"), "Canto" (with "-- scegli il canto --"), and "Versi" (with "N oppure N-N oppure N-N,N-N").
- Rubrica:** A large empty text input field.
- Glossa:** A large empty text input field.
- Livello del testo:** A dropdown menu with "-- scegli il livello --".
- Contesto interno:** A dropdown menu with "-- scegli il luogo dantesco --".
- Rapporti con la tradizione dantesca:** A text input field with a yellow "Aggiungi" button and icons to the right.
- Rapporti extra-danteschi:** A text input field with a yellow "Aggiungi" button and icons to the right.
- SOGGETTO:** A text input field containing "Nessuno", a yellow "Aggiungi" button, and icons.
- MACROSOGGETTO:** A text input field containing "Nessuno", a yellow "Aggiungi" button, and icons.
- PAROLE CHIAVE:** A text input field containing "Nessuna", a yellow "Aggiungi" button, and icons.

Figura 7.

tuale commento a margine), tra versi, miniatura e possibile esegesi: è il luogo, questo, in cui si indaga e si prova a interpretare quel dialogo fittissimo tra testo primario e interpretazione visuale e verbale, tra Autore e lettori.

Appare piuttosto evidente, in questo caso, lo sforzo di offrire ai fruitori del database una decrittazione il più possibile analitica dei rapporti testo/immagine, dalla imprescindibile informazione che lega quest'ultima ad uno specifico luogo della *Commedia* (per cui sarà possibile, inoltre, distinguere tra i due livelli diegetici fondamentali, «storie prime» vs «storie seconde», nonché spuntare un'ulteriore specificazione del contesto interno, dalla «Selva» all'«Empireo - Candida Rosa»), a relazioni possibili con testi di corredo come rubriche e glosse (che saranno integralmente trascritte negli appositi campi liberi). Non può mancare, per altro, una categorizzazione dei macrosoggetti (in Fig. 8 alcuni esempi), dei soggetti (dall'«Albero genealogico degli dei» a «Virgilio, in compagnia di Dante, ordina al minotauro di andarsene», da «Cerbero scuola un dannato» a «Virgilio mostra a Dante le anime del Limbo», ma la lunga lista già disponibile potrebbe essere facilmente implementata dalla visione di *hapax* iconografici non ancora considerati), nonché delle parole chiave (come «spiriti magni», «Antenora» e tanti altri esempi legati naturalmente all'iconografia oggetto di descrizione). Tutto ciò per permettere, a chi si approccerà al sistema, una ricerca per soggetto il più possibile esaustiva.

Nell'anno del settecentenario del 2021, dunque, sarà auspicabilmente disponibile per la comunità di studiosi e appassionati del testo dantesco, in un'ottica in

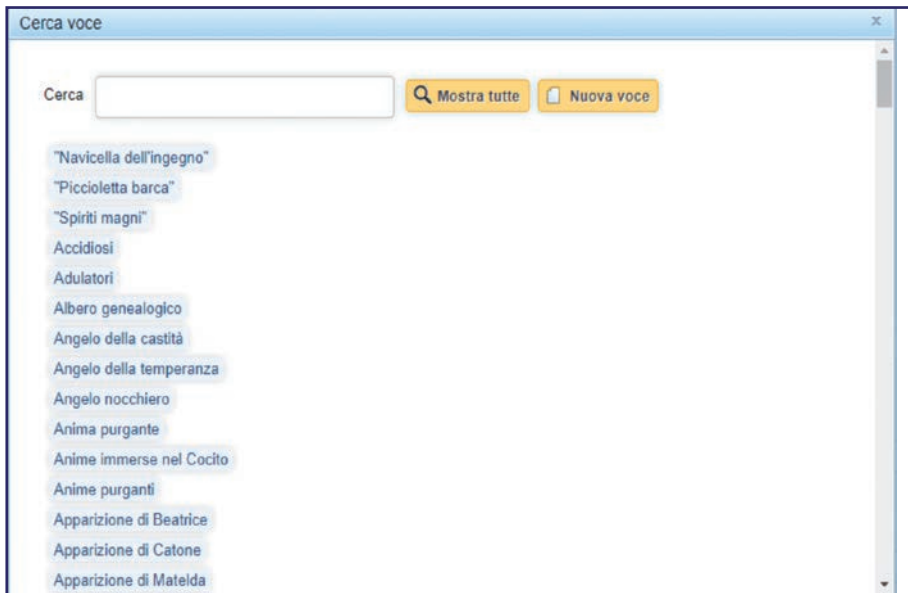


Figura 8.

cui la multidisciplinarietà degli approcci è persino scontata, il database IDP completo in tutte le sue funzioni⁷, interoperabile, open access, che potrà costituire, così come negli intenti di chi porta avanti questa sfida, un utile strumento per comprendere le complesse dinamiche di distribuzione delle «singolarità in un campo di relazioni e differenziazioni»⁸, così variegato come l'iconografia della *Commedia*.

Born in 2015 and linked to the Department of Humanities of the University of Naples "Federico II" and the Department of Humanities and Cultural Heritage of the University of Campania "Luigi Vanvitelli", as well as to various national and international Institutions, IDP -Illuminated Dante Project aims to provide an archive of Dante's illuminated manuscripts between the fourteenth and fifteenth centuries, with its open access database, which will be a useful tool for understanding the complex mechanisms that regulate the relationships between Dante's verses and iconography. This paper presents the structure of the IDP database, starting from the articulations of the back-end.

⁷ Di seguito il link ai manoscritti già disponibili: <<http://www.dante.unina.it/public/pagine/manoscritti>>.

⁸ Jérôme Baschet, *L'iconografia medievale*, Milano: Jaca Book, 2014, p. 162.

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2020

Proposta per una valutazione automatica della completeness dei metadati nel contesto delle biblioteche digitali

«Digitalia» 2-2020
DOI: 10.36181/digitalia-00023

Matteo Lorenzini - ETH Zurigo; Fondazione Bruno Kessler; Università degli Studi di Trento
Marco Rospocher - Università degli Studi di Verona
Sara Tonelli - Fondazione Bruno Kessler

Grazie all'utilizzo dei metadati è possibile accedere ad un vasto numero di risorse rese disponibili attraverso archivi e biblioteche digitali.

Normalmente i metadati sono strutturati secondo uno schema standardizzato e garantiscono l'interoperabilità e l'identificazione di un oggetto digitale facilitando l'accesso a determinati tipi di risorse. Tuttavia, una bassa qualità dei metadati e delle informazioni che questi rendono disponibili, come ad esempio la mancanza di alcuni elementi rispetto allo schema di metadati utilizzato rappresenta tutt'ora un problema comune a molti repositories.

In questo contributo presentiamo sia il nostro lavoro, allo stato attuale in corso, che ha come scopo la valutazione automatica della qualità dei metadati nelle digital libraries, che un'analisi preliminare sulla metadata completeness delle risorse presenti nell'aggregatore di metadati "CulturalItalia".

1. Introduzione

Negli ultimi anni, nel dominio dei beni culturali, abbiamo assistito ad un considerevole aumento sia delle risorse digitalizzate che, di conseguenza, ad un aumento dell'utilizzo di archivi digitali per la loro fruizione e per il loro mantenimento. Di conseguenza, i metadati sono diventati l'elemento fondamentale attraverso il quale gli utenti possono accedere ad un determinato bacino di informazioni ed aumentare la loro conoscenza riguardo ad un particolare dominio.

Tuttavia, nonostante il massiccio utilizzo dei metadati ed il loro ruolo chiave nel garantire l'accessibilità alle risorse culturali, il processo di controllo e verifica della loro qualità manca ancora di una chiara definizione. In letteratura¹, la *metadata quality* è stata presentata come un modo di misurare quanto un oggetto digitale soddisfi un determinato scopo². In tal senso, il *Metadata Quality Framework* svi-

¹ Pennok 2007.

² Secondo Pennok, nel dominio dei beni culturali sono identificabili 3 scopi principali che un oggetto digitale dovrebbe garantire: Preservation, Registering, Discovery.

luppato da Bruce e Hillmann³ è considerato il cardine sul quale poggia la discussione relativa alla *metadata quality*.

Per valutare la qualità dei metadati il *framework* propone sette metriche qualitative: *Completeness, Accuracy, Conformance, Logical Consistency and Coherence, Accessibility, Timeliness, Provenance*.

Sebbene non vi siano casi d'uso in cui queste metriche siano state applicate nell'analisi di un archivio, sarebbero sicuramente di aiuto nel caso di una valutazione sistematica sui metadati applicandole, per esempio, ad una digital library come Europea⁴.

Facendo riferimento al quadro generale poche sono le casistiche di controllo automatico dei metadati⁵. Possiamo citare il lavoro fatto da Ochoa e Duvall⁶ dove, per ciascuna delle sette metriche del *Metadata Quality Framework*, viene ipotizzata una funzione in modo da rendere la valutazione quantitativa e non più qualitativa ed il metodo ipotizzato da Peter Kiraly⁷ e parzialmente applicato alla digital library Europea.

In entrambi i casi non vengono considerati tre importanti fattori:

1. Il processo di creazione dei metadati: spesso svolto manualmente da operatori specializzati seguendo le linee guida messe a disposizione dall'archivio che poi provvederà all'*ingestion*. Per cui, le definizioni dei *metadata elements* usati per la descrizione dell'oggetto potrebbero essere interpretate in maniera diversa a seconda del punto di vista dell'operatore.

2. Processo di aggregazione: il processo di aggregazione di risorse digitali viene fatto quando dei metadati messi a disposizione da uno o più providers sono acquisiti in un unico *repository*. Poiché l'aggregazione dei metadati, nel dominio dei beni culturali, è cresciuta esponenzialmente durante gli ultimi anni, il processo di controllo dei metadati dovrebbe essere ricontestualizzato nell'ottica di verificare informazioni non omogenee tra di loro e su larga scala.

3. Contesto: l'utilizzo dei metadati per descrivere un oggetto varia a seconda del dominio o del contesto in cui viene descritto l'oggetto. Per esempio, un archeologo e un filologo hanno diverse percezioni di un'epigrafe: per il primo, un'epigrafe documenta un reperto archeologico, mentre per il secondo l'epigrafe rappresenta un testo da analizzare.

Anche la definizione di metadato definita da NISO⁸ sottolinea l'importanza del contesto descrivendo il metadato come un'entità non statica: hanno diverse interpretazioni e dovrebbero essere considerate in base al loro contesto.

Gli esistenti lavori sulla *metadata quality* presentano inoltre, secondo il nostro pa-

³ Bruce 2004.

⁴ <https://www.europeana.eu/en>.

⁵ Ochoa – Duvall 2009, Ostojic – Sugimoto – Durco 2017.

⁶ Ochoa – Duvall 2009.

⁷ Kiraly 2015.

⁸ <https://www.niso.org>.

rere, alcune limitazioni per quanto riguarda la granularità dell'analisi in quanto si concentrano o su una specifica metrica oppure focalizzano l'analisi relativamente ad un unico *repository*. Per valutare la *completeness* di un *repository* sono stati presentati quindi due diverse tipologie di approccio:

- La presenza o assenza dei *metadata elements*: affrontata come un problema binario assegnando quindi valore 0 o 1 a seconda della presenza o assenza di un particolare *metadata element*.
- Per ciascun *metadata element* viene attribuito un valore/peso arbitrario a seconda dell'importanza rispetto al profilo di metadati utilizzato.

Questi approcci, anche se consentono al curatore di misurare la qualità del singolo record o, più generalmente, di un intero *dataset*, non considerano se una bassa qualità della risorsa o del *dataset* possa dipendere dalla mancanza di pochi *metadata element* con uno *scoring* alto o dalla mancanza di molti *metadata element* con uno *scoring* basso.

La nostra soluzione che andiamo a proporre si basa sull'assunzione che queste metriche debbano permettere ai curatori di definire in un modo flessibile quali *metadata element* possano essere considerati come più interessanti nella valutazione complessiva della *completeness*, al fine di omogeneizzare questa metodologia di valutazione tra diversi *repositories* e per permettere una più dettagliata analisi dei *metadata elements*.

Questo rappresenta il maggior contributo di questo articolo: proponiamo un modo flessibile per valutare la *completeness* considerando sia i metadati obbligatori che quelli facoltativi di un particolare *metadata schema* così come il dominio di una collezione digitale.

Presentiamo inoltre una valutazione della *metadata completeness* prendendo come caso d'uso il *repository* di metadati *CulturaItalia*⁹ evidenziando come, con l'aiuto di una rappresentazione grafica, il nostro approccio possa supportare gli esperti per risolvere il problema della valutazione *metadata quality*.

2. La metodologia proposta

Il nostro obiettivo a lungo termine è quello di sviluppare un *framework* che automaticamente verifichi la qualità dei metadati di un *repository* considerando diverse metriche. Lo sviluppo di questo *framework* è legato allo svolgimento delle seguenti attività:

- Definizione delle metriche di valutazione, catturando lo stato dei *metadati* sia a livello del singolo oggetto digitale (per esempio controllare la qualità dei metadati su un singolo record in un *dataset*) che a livello dell'intero *repository*.

⁹ <http://www.culturaitalia.it>.

- Definizione di algoritmi per calcolare le metriche sopra descritte e, possibilmente, restituire suggerimenti su come riparare la bassa qualità dei metadati.

In questo articolo presentiamo i primi risultati relativi alla *completeness*.

In termini generali, la *completeness* viene calcolata come il rapporto degli elementi usati per descrivere l'oggetto rispetto al profilo di metadati utilizzato. In questa valutazione, diverse variabili dovrebbero essere prese in considerazione: gli elementi che sono obbligatori e quelli che sono opzionali, il contesto ed il dominio della collezione, così come le preferenze dei curatori quando viene valutata la *completeness*.

La metodologia che proponiamo viene strutturata come segue:

- Dato un *repository* che deve essere valutato, i *metadata elements* sono divisi in gruppi, rappresentanti la loro importanza (per esempio elementi obbligatori / raccomandati / opzionali)
- Per ciascun oggetto *O* contenuto nel *repository*, la *completeness score* $Cg(O)$ viene calcolata separatamente per ciascuno dei gruppi (*G*) definiti: il numero dei *metadata element* utilizzati per descrivere ciascun oggetto, viene diviso per il numero totale dei *metadata element* presenti nel gruppo di cui andiamo a calcolare la *completeness*. Per esempio, se un oggetto viene descritto utilizzando 3 dei 10 elementi obbligatori, il valore della *completeness* sarà 0.3. Lo score della *completeness* sarà dunque un numero reale in una scala da 0 a 1. Più il valore è vicino ad 1, più completa è la descrizione per quello specifico gruppo di metadati.
- La *completeness score* (una per ogni gruppo di metadati) viene calcolata relativamente ad ogni singolo record. Per valutare la completezza generale del *dataset*, per ciascun gruppo di metadati viene definito un grafico separato dove sull'asse X, utilizzando 10 intervalli, viene rappresentato lo score *range* relativo alla *completeness* (0-0.1;0.1-0.2;0.2-0.3;..0.9-1.0) mentre sull'asse Y viene rappresentata la percentuale degli oggetti relativi ad ogni *completeness score*.

La valutazione della *metadata quality* diversificata per gruppi di importanza e di dominio offre al *metadata curator* la possibilità di avere un quadro completo riguardo allo stato generale del dataset o *repository*. Per cui, il curatore, può intervenire sugli oggetti con una bassa qualità dei metadati valutando i diversi fattori che hanno un impatto sul risultato finale.

3. Caso d'uso: Culturalitalia

Culturalitalia è un *repository* di metadati nel dominio dei beni culturali e rappresenta uno degli aggregatori nazionali della digital library europea Europea.

Gestito dall'ICCU, accoglie circa 4.500.000 risorse che includono immagini, contenuti audio visivi e testuali. I metadati sono resi disponibili attraverso la piattaforma dati.culturalitalia.it¹⁰ utilizzando lo schema di metadati PICO¹¹, caratterizzato da 94 elementi allineati con Dublin Core. Gli elementi PICO sono stati suddivisi in Obbligatori (8 elementi), raccomandati (10 elementi) e opzionali (76 elementi).

Per questo articolo sono stati presi come casi d'uso due *datasets*: MuselD-Italia (76.828 record) e Regione Marche (90.602 record). Le risorse dei due *datasets* sono per lo più Opere d'Arte Visiva.

Seguendo la metodologia presentata nel paragrafo uno, la suddivisione in gruppi dei metadati è stata fatta prendendo come riferimento le linee guida tecniche¹² ed i documenti di *mapping* resi disponibili attraverso la documentazione tecnica del portale. Sono stati quindi definiti i seguenti 4 gruppi:

- Elementi obbligatori (8 elementi): gli elementi obbligatori del PICO *profile*, *dc:title*, *dc:identifier*, *dc:subject*, *dc:type*, *pico:preview*, *dc:isReferencedBy*, *dcterms:license*, *pico:licenseMetadata*.
- Raccomandati (10 elementi): gli elementi raccomandati del PICO *profile* come ad esempio *dc:description*, *pico:author*, *dcterms:spatial*.
- Di dominio: gli elementi opzionali del PICO *profile* che sono però rilevanti per uno specifico dominio. Nel caso di Opere d'arte visiva abbiamo identificato i seguenti 11 elementi: *pico:commissioner*, *pico:matherialAndTechnique*, *dcterms:created*, *dcterms:isPartOf*, *dcterms:alternative*, *dcterms:modified*, *dc:contributor*, *dc:coverage*, *dcterms:bibliographicCitation*, *pico:printer*, *dcterms:replaces*.
- Opzionali: (76 elementi): i rimanenti elementi opzionali del PICO *profile* come per esempio *dcterms:bibliographicCitation*, *pico:commissioner*, *pico:performer* ecc.

3.1 Risultati

Prima di analizzare nel dettaglio la *completeness* dei *datasets* in esame, utilizzando la suddivisione in gruppi dei metadati, abbiamo analizzato la frequenza generale di utilizzo degli elementi PICO nelle risorse presenti in MuselD-Italia e Regione Marche. Le figure 1 e 2 rappresentano la percentuale dei *records* nei due *datasets*.

¹⁰ <http://dati.culturalitalia.it>.

¹¹ <http://www.culturalitalia.it/opencms/export/sites/culturalitalia/attachments/documenti/picoap/picoap1.0.xml>.

¹² http://www.culturalitalia.it/opencms/documentazione_tecnica_it.jsp.

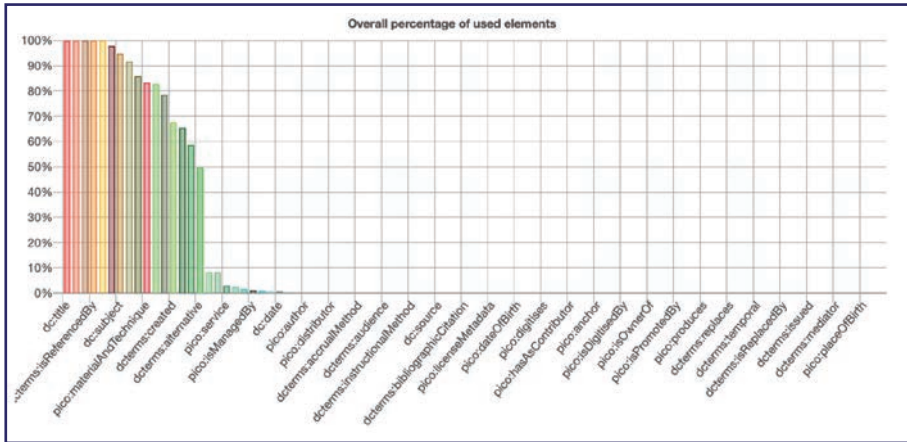


Figura 1. Percentuale degli elementi PICO per record nel dataset MuseiD-Italia

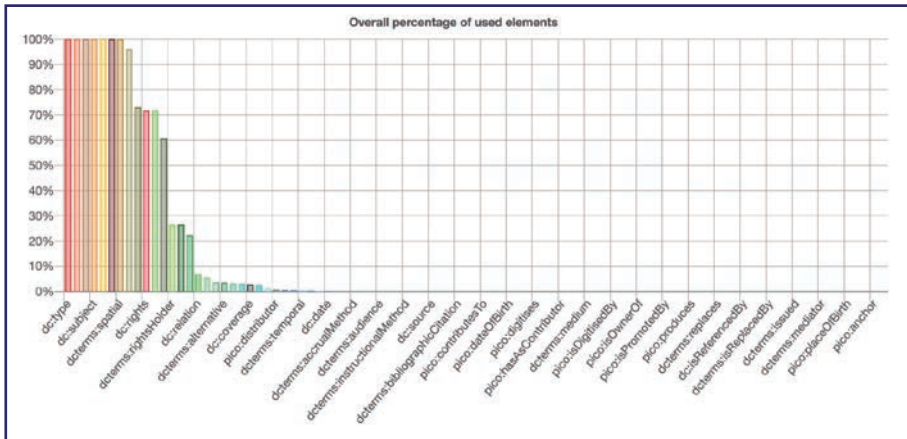


Figura 2. Percentuale degli elementi PICO per record nel dataset Regione Marche

Possiamo notare che molti casi (ma non tutti) gli elementi obbligatori sono valorizzati in tutti i record del *dataset*. Diverso è invece il quadro ottenuto dalla verifica degli elementi che caratterizzano i singoli gruppi dove la percentuale degli elementi valorizzati è sostanzialmente più bassa.

Successivamente, utilizzando il raggruppamento definito nella sezione 3, abbiamo analizzato, per ciascun record del *dataset*, 4 diverse *completeness* (una per ciascun gruppo di metadati) e, separatamente, per ciascun gruppo di metadati abbiamo analizzato la distribuzione del grado di *completeness* attraverso un istogramma come descritto nella sezione 2, figure 3 e 4.

I grafici dimostrano che i risultati ottenuti per la *completeness* nei due *datasets* sono generalmente bassi. Per esempio, nel dataset MuseiD-Italia (Fig. 3 in alto a sinistra), la maggior parte dei *records* (~92%) ha un valore di *completeness* per i

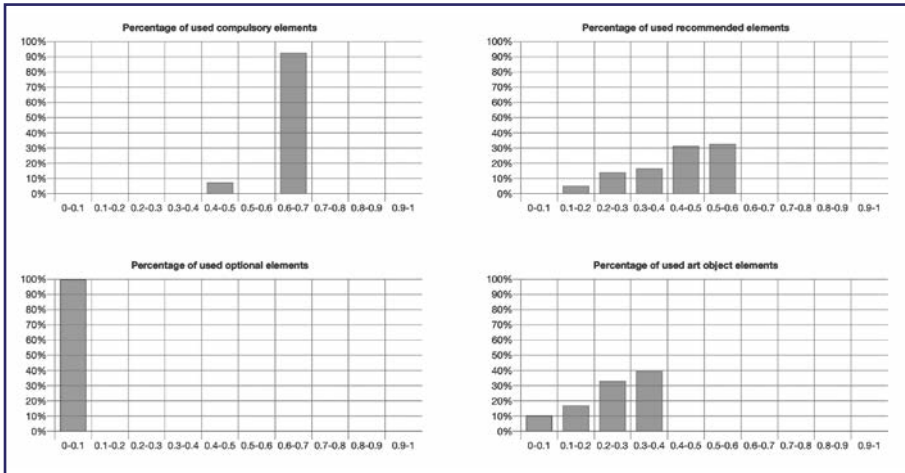


Figura 3. *Completeness per il dataset MuseiD-Italia*

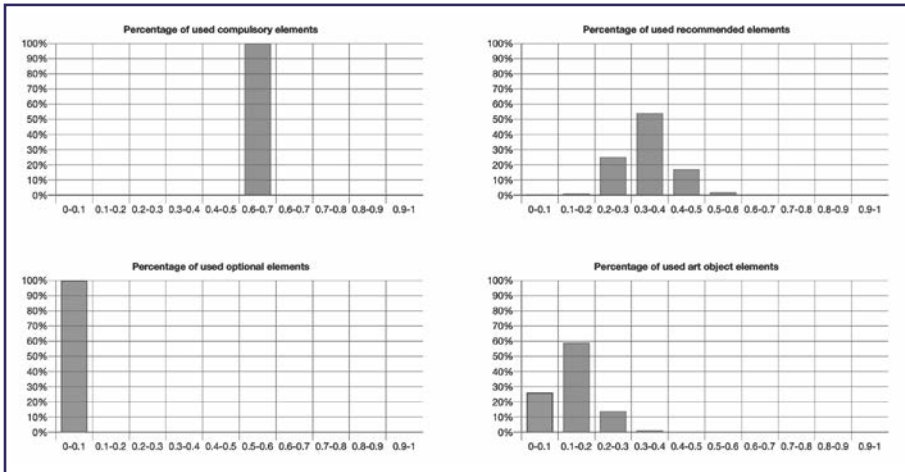


Figura 4. *Completeness per il dataset Regione Marche*

campi obbligatori in un range tra 0.6 e 0.7, nel caso della Regione Marche (Fig. 4 in alto a sinistra) tutti i record raggiungono uno score per gli stessi gruppi di metadati tra 0.5 e 0.6. Lo stesso può essere osservato nella valutazione sul dominio specifico (tra 0.1 e 0.4). I record del *dataset* quindi raramente usano gli elementi opzionali, mentre l'utilizzo degli elementi raccomandati e specifici varia rispetto al dominio.

Abbiamo sottolineato che, grazie ad un'analisi basata su 4 diversi raggruppamenti di metadati, quello che viene restituito al curatore è un'analisi precisa e comprensiva relativa al grado di *completeness* dell'intero *dataset* permettendone anche un confronto rispetto ai 4 raggruppamenti che sono stati definiti.

4. Conclusioni

Applicando concretamente il metodo proposto in due *datasets* come MusEID-Italia e Regione Marche, abbiamo introdotto un metodo innovativo e dettagliato di analizzare i *metadata element* sulla base di diversi raggruppamenti considerando la loro rilevanza rispetto al *dataset*.

In questo modo, il curatore dei metadati può identificare in maniera immediata quali siano i problemi all'interno del *repository* ottimizzando il processo di verifica.

Metadata allow the access to a wide variety of cultural heritage resources made available through repositories, digital libraries and catalogues. Usually taking the form of a structured set of descriptive elements, metadata assist in the identification, location, processing, tracking, preserving, sharing and retrieval of information, while facilitating content and access management. However, low metadata quality, such as the lack of mandatory information, incorrect information or inconsistency, is still an open issue in many repositories. In this paper we present our ongoing work aiming at automatizing the metadata quality analysis, and the preliminary results on metadata completeness for the metadata national aggregator CulturalItalia.

L'ultima consultazione dei siti web è avvenuta nel mese di dicembre 2020

RIFERIMENTI BIBLIOGRAFICI

- Bruce - Hillmann 2004 Thomas R. Bruce – Diane I. Hillmann. *The continuum of metadata quality: defining, expressing, exploiting*. In: *Metadata in practice*, ed. by D. I. Hillmann, E. L. Westbrook. Chicago: American Library association, 2004, p. 238-256
- Buonazia – Masci 2007 Irene Buonazia – Maria Emilia Masci. *Il pico application profile. Un dublin core application profile per il portale della cultura italiana*. Intervento tenuto al Seminario nazionale di studi *Interoperabilità di contenuti e servizi digitali: metadati, standard e linee guida*. Roma, 3 aprile 2007
- Buonazia – Masci – Merlitti 2007 Irene Buonazia – Maria Emilia Masci – Davide Merlitti. *The project of the Italian culture portal and its development. A case study: Designing a Dublin core application profile for interoperability and open distributions of cultural contents*. In: *Openness in Digital Publishing: Awareness, Discovery and Access - Proceedings of the 11th International Conference on Electronic Publishing held in Vienna - ELPUB 2007, Vienna, Austria, June 13-15, 2007*. ELPUB, 2007, p. 393-404
- Király 2015 Péter Király. *A metadata quality assurance framework*. 2015. <https://pkiraly.github.io/metadata-quality-project-plan.pdf>
- Király – Büchler 2018 Péter Király – Marco Büchler. *Measuring completeness as metadata quality metric in Europeana*. In: IEEE International Conference on Big Data (Big Data), 2018, p. 2711-2720.
- Neumaier – Umbrich – Polleres 2016 Sebastian Neumaier – Jürgen Umbrich – Axel Polleres. *Automated quality assessment of metadata across open data portals*. «Journal of Data and Information Quality (JDIQ)», 1 (2016), p. 1-29.
- Ochoa – Duvall 2009 Xavier Ochoa – Erik Duvall. *Automatic evaluation of metadata quality in digital repositories*. «International journal on digital libraries» 10 (2009), n. 2-3, p. 67-91
- Ostojic – Sugimoto – Durco 2017 Davor Ostojic – Go Sugimoto – Matej Durco. *The curation module and statistical analysis on VLO metadata quality*. Linköping Electronic Conference Proceedings, 2017, p. 90-101. <<https://ep.liu.se/ecp/136/007/ecp17136007.pdf>>
- Pennok 2007 Maureen Pennok. *Digital curation: a life-cycle approach to managing and preserving usable digital information*. «Library & Archives», (2007), p. 34-45.
- Tani – Candela – Castelli 2013 Alice Tani – Leonardo Candela – Donatella Castelli. *Dealing with metadata quality: The legacy of digital library efforts*. «Information Processing & Management», 6 (2013), p. 1194-1205.
- Valentine et al. 2017 Charles Valentine – Juliane Stiller – Péter Király – Werner Bailer – Nuno Freire. *Data Quality Assessment in Europeana: Metrics for Multilinguality*. In TDDL/MDQual/Futurity, 2017. <<http://ceur-ws.org/Vol-2038/paper6.pdf>>.



Documenti e discussioni

L'evoluzione del ruolo delle Biblioteche Digitali Scientifiche (BDS)

Costantino Thanos

Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo"

La Rete di Eccellenza DELOS (2004 – 2008), finanziata dal Sesto Programma Quadro (6FP) della Commissione Europea, nell'arco della sua vita ha sviluppato due definizioni/visioni per la Biblioteca Digitale Scientifica (BDS):

Prima definizione/visione (San Cassiano Workshop 2001):

«Le Biblioteche Digitali dovrebbero consentire a qualsiasi cittadino di accedere alla conoscenza umana in qualunque momento e da qualunque luogo in un modo amichevole, multimodale, efficiente ed efficace superando barriere linguistiche, culturali, e di distanza ed utilizzando dispositivi interconnessi all'Internet».

Da notare che in questa Visione l'enfasi viene posta sull'"accesso" che a quel tempo era il principale obiettivo dell'internet e del web.

Seconda definizione/visione (Corvara, Workshop 2004):

«Le Biblioteche Digitali hanno il potenziale di diventare i depositi della conoscenza universale e i condotti di comunicazione per il futuro, un normale veicolo attraverso il quale tutti potranno accedere, discutere e accrescere/migliorare informazioni di tutte le forme».

Questa seconda visione tiene conto della rapida evoluzione dell'internet e del web da piattaforma di pubblicazione, dove l'aspetto im-

portante era l'accesso all'informazione, a piattaforma di "social media", dove l'aspetto importante è la creazione e lo scambio d'informazioni e la collaborazione.

In tutte e due le definizioni/visioni era implicito che il record scientifico era composto, fondamentalmente, dalla letteratura scientifica, dalla letteratura grigia e dai giornali scientifici. Essenzialmente, il record scientifico era di natura testuale.

Negli ultimi anni, significativi progressi tecnologici hanno modificato radicalmente la composizione del record scientifico. Alludiamo a progressi tecnologici che hanno permesso la raccolta (attraverso osservazioni), la produzione (attraverso esperimenti), e il calcolo (attraverso simulazioni) di enormi volumi di dati scientifici noti come big data. Altri progressi tecnologici hanno permesso la digitalizzazione di questi dati su larga scala. Di conseguenza, il record scientifico moderno è composto, fondamentalmente, da un gran numero di collezioni di dati scientifici (datasets) distribuite su scala globale e contenenti enormi volumi di dati dell'ordine di terabytes. Dati che presentano alcune importanti caratteristiche, per esempio dinamicità, grande dimensionalità e connettività. Queste collezioni di dati sono poco connesse fra loro e, di conseguenza, anche il record scientifico risulta poco connesso.

Le conseguenze di questa radicale modifica del record scientifico sono due, una di natura culturale e l'altra di natura tecnologica. La prima, di natura culturale, riguarda l'emergere

di un nuovo paradigma del metodo scientifico dovuto al fatto che gran parte della conoscenza scientifica prodotta dai ricercatori si è trasferita dalla letteratura ai dati. Di conseguenza, è nei dati che va cercata la conoscenza scientifica prodotta da osservazioni/esperimenti/simulazioni condotte dai ricercatori. Il nuovo paradigma, noto come data science, è basato sull'intenso utilizzo dei dati scientifici. La seconda conseguenza, di natura tecnologica, riguarda l'inadeguatezza di alcune tecnologie tradizionali riguardanti, principalmente, la ricerca (query processing) e gestione dei dati (data management) nel nuovo contesto tecnologico creato dai big data. Questa inadeguatezza ha motivato lo sviluppo di nuove tecnologie nella ricerca dei dati (data seeking) e nella loro interconnessione (data linking). Inoltre, nuovi approcci alla creazione di conoscenza stanno emergendo basati su tecnologie che supportano l'esplorazione di spazi dati interconnessi.

Infatti, il nuovo paradigma di metodo scientifico imporrebbe ai ricercatori/studiosi di esplorare il record scientifico creando percorsi di conoscenza (knowledge patterns) che permetterebbero loro di acquisire nuove conoscenze e/o formulare nuove ipotesi circa un fenomeno sotto studio. Questa esigenza, però, si scontra con il fatto che il record scientifico è poco connesso.

In questo emergente contesto culturale-tecnologico il ruolo della BDS è destinato ad un radicale cambiamento. In sostanza, la BDS non è più soltanto un deposito della conoscenza scientifica contenuta nella letteratura, ma viene ad assumere anche la funzione di una infrastruttura che supporta i ricercatori/studiosi nelle loro attività richieste dal nuovo paradigma di metodo scientifico. Quindi, la BDS deve supportare i ricercatori/studiosi nella loro attività che mira a scoprire e ad accedere alla conoscenza confinata nei datasets che si custodiscono in depositi disciplinari (domain-specific data centers) distribuiti su scala globale. La

BDS deve fornire servizi che permettono ai ricercatori/studiosi di scoprire esistenti relazioni fra dataset globalmente distribuiti e stabilire legami (links) fra questi creando così grafi di conoscenza, dove i nodi dei grafi sono dataset e gli archi rappresentano relazioni (semantiche/temporali/spaziali/causali) fra loro. La moderna BDS, in sintesi, deve fornire i strumenti che permettano ad un ricercatore/studiose sia di creare percorsi di conoscenza, sia di navigare in uno spazio dati interconnesso.

Una funzione molto importante, che in questo suo nuovo ruolo la BDS deve fornire, è quella della Pubblicazione dei Dati. Per pubblicazione dei dati si intende un processo che ha come obiettivo quello di permettere ai ricercatori/studiosi di scoprire, accedere, capire, riutilizzare e riprodurre i dati scientifici contenuti nelle varie collezioni dati distribuite su scala globale. Fasi importanti di questo processo sono la registrazione dei dati ed il loro potenziamento semantico.

Il nuovo ruolo della BDS, sopra tratteggiato, è destinato nel prossimo futuro a diventare prevalente rispetto al suo ruolo tradizionale, quello che la considera come un deposito di letteratura scientifica.

Un'ultima fondamentale considerazione riguarda il contesto politico all'interno del quale il nuovo ruolo della BDS sta emergendo. Questo contesto consiste nella politica di Scienza Aperta (Open Science). Per Scienza Aperta si intende una politica che prescrive la condivisione aperta di tutti i risultati della ricerca al più presto possibile compatibilmente con i tempi del processo della indagine/ricerca.

Tale politica è stata promulgata dalla Commissione dell'Unione Europea ed è diventata la politica che regola tutte le attività di ricerca condotte con finanziamenti Europei.

È evidente che il nuovo ruolo della BDS, sopra delineato, definisce anche un nuovo settore di ricerca. Infatti, affinché una BDS possa con efficacia svolgere il suo nuovo ruolo è necessario lo sviluppo di importanti attività di ricer-

ca. Qui di seguito elenchiamo una lista, ovviamente non esaustiva, di attività di ricerca necessarie, al nostro avviso, per lo sviluppo della BDS della prossima generazione.

- Per la creazione di un record scientifico interconnesso è fondamentale una ricerca avente come obiettivo l'allargamento del numero di tipi di relazione fra dataset (semantica, causale, temporale, spaziale ecc.) e di "assiomatizzarle" mediante anche nuove ontologie, specifiche per ogni disciplina scientifica. È, quindi, necessario anche lo sviluppo di motori di inferenza in grado di inferire relazioni di vario tipo fra dataset.
- Indirizzare la ricerca su tematiche riguardanti la citazione ed il collegamento di dati scientifici dinamici, in quanto una importante caratteristica dei big data è la loro dinamicità. Una tecnologia di citazione dinamica per i datasets scientifici è di fondamentale importanza per l'accesso alla conoscenza che si trova confinata nei big data.
- Sviluppare nuovi paradigmi di ricerca di dati (data seeking) e di creazione di cono-

scenza (knowledge creation) basati su approcci esplorativi che sfruttano anche la tecnologia di *machine learning* in quando i tradizionali approcci (query processing e data mining) sono inadeguati nel contesto dei big data.

- Indirizzare la ricerca verso la definizione di metadati, specifici per ogni disciplina, semanticamente potenziati.
- Indirizzare la ricerca su tematiche riguardanti la gestione (management) e attraversamento di grafi (graph data).
- Sviluppare ricerche verso tecniche di *data assimilation* e *correlation*.
- Sviluppare nuovi modelli di pubblicazione.

Ovviamente, come già accennato, questa non è una lista esaustiva ed in particolare gli aspetti multimediali dei big data meritano attento studio.

Una Conferenza sulle Biblioteche Digitali Scientifiche può costituire un forum dove queste tematiche vengono dibattute e dove relativi risultati vengono presentati.

33 million objects and counting: a conversation about the Deutsche Digitale Bibliothek

Cosmina Berta - Gerke Dunkhase
Deutsche Digitale Bibliothek (DDB)

Since 2014, the year when the Deutsche Digitale Bibliothek (DDB) was officially launched, the number of digital objects offered on www.deutsche-digitale-bibliothek.de grew from 8 million records to 33 million.

We would like to talk a bit about the development of the DDB, its structure, and what is next in store. Cosmina Berta (Servicestelle and Data Management) and Gerke Dunkhase (Head of Technology, Development and Service) answer Digitalia's questions.

Digitalia: What is the goal of the Deutsche Digitale Bibliothek?

Cosmina Berta: The goal of the Deutsche Digitale Bibliothek (DDB) is to offer everyone unrestricted access to Germany's cultural and scientific heritage, to make it possible for anyone to access its unique cultural heritage and knowledge conveniently via one central point of entry. That means millions of books, archival items, images, music, films and scores, from all over Germany.

In the present day, more than ever, having this entire wealth of information at our fingertips, centrally accessible from anywhere in the world, independent of time and location, is a fundamental improvement of our opportunities regarding research, education, use of creativity. It is also a wonderful possibility to stay connected with the world of cultural heritage. This vision is continuously developing, and we have some way to go until it is realized in its full potential.

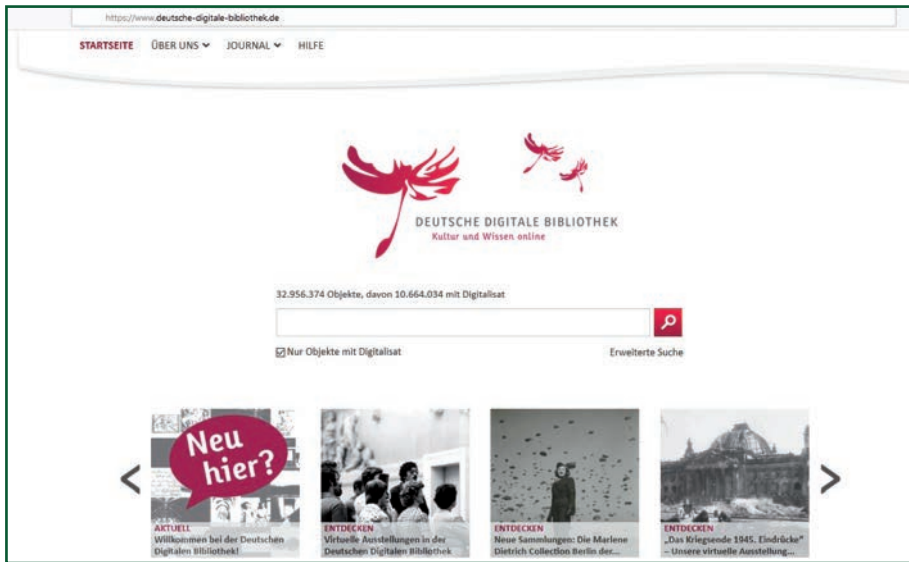
Digitalia: How is the network built?

Cosmina Berta: Culture connects, culture contributes to growth, culture inspires, culture makes the world a better place – and the DDB aims to provide access to massive amounts of cultural data, in order to make our heritage widely available for everyone. But it's not only data! The DDB also provides innovation and interoperability, because we live in a world where the technical landscape keeps evolving. These are strategic tasks that need to be built on a solid foundation, and that is why the underpinning foundation of the DDB is a competence network of cultural and scientific institutions, financially supported by federal, state and local authorities, all contributing to the joint task of developing, advancing the DDB.

The support of the network institutions and their work together brought the DDB to where it is today. 33 million objects and counting! We know our tasks are not easy, we try to keep learning from the pain points and we know that recognising and addressing challenges will allow us to move forward.

Digitalia: How are the partner institutions organized?

Cosmina Berta: The administrative office of the DDB competence network is located at the Stiftung Preußischer Kulturbesitz (Prussian Cultural Heritage Foundation) in Berlin and the DDB technical development and service desk are located in the Deutsche Nationalbibliothek in Frankfurt.



Bringing together and linking up content from cultural and scientific institutions and also earning their trust to store and manage their metadata is an enormous technical and logistic undertaking.

The communities play a very important role in the cultural sector, so for a cross domain national aggregator like the DDB it is crucial to win their support and get them involved. So how have we done this? It is all about the specialized domain desks or the “Fachstellen” in German.

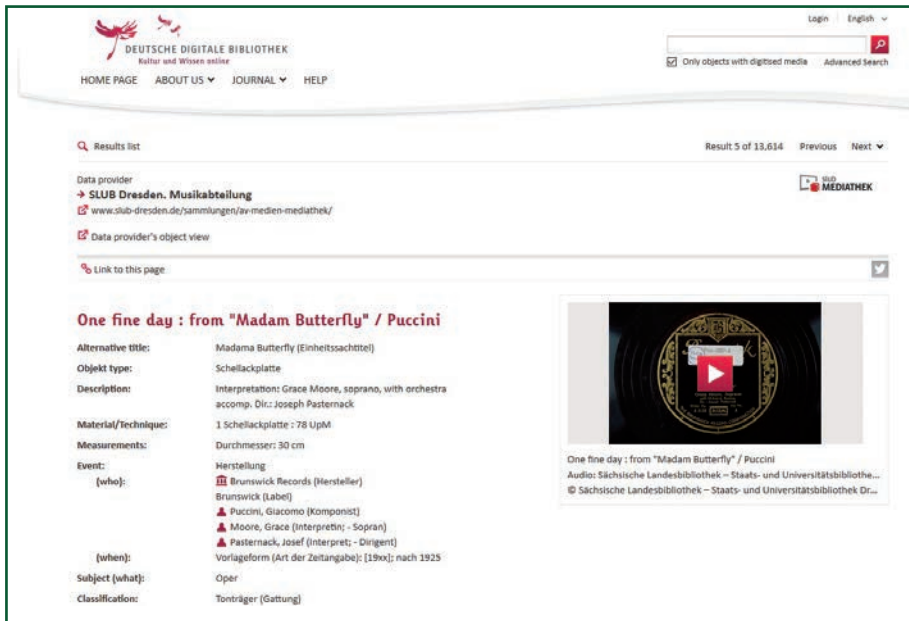
In order to ensure the best possible support for our data partners, the DDB set up domain-specific desks. You are an archivist? Then you can turn for support to our Archive domain desk in the Landesarchiv Baden-Württemberg. For a curator in a museum, our Museum domain desk at the Institute for Museum Research is ready to assist.

All in all, there are seven domain desks supporting the DDB, established within prestigious institutions in the respective sectors, offering professional support for institutions from all cultural areas on their way to participate in the DDB: Museum (at the Institute for Museum Research in Berlin), Media-Audio, Media-Image (both at the Saxon State and

University Library in Dresden), Media-Film (at the German Film Institute and Film Museum), Archive (Landesarchiv Baden-Württemberg), Library (at the Göttingen State and University Library).

An important point to make here is that the domain desks not only answer data related questions or explain workflows. They also offer all institutions, big or small, the opportunity to work with and learn from leading specialists, they enable them to learn new things and get inspired through community work at conferences, workshops and fairs.

And there is the Servicestelle or the service desk. I would like to use a metaphor here: If the cultural institutions that make their data available in our portal are the lifeblood of the DDB, I like to call the Servicestelle the heart of the DDB. That is because it has the central roles of coordinating the work of the domain desks, and to further the development of all data processes, establishing requirements for data processing tools. It also acts as first point of contact and consults and supports new and existing data partners. Not an easy task in innovative and often unstructured fields like the cultural and scientific sectors! We need to permanently step up our game, to



stay in contact with the latest developments in the cultural sector, to be able to deal with new and challenging tasks.

Digitalia: What future developments are planned for the DDB?

Gerke Dunkhase: On the content level, the DDB is busy preparing a new strategy process. One main point to be addressed is: How does the DDB react to new developments in the field of digitisation of cultural assets or in the field of artificial intelligence? These are questions that our teams in Frankfurt and Berlin will be dealing with this year. The aim is to develop a five year strategy for the DDB by the end of the year.

A second important point is to conduct user and usability studies. We have just conducted a large-scale online survey via the DDB portal, and the market research institute we commissioned to do this is currently conducting individual interviews and usability tests with selected users. We want to use the results to consistently align the products and development projects of the DDB with the requirements and expectations of our users.

Also, one of the DDB’s most important projects this year is the Newspaper Portal. The project, which is funded by the German Research Foundation (DFG), aims to make digitized historical newspapers from the 17th to the 20th century accessible via a separate section within the DDB. There will be full texts for most of the historical newspapers. This means that not only the meta-data but also, in contrast to the DDB, the full text contents of the newspapers can be searched. In the medium term, regional newspaper portals are also to be integrated into the DDB Newspaper Portal, so that in the future there will be a uniform access point for historical newspapers throughout Germany. In short: a national newspaper portal. With its corpus of historical newspapers, the Newspaper Portal is also an important service for the sciences. The public release of the Newspaper Portal is planned for December 2020.

Another important future topic for the DDB is authority files or vocabularies. The entity pages for persons and organizations in the DDB are already based on the Gemeinsame

Normdatei (Integrated Authority File), GND for short. Currently, authority files are primarily used in libraries and partly in archives. The project GND4C - GND for culture data, in which the DDB is a project partner, aims to win over cultural institutions from other sectors for the use of authority data. By enriching metadata with GND data, the appeal of the DDB can be significantly increased, either by linking object pages with each other or by improving the search quality.

Of course, it is an important goal to further increase the number of data partners and data sets in the DDB. In the past, the ingest - or data import - was a bit of a bottleneck at times, high demand of data ingestion simply prevented data from being published on the main portal. By switching to a new architecture, optimizing the data processing workflows and developing the administration com-

ponent DDBdash, we were able to significantly increase the data flow. To further boost the efficiency of our data supply processes, the Aggregator Fachstelle (aggregator desk) was set up in October 2019. With this specialized unit, we want to increasingly address aggregators as data providers and win them over for the DDB.

Improving the quality of the data is just as important as increasing the quantity. One way of doing this is having the domain desks and the service desk preprocess the data, for example by enriching it with authority data.

The DDB is to become the central point for researching digitised cultural assets while at the same time establishing references and linking together objects of different cultural institutions. Above all, the DDB should be a tool that can be used and worked with. These are the goals we are working towards.

Evaluating a Semantic Portal for the “Mapping Manuscript Migrations” Project

Toby Burrows - Oxford e-Research Centre, University of Oxford
Nicole Bergk Pinto - Mahaut Cazals - Alexandre Gaudin - Hanno Wijsman
 Institut de recherche et d'histoire des textes (CNRS-IRHT)

Questo articolo riporta una valutazione del portale semantico Mapping Manuscript Migrations (MMM), che combina in una struttura Linked Open Data dati su manoscritti medievali e rinascimentali provenienti da diverse fonti. Il gruppo di ricerca ha esaminato inizialmente l'esperienza utente per i nuovi fruitori del portale:

- *Il progetto MMM è presentato con sufficiente chiarezza per i nuovi utenti?*
- *Le istruzioni per navigare nel portale sono reperibili con facilità?*
- *L'interfaccia è intuitiva? La navigazione è facile? Quanto è semplice fare ricerche?*

Sono stati poi valutati i modi in cui le ricerche potevano essere inquadrare ed eseguite, insieme alla presentazione e all'uso dei risultati, incluso l'esame dei problemi derivanti dal collegamento incrociato di dati disparati e la ricerca di eventuali errori o imprecisioni evidenti nei dati. È stata inoltre esaminata l'efficacia delle visualizzazioni basate su mappe prodotte dal software Sampo-UI. Le raccomandazioni derivanti dalla valutazione sono state utilizzate come base per migliorare la funzionalità di Sampo-UI, aggiungendosi alla guida in linea fornita agli utenti del portale e aumentando la quantità e il tipo di informazioni disponibili agli utenti avanzati, soprattutto in relazione all'ambito e copertura dei dati MMM.

Context

Mapping Manuscript Migrations (MMM) is a project funded between 2017 and 2020 by the Digging into Data Challenge of the Trans-Atlantic Platform research funding consortium. The main goal of the project is to combine data from several disparate sources about medieval and Renaissance manuscripts, and to use the aggregated data to explore a range of research questions about manuscript history and provenance. The project took data from three existing databases (the Schoenberg

Database of Manuscripts¹, Medieval Manuscripts in Oxford Libraries², and Bibale³) and turned them into Linked Open Data (LOD)⁴. This involved transforming them into RDF triples and mapping them to a newly developed unified data model, drawing on the CIDOC-CRM and FRBR₀₀ ontologies. Vocabularies for the main classes of entity (manuscripts, actors, places, and works) were reconciled across the three data sources using a mixture of automatic and semi-automatic methods.

¹ <https://sdbm.library.upenn.edu/>.

² <https://medieval.bodleian.ox.ac.uk/>.

³ <http://bibale.irht.cnrs.fr/>.

⁴ Full details of the technical aspects of the MMM project can be found in the MMM *White Paper*, available here: <<https://diggingintodata.org/file/1281/download?token=x59u8fFQ>>.

The aggregated data (nearly 22.5 million RDF triples) have been made available in several different ways. A copy of the dataset has been published through the Zenodo repository⁵. The data are hosted on the Linked Data Finland platform, from which they can be queried through a SPARQL endpoint or inspected directly⁶. A semantic portal has also been implemented using the Sampo-UI framework, through which the 217,000 manuscripts and other entities can be searched and browsed, using a combination of filters and map-based visualizations⁷. Result sets from the portal can also be downloaded in the form of CSV files via the SPARQL query service Yasgui⁸.

The project convened an initial focus group of Oxford University researchers in 2017 and asked them to identify desirable features and functionality for the kind of data discovery service envisaged. It also gathered a set of research questions for testing, data modelling, and evaluation purposes. A User Group was established within the project, consisting of librarians, curatorial staff, and manuscript researchers from the project, with the aim of guiding and testing the implementation and customization of the semantic portal, as the user interface to the aggregated data. An early version of the portal was presented at a workshop during the 2019 “Digital Humanities” conference in Utrecht, and useful feedback was obtained from the researchers, librarians, and DH specialists who attended.

Once the User Group had signed off on the portal in early 2020, a more formal evaluation

was carried out. The three researchers who carried out this evaluation were employed by the Institut de recherche et d’histoire des textes (IRHT) for the final months of the project, but had not been involved in the design or implementation of the portal and therefore brought fresh eyes to the way in which it worked. Their report was then used to make additional modifications and improvements to the portal, and also served as the basis for developing a set of Frequently Asked Questions for the project’s Web site⁹.

Methodology

The evaluation team consisted of three early career researchers: two historians and one philologist. Their profile was generally similar to that of the intended users of the MMM portal: qualifications and research interests in the field of manuscript studies, with varying degrees of computer skills, but not specialists in manuscript curation or Digital Humanities. They did have some experience in adding records to the Bibale database, and consequently made some use of Bibale as a point of comparison. They were not given any specific training in the use of the MMM portal, and were asked to approach it as new users, drawing only on the “Help” information available on the site itself. They were asked to comment on what worked in the portal, what did not appear to work, and what was unclear about the design and the instructions.

Their starting-point was the set of 26 research questions previously assembled by the project team (see Table 1). Some of these

⁵ <https://zenodo.org/record/4019643>.

⁶ <http://ldf.fi/mmm/sparql>.

⁷ Esko Ikkala – Eero Hyvönen – Heikki Rantala – Mikko Koho, *Sampo-UI: A Full Stack JavaScript Framework for Developing Semantic Portal User Interfaces*, submitted to «Semantic Web Journal», (2020), <<http://www.semantic-web-journal.net/content/sampo-ui-full-stack-javascript-framework-developingsemantic-portal-user-interfaces>>.

⁸ <https://yasgui.triply.cc/>.

⁹ <http://blog.mappingmanuscriptmigrations.org/frequently-asked-questions/>.

questions were very specific, reflecting the interests of MMM project team members or the Oxford focus group, while others were more generic. The source for this latter group was a list of “Requêtes intéressantes” produced by the French Biblissima project¹⁰. The evaluation team also searched the portal on

the basis of their own professional and personal interests. They investigated topics that they were specifically familiar with, such as the collection of Claude Fauchet; the manuscripts of the “Tournement Antechrist” and “Le Roman de le Rose”; and the Montpellier area in the medieval period.

- [A1] How many manuscripts from pre-1600 produced in European countries survive?
- [A2] How many manuscripts survive that contain Spanish texts written in gothic rotunda were produced in Castile for an abbey or convent? Then show me those which were owned during the nineteenth century by English private collectors; Then show me those which are now owned by an institution in North America.
- [A3] What French collectors purchased manuscripts since the end of the Wars of Religion (after 1598)? Where are their manuscripts now?
- [B1] How many manuscripts containing texts by Ramon Llul were sold in the 19th century?
- [B2] Where are Ramon Llul manuscripts today?
- [B3] Who collects manuscripts with texts by Ramon Llul?
- [B4] How many times do texts by Ramon Llul's appear with texts by Albertus Magnus in the same manuscript?
- [C1] What was the most popular text by a medieval author in France in the 17th Century?
- [C2] Did Sir Thomas Phillipps own a 13th-century Bible with historiated initials?
- [F1] Combien de manuscrits enluminés se trouvent dans une collection particulière? (volumétrie)
- [F2] Quelle est la dynamique dans l'évolution des acquisitions et des dons? (étude diachronique)
- [F3] Qui sont les donateurs et les propriétaires d'une collection?
- [F4] Faire des recherches par sujet, par technique, par langue, par artiste voire par pigments (plus d'encre d'or, argent et pourpre) dans une collection.
- [F5] Particularités d'une collection (sujet, technique, lieu de production etc.)? Quelles en sont les lacunes? Quelles en sont les dominantes?
- [F6] Vie d'une collection, vie d'un livre enluminé?
- [F7] Quels manuscrits sont probablement perdus?
- [F8] Quel manuscrit a été vendu et n'est pas identifié au sein d'une collection à l'heure actuelle? (catalogue de vente)
- [G1] Quelles copies d'un texte sont enluminées?
- [G2] Quelle position occupe une copie dans l'histoire de la transmission d'un texte? Y a-t-il des exemplaires uniques des oeuvres?
- [G3] Histoire de la transmission des images? [Outside the scope of the MMM dataset]
- [G4] Quelles sont les versions existantes d'une oeuvre? Qui a fait une traduction française d'un texte ancien? Quand?
- [G5] Quelles sont les différentes publications existantes [manuscript copies] d'un texte? (date, lieu de production, personne(s) responsable(s) etc.)
- [H1] How many manuscripts were produced in Northern Italy and/or Lombardy?
- [H2] How many manuscripts were produced in the Low Countries?
- [H3] How many manuscripts were produced in London in the 15th century?
- [H4] How many manuscripts formerly owned by Sir Thomas Phillipps are in British libraries?
- [H5] What is the average number of folios in a Book of Hours?
- [H6] Which collectors bought manuscripts from Wilfrid Voynich? Where were they located? What do we know about the kind of manuscripts he sold, and their earlier histories?

Table 1. *Mapping Manuscript Migrations research questions*

¹⁰ [https://doc.biblissima.fr/ontologie-biblissima - méthodologie.](https://doc.biblissima.fr/ontologie-biblissima-methodologie)

The evaluation team focused initially on the user experience for new users of the portal:

- Is the MMM project presented with sufficient clarity for new users?
- Can one easily find instructions to browse the portal?
- Is the interface intuitive? Is browsing easy? How simple is it to make queries?

They then went on to evaluate the ways in which queries can be framed and executed, together with the presentation and use of the results. This included examining issues arising from the cross-linking of disparate data, and looking for any noticeable errors or inaccuracies in the data. They also examined the effectiveness of the map-based visualizations produced by the Sampo-UI software.

Findings

The findings from this evaluation were, in general, very positive and enthusiastic. Browsing the portal is easy and does not require specific computer skills. Basic queries (e.g., finding a manuscript, a collection, a place, etc.) are very straightforward. Using the filters and the map visualizations is very intuitive, even enjoyable. But the findings of most interest to the project team concerned areas where improvements could be made, or where clearer explanations were required.

Some of these related to specific features of the Sampo-UI software. Moving the date slider with precision was felt to be difficult, for example. Instead of displaying all the many columns of information about each manuscript, by default, it would be helpful to be able to select which columns to display. Pie charts are provided as a way of visualizing the results of some filters (such as owners), but not for others (such as production dates). But their value might be limited when the percentages are too small to display. The percentage of French-language manuscripts in

the Yale University Library is labelled as 6.6% in the appropriate pie chart, for instance. But the percentages for the least-represented languages in the Yale collections are too small and cannot be displayed.

Selecting and combining different elements within a filter was thought to lack certain options. Filtering by “Owners” to find manuscripts owned by either Thomas Phillipps or Thomas Thorpe, for example, was possible, but not manuscripts owned by both Phillipps and Thorpe. Some combinations for filtering by the hierarchical lists of places were not possible, especially with overlapping regions or terms like Languedoc, Occitanie, and Southern France. The other approach to filtering – drawing and using a “Bounding Box” on the map – was not necessarily an effective alternative.

In the tabular display of information relating to manuscripts, there are separate columns for “Owner” and “Transfer of Custody”. This type of presentation was felt to make it difficult to connect these provenance events with the relevant owners of a manuscript. Because the connections between columns are not displayed – except inasmuch as they relate to the same manuscript – the only way to see these relationships is to click on a specific element, such as a “Transfer of Custody” event. This will display all the available information about that event, including the persons and organizations involved, in the “Custody surrendered by” and “Custody received by” fields.

The export tool based on the underlying SPARQL queries was found to be relatively simple to use, but the resulting CSV file was less easy to understand. A manuscript like “Montpellier (F), BU Historique de Médecine, H 069”, for example, produces a spreadsheet with six lines rather than one, together with a total of 84 columns. Because there is so much information about each manuscript, often

with multiple values for the same element (e.g., production date), it would be difficult and possibly misleading to combine all the information on to a single row. Once the CSV file has been exported and downloaded, it is possible to combine multiple rows using software like OpenRefine or Google Sheets.

A number of the findings appeared to relate to the Sampo-UI software but were actually connected to the nature and structure of the source data, and the ways in which the data had been mapped and harmonized. When the manuscripts are filtered by production place and the results are presented in a table, the total number of results may differ. Filtering for “Europe” as “Production Place” shows a total of 86,332 in the filter, but 73,665 in the table of results. This is because the first is the number of places and the second is the number of manuscripts. A manuscript may have multiple production places assigned to it, either because of disagreements between the data sources or because of uncertainty about its origins: Southern France or Northern Italy?

Manuscripts belonging to a specific person or organization can be filtered by “Collections” as well as by “Owners”, but these produce different results. Claude Fauchet as an “Owner” owns nine manuscripts but his “Collection” contains only one manuscript; Vossius owns seventy manuscripts, though his collection only shows five of them. This is because the “Collection” information is derived exclusively from the Bibale database; the other data sources do not include the separate concept of a collection in their data model for provenance histories. “Owner” information, on the other hand, may come from any of the three data sources and will invariably have the fullest picture of a specific person’s manuscript holdings.

The different stages in the migration of a manuscript are not shown on the map visual-

izations. They only show the place of production and the last-known location, together with the arc joining these two places. This is because the information in the source datasets is simply not full enough or sufficiently well-structured to arrange the intervening places in the history of a manuscript’s ownership over the centuries in chronological order. The available information is displayed in the events and owners columns in the tabular presentation of data for each manuscript, however, and can also be found through more complex SPARQL queries. The use of last-known location rather than current location was also queried. Although the current location is obvious for manuscripts from the Oxford catalogue, Bibale does not necessarily record a current location, while the Schoenberg Database focuses on last-known locations associated with or recorded in catalogue entries.

The evaluation noted that some authors had multiple entries, while other authors had only one entry. This is because many authors appear in different forms in the data sources. The MMM project tried to harmonize as many of these as possible, but there are still some authors who have not been harmonized. As a result, they appear as two or more different names in the portal. Searching the “Author” filter and selecting the different forms there will produce a combined set of results. Authors may be linked to many works, even though they are the author of only one of these works. This arises from the structure of some entries in the Schoenberg Database, where multiple works and multiple authors in a single manuscript are not individually linked. In these cases, the author is labelled “possible author”.

The evaluation also noted some repetitions and inconsistencies in some elements of the manuscript descriptions. This is usually because the MMM project has combined data

about the same feature from two or three different sources. For production dates and places, in particular, it is quite often the case that (for example) the Schoenberg Database has multiple differing values for the same manuscript from different sales catalogues. The MMM project chose to display all these variations, together with their source, rather than attempting to merge them or identify a single “most-accurate” value.

More generally, some of the findings related to the overall scope and purpose of the MMM project and its data. While the researchers found the portal’s design to be attractive and inviting for new users, they also commented that it was hard to know, at a glance, what the aims of the portal were, what data were incorporated in it, and what its limits were. They were unsure to what extent the data had been harmonized, how complete or incomplete the manuscript data were, and what information had not been mapped from the source datasets.

These issues were demonstrated through specific research inquiries devised during the evaluation. Manuscripts with texts relating to the Carolingian Duke Guillaume d’Orange cannot easily be found in a single search of works, partly because the MMM data are not oriented towards mapping textual traditions specifically and partly because it proved difficult to reconcile the titles of works across the data sources. There are relatively few results with manuscripts produced in the south of France during the Middle Ages, simply because this reflects the coverage in the data sources.

The evaluation report suggested the provision of fuller information to guide users in

their expectations of the portal and the data. This might include explaining the nature of the data, and the possibilities and limitations arising from such factors as the nature of the data in the various constituent databases, the processes for harmonization and reconciliation of the data, and the way in which the portal was built, including the data model.

Outcomes

The evaluation report by the three IRHT researchers presented specific recommendations and suggestions for addressing the issues identified. The most important of these concerned the need to provide more explanatory and contextual information about the portal. As a result, the “Info” section of the portal has been expanded with links to the project’s technical documentation on GitHub¹¹ and to the project’s Web site and blog¹², where a set of Frequently Asked Questions has been added. The FAQ covers general questions about the scope and purpose of MMM and the nature of its data, as well as more specific questions about the use and functionality of the MMM portal¹³. The GitHub documentation includes a SPARQL tutorial based on the MMM data and research questions¹⁴.

Several recommendations suggested specific improvements to the functionality of the portal. Where feasible, these were added to the workplan for the Sampo-UI software, and most have subsequently been implemented. They included the following items:

- Combining values from within one filter: an enhancement will give users the choice of using “AND” within a filter (to find manuscripts owned by both Phillipps and Beatty, for example) instead of the default “OR”

¹¹ <https://mapping-manuscript-migrations.github.io/>.

¹² <http://blog.mappingmanuscriptmigrations.org/>.

¹³ <http://blog.mappingmanuscriptmigrations.org/frequently-asked-questions/>.

¹⁴ https://mapping-manuscript-migrations.github.io/sparql/sparql_tutorial.html.

combination (manuscripts owned by either Phillipps or Beatty);

- Reducing the number of columns which display in the results table: an enhancement will allow users to select columns for display, as an alternative to the default display of all columns;
- Extending the display of percentages in pie charts to all the values;
- Replacing the date slider with “from” and “to” boxes, for selecting date ranges for manuscript dates of production;
- Providing an option to bypass the automatic selection of all subsidiary places when filtering manuscripts by place of production.

Other issues raised in the evaluation report were related to the nature of manuscript provenance data more generally, and the ways in which such data are recorded and structured. These arose partly because of the inherent incompleteness and uncertainty of the data, and partly because of limitations and assumptions in the data models of the source datasets, which made some elements of them difficult to map with sufficient granularity or specificity. Some compromises had to be made in the MMM data modelling and mapping processes in order to find common semantic ground across three very different sets of complex data.

The project has produced a White Paper which surveys these issues in more detail, and has published its data model and its data for other projects or users to work with and improve on¹⁵. It is also developing specific recommendations for enhancing the TEI encod-

ing of provenance in manuscript catalogues, with the aim of improving the mapping process to event-based data models like that employed by MMM.

Conclusion

The overall conclusion of the evaluation report was that the MMM portal is an excellent tool, and very easy to use. Nevertheless, the report also showed that it is important to understand and acknowledge the scope and parameters of the MMM data underlying the portal, and the inherent limitations which result from them. The source data have not been corrected or amended, and they have been harmonized rather than merged. Some elements, especially those unique to one of the sources, have not been mapped. But links to the original records in the data sources have been provided, so that the full information about any given manuscript in that source can be easily found.

The portal is a very rich tool for exploring the history of books, of cultural exchanges, of libraries, and of collections. This tool also makes it possible to follow the history of a given manuscript, and to conceive of the manuscript as a source (text additions, transformation, reason for its preservation, and so on). It is easy to create corpora organized by collection or owner, but more difficult with regard to geographical areas or historical periods. It is also possible to create corpora according to production places (when this information exists), and to study the vitality of specific scriptoria.

¹⁵ <https://diggingintodata.org/file/1281/download?token=x59u8ffQ>.

This paper reports on an evaluation of the semantic portal Mapping Manuscript Migrations (MMM), which combines data from several different sources about medieval and Renaissance manuscripts, in a Linked Open Data framework. The evaluation team looked initially at the user experience for new users of the portal:

- *Is the MMM project presented with sufficient clarity for new users?*
- *Can one easily find instructions to browse the portal?*
- *Is the interface intuitive? Is browsing easy? How simple is it to make queries?*

They then went on to evaluate the ways in which queries could be framed and executed, together with the presentation and use of the results. This included examining issues arising from the cross-linking of disparate data, and looking for any noticeable errors or inaccuracies in the data. They also examined the effectiveness of the map-based visualizations produced by the Sampo-UI software. The recommendations arising from the evaluation were used as the basis for improving the Sampo-UI functionality, adding to the online help given to users of the portal, and increasing the amount and type of information available to advanced users, especially in relation to the scope and coverage of the MMM data.



Archivissima 2020: Gli archivi in digitale

Ilaria Scalmani

Archivio centrale UDI

Si è svolta il primo weekend di giugno *Archivissima 2020 – Digital edition*, la terza edizione di una full immersion negli archivi italiani. Un'intensa programmazione che ha visto dalle 18 del 5 giugno per tutta la notte più di 200 archivi pubblici e privati partecipare all'iniziativa.

Durante la "notte degli archivi" un folto palinsesto sulla pagina FB di Archivissima ha mandato in onda ogni 5 minuti trasmissioni sugli archivi. Il tema di quest'anno erano le donne. Ogni ente ha dato libero sfogo nel raccontare profili o storie di donne nei vari ambiti museali, istituzionali, sportivi, politici ecc. È stato così possibile assistere in diretta a dibattiti, video, ricostruzioni che avessero al centro la donna.

Ma accanto alla Notte degli archivi del 5 giugno, patrocinata dall'Anai, la manifestazione è andata avanti tutto il week end con materiali digitali accessibili anche sul sito di Archivissima¹, dove si possono ascoltare i podcast presentati da testimoni e voci d'eccezione.

Un festival aperto a tutti e tutte, dai bambini ai più grandi, toccando le mille sfaccettature della vita quotidiana, sociale, lavorativa, culturale, aprendo le porte a luoghi che spesso non si conoscono o a cui non si è avuta mai la possibilità di accedere.

È stato così stimolante "entrare" nell'Archivio del Museo Egizio di Torino per sbirciare i documenti contenenti la storia della scoperta della tomba di Nefertari, una delle donne più importanti della storia egizia; "andare" a

Roma nell'Archivio Centrale dell'UDI per vedere i documenti, le fotografie, i manifesti e i cimeli che raccontano la storia delle donne che hanno fatto l'Italia; "proseguire" nell'Archivio di stato di Agrigento per entrare nel mondo della scuola attraverso i registri di classe (anno scolastico 1938/39) e leggere le parole delle insegnanti attraverso i giudizi e le valutazioni delle donne in piena epoca fascista. Un vero e proprio viaggio negli archivi italiani da Nord a Sud.

La manifestazione, come nelle edizioni precedenti, era stata pensata in presenza, tenendo per una notte bianca tutti gli archivi aderenti aperti, in via eccezionale, al pubblico. Tuttavia il Covid 19 ha cambiato le regole e pertanto è stato deciso, un po' per scommessa, di trasportare tutto sul digitale e quindi online. La scommessa è stata vincente e questi sono i numeri che lo dimostrano:

- 73.000 visualizzazioni di video durante La Notte degli Archivi
- 1.000 visualizzazioni medie per ogni video
- 37.700 accessi al sito e ai podcast nei soli giorni del festival
- 125.000 utenti unici
- 207.000 visite ai contenuti prodotti nei soli giorni del festival
- 125 podcast inediti prodotti
- 95 video inediti prodotti
- circa 200 archivi partecipanti

Dato l'importante risultato si è pensato di mantenere tutti i materiali consultabili sul sito, nel quale a breve verranno integrati i video

¹ www.archivissima.it.

prodotti e trasmessi durante la Notte degli archivi sulla pagina facebook. Ciò creerà un archivio degli archivi arricchito di materiali e informazioni eterogenee e differenziate.

L'interessante offerta che è stata proposta sia per i video e i podcast, che per i dibattiti sul tema degli archivi che sono succeduti nei giorni a seguire, per la prima volta grazie al digitale è arrivata in molte case: ciò ha acceso un no-

tevole interesse, dimostrando come le nuove tecnologie possono interagire con la memoria dando visibilità alle "carte polverose", creando momenti di riflessione e comparazione tra passato e presente.

Tutto ciò permetterà per l'edizione del prossimo anno, già confermata, di ampliare il novero degli archivi dando sempre più visibilità a luoghi troppo spesso sottovalutati e poco conosciuti.

Segnalazioni

Un assaggio di Leopardi. Il progetto BDL – Biblioteca Digitale Leopardiana

Gioele Marozzi

Ph.D. Student. Scuola di Dottorato dell'Università degli Studi di Macerata

Per il triennio accademico 2018/2021, la Scuola di Dottorato dell'Università di Macerata ha avviato il progetto "Biblioteca Digitale Leopardiana", nell'ambito del programma di ricerca applicata Eureka della Regione Marche e con il cofinanziamento di Filippetti Spa, azienda leader nel settore della trasformazione digitale. La ricerca è condotta sotto l'egida del Centro Nazionale di Studi Leopardiani di Recanati e si inserisce nel quadro delle attività promosse per la celebrazione dei 200 anni della composizione de *L'infinito* (1819) da un apposito Comitato Nazionale. Prevede la catalogazione, la metadattazione e la digitalizzazione degli autografi leopardiani conservati nei principali fondi extra-napoletani, in sintonia con quanto si sta facendo presso la Biblioteca Nazionale Vittorio Emanuele III di Napoli sulle carte leopardiane, nella quale un gruppo di lavoro sta realizzando un'analoga operazione sull'imponente Fondo ivi conservato, composto da circa 2.000 carte oltre alle 4.526 dello *Zibaldone*.

La ricerca rientra nel più generale progetto di allestimento di una biblioteca digitale d'autore, che comprenda diverse tipologie di documenti di interesse leopardiano: dalla produzione manoscritta alle edizioni, dai postillati delle opere all'iconografia e a una selezione di testi di autori vari che, disponibili in full-text, consentano di stabilire interrelazioni con l'attività di Giacomo Leopardi e costituire una fonte di primario interesse per gli studi letterari in generale e per quelli legati alla leopardistica in particolare.

Oggetto specifico del programma condotto a Macerata come percorso di formazione PhD coordinato da Laura Melosi per la Cattedra Giacomo Leopardi e da Dante Ciantra per Filippetti Spa, è la descrizione, catalogografica e digitale, del ricco patrimonio di manoscritti autografi di Leopardi che non sono conservati presso la Biblioteca Nazionale di Napoli. Le dinamiche del lavoro leopardiano, la spedizione delle lettere, il collezionismo privato, doni e vere e proprie spoliazioni hanno condotto nel tempo alla dispersione di una notevole quantità di documenti autografi di Giacomo Leopardi, che oggi si trovano disseminati in oltre 80 istituti non soltanto del territorio italiano, ma anche in Polonia, Gran Bretagna, Germania, Svizzera e Stati Uniti. Tale situazione crea non poche difficoltà sia dal punto di vista logistico, giacché non è mai stata realizzata una mappatura complessiva dei manoscritti e spesso risulta difficoltoso avere notizia di tutte le localizzazioni, sia per le puntuali necessità di ricerca, dal momento che chiunque volesse consultare i documenti dovrebbe necessariamente spostarsi di istituzione in istituzione, senza avere formalmente la possibilità di mettere a confronto carte custodite in enti diversi e tra loro anche molto distanti. Si consideri, ad esempio, il caso dell'*Epistola al conte Carlo Pepoli*, attestata da due autografi conservati rispettivamente presso l'Archivio Storico del Comune di Visso e la Biblioteka Jagiellońska di Cracovia: senza uno strumento digitale, la

comparazione diretta dei due documenti è di fatto impossibile. L'allestimento del database, cuore del progetto, vuole contribuire precisamente a sanare - almeno in forma virtuale - questa situazione, ricongiungendo seppure soltanto in digitale una *libreria* che altrimenti non potrebbe essere in alcun modo indagata sinchronicamente.

In tale scenario, il progetto si concretizza nella sinergia di tre fasi distinte, ciascuna gestita con strumenti e competenze specifiche. Poiché per valorizzare qualsiasi bene culturale il requisito imprescindibile è conoscere le caratteristiche, la prima attività da compiersi - realizzata già per una parte dei documenti - è la ricognizione e la catalogazione del materiale autografo leopardiano. In assenza di uno strumento, analogico o digitale, che dia conto della localizzazione dei manoscritti, si rivela necessario il monitoraggio attento e costante di cataloghi (anche d'asta), di edizioni critiche, di saggi, studi e lavori che forniscano informazioni a proposito della collocazione fisica degli autografi di Giacomo Leopardi; l'individuazione degli istituti collettori è propedeutica e necessaria all'avvio delle procedure di catalogazione, svolte attraverso il software Manus OnLine (MOL), sviluppato e messo a disposizione dall'Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU). Attraverso la scheda Manus è possibile, infatti, valorizzare metadati specifici in relazione al supporto di cui consta il manoscritto e al testo che esso veicola, individuando anche aspetti apparentemente secondari, quali lo spessore della carta, la filigrana o la presenza di annotazioni non autografe, che potrebbero contribuire in maniera sostanziale a ricostruire la storia tanto del manoscritto quanto dell'opera o dello scritto che esso trasmette.

Contestualmente alla catalogazione, è avviata anche la seconda attività del progetto, dedicata alla digitalizzazione delle carte, anch'essa condotta in stretta relazione con

l'ICCU, dal momento che gli standard utilizzati in sede di acquisizione delle immagini seguono il protocollo di Internet Culturale, sul cui portale le riproduzioni verranno ospitate e messe a disposizione degli utenti.

La terza e ultima fase del progetto, condotta in collaborazione con il partner aziendale Filippetti Spa del Gruppo Filippetti, riguarda l'allestimento di uno specifico database, dedicato in prima istanza allo studio approfondito dell'opera di Giacomo Leopardi. La base dati, infatti, non si propone di implementare soltanto la visualizzazione delle immagini e la presentazione delle schede di catalogo dei singoli manoscritti, ma intende offrire ai ricercatori anche dettagli aggiuntivi su documenti dotati di peculiare interesse storico-letterario o filologico, arrivando a esperienze inedite di ricerca garantite, ad esempio, dalla presentazione di specifici percorsi tematici. Premessa necessaria per lo sviluppo di quest'ultima fase del progetto è l'allestimento di un apposito motore di ricerca, capace di cogliere i metadati dei documenti e di metterli in relazione tra loro per garantire risultati accurati e in costante aggiornamento, anche in virtù del popolamento del database con notizie e documenti sempre nuovi.

Relativamente all'impatto atteso, un progetto siffatto ha come ricaduta la disseminazione delle conoscenze e la diffusione della leopardistica a livello non soltanto locale, ma anche internazionale; dal punto di vista tecnico-scientifico, inoltre, i percorsi di ricerca su Giacomo Leopardi potranno giovare di una piattaforma testuale e di immagini collocabile accanto a progetti di altrettanto prestigio, come quello dedicato allo *Zibaldone* e realizzato da un team di ricercatori afferente all'Università di Princeton; il tutto, a conferma della vocazione dell'Università di Macerata a un umanesimo che innova. Nell'attuale assetto sociale, infatti, che ha subito e continua a subire anche l'impatto della diffusione del COVID-19, la digitalizzazione del patrimonio culturale conservato presso vari istituti italiani

e, nel caso di specie, internazionali si lega strettamente a uno dei pilastri su cui il Ministero per i beni e le attività culturali e per il turismo fonda la propria attività. Le nuove tecnologie possono essere affiancate proficuamente alla tradizionale frequenza dei luo-

ghi della cultura per continuare a comunicare valore e per incentivare la conoscenza degli enti e dei beni stessi, nell'ottica di una fruizione sempre più ibrida della nostra cultura che non faccia dimenticare, al contempo, le radici da cui essa trae la propria linfa vitale.





Pubblicazioni

Futuro delle memorie digitali e patrimonio culturale: atti del convegno internazionale, Firenze, 16-17 ottobre 2003, a cura di Vittoria Tola e Cecilia Castellani, Roma, ICCU (2004)
ISBN 978-88-7107-110-7

Memorie digitali: rischi ed emergenze, a cura di Alessandra Ruggiero e Vittoria Tola, Roma, ICCU (2005)
ISBN 978-88-7107-111-5

Linee guida per registrazioni d'autorità e di rinvio, Roma, ICCU (2005)
ISBN 978-88-7107-105-0

Linee guida per la digitalizzazione del materiale fotografico, a cura del Gruppo di lavoro sulla digitalizzazione del materiale fotografico, Roma, ICCU (2005)
ISBN 978-88-7107-112-3

EAD: Descrizione archivistica codificata: dizionario dei marcatori, a cura di Giovanni Michetti, Roma, ICCU (2005)
ISBN 978-88-7107-115-8

MAG: metadati amministrativi e gestionali: manuale utente, a cura di Elena Pierazzo, Roma, ICCU (2006)
ISBN 978-88-7107-117-4

Linee guida per la digitalizzazione del materiale cartografico, a cura del Gruppo di lavoro sulla digitalizzazione del materiale cartografico, Roma, ICCU (2006)
ISBN 978-88-7107-116-6

Inter omnes: contributo allo studio delle marche dei tipografi e degli editori italiani del XVI secolo, Roma, ICCU (2006)
ISBN 978-88-7107-114-X

Le edizioni italiane del XVI secolo: censimento nazionale, Vol. 5: D, Roma, ICCU (2006)
ISBN 978-88-7107-113-1

Linee guida per la digitalizzazione di bandi, manifesti e fogli volanti, a cura del Gruppo di lavoro sulla digitalizzazione di bandi, manifesti e fogli volanti, Roma, ICCU (2006)
ISBN 978-88-7107-118-2

Miniature e decorazioni dei manoscritti, a cura di Emilia Ambra, Angela Dillon Bussi, Massimo Menna, Roma, ICCU (2006)
ISBN 978-88-7107-119-0

Le edizioni italiane del XVI secolo: censimento nazionale, Vol. 6: E-F, Roma, ICCU (2007)
ISBN 978-88-7107-125-1

Il libro italiano del XVI secolo: conferme e novità in EDIT16: atti della giornata di studio, a cura di Rosaria Maria Servello, Roma, ICCU (2007)
ISBN 978-88-7107-124-4

OAIS: Sistema informativo aperto per l'archiviazione, a cura di Giovanni Michetti, Roma, ICCU (2007)
ISBN 978-88-7107-126-8

ISBD(CR): International Standard Bibliographic Description for Serials and Other Continuing Resources: revisione dell'ISBD(S), Roma, ICCU (2007)
ISBN 978-88-7107-104-2

Regole italiane di catalogazione: REICAT, a cura della Commissione permanente per la revisione delle regole italiane di catalogazione, Roma, ICCU (2009)
ISBN 978-88-7107-127-5

Requisiti funzionali per i dati di autorità: un modello concettuale, a cura di Glenn E. Patton, Roma, ICCU (2010)
ISBN 978-88-7107-129-9

International Standard Bibliographic Description (ISBD), IFLA, Roma, ICCU (2010)
ISBN 978-88-7107-128-2

Guida alla catalogazione SBN. Musica: musica e libretti a stampa, registrazioni sonore, video e risorse elettroniche musicali, Roma, ICCU (2012)
ISBN 978-88-7107-130-5

ISBD: International Standard Bibliographic Description, edizione consolidata, Roma, ICCU (2012)
ISBN 978-88-7107-131-2

Titolo Uniforme Musicale: norme per la redazione, Roma, ICCU (2014)
ISBN 978-88-7107-133-6

Per l'acquisto delle edizioni ICCU rivolgersi a:

ICCU Ufficio pubblicazioni
Viale Castro Pretorio, 105
00185 Roma
T +39 06 49210425
F +39 06 4959302
e-mail: venditapubbl@iccu.sbn.it
<https://www.iccu.sbn.it>



ICCU

Istituto centrale per il catalogo unico
delle biblioteche italiane e per le informazioni bibliografiche
www.iccu.sbn.it

Copyright © ICCU - Roma

2020 - Istituto Poligrafico e Zecca dello Stato S.p.A. - Roma
Finito di stampare nel mese di dicembre 2020
presso lo stabilimento IPZS di Via Salaria, 691 Roma