

Il progetto di digitalizzazione della Biblioteca Estense Universitaria di Modena

Venerdì 15 marzo 2019, in Sala Solari, è stato presentato un grande progetto di digitalizzazione della Biblioteca Estense Universitaria di Modena. L'evento, a cura di MLOL, si è svolto durante il tradizionale convegno bibliotecario delle Stelline.

Sono intervenuti: Martina Bagnoli (direttore Gallerie Estensi), Paolo Cavicchioli (presidente Fondazione Cassa di Risparmio di Modena), Giulio Blasi (MLOL), Luca Panini (Franco Cosimo Panini Editore), Sauro Salvadori (Hyperborea), Angelo Perduca (Mida), Andrea Zanni (MLOL).

Il progetto di digitalizzazione dei fondi della Biblioteca Estense Universitaria di Modena (BEU), appena avviato dalle Gallerie Estensi, dirette da Martina Bagnoli, è stato finanziato dalla Fondazione Cassa di Risparmio di Modena (FCRMO) attraverso un bando da un milione e 300 mila euro. Il progetto *digital humanities* prevede il restauro, la digitalizzazione in alta definizione e la catalogazione del fondo musicale e cartografico della biblioteca, oltre alla metadattazione e al recupero del patrimonio già digitalizzato. Prevede inoltre l'archiviazione digitale delle immagini e la realizzazione di una biblioteca digitale ad accesso libero, con interfacce per l'annotazione e il *crowdsourcing*, realizzate grazie al protocollo IIIF.

Nel 2017 le Gallerie Estensi, con il supporto della FCRMO, avevano già avviato tre assegni di ricerca su tre dei fondi principali delle collezioni della biblioteca: fondo musicale estense, fondo cartografico, fondo Muratoriano. Su questi nuclei documentali poi si è concentrato il nuovo bando di digitalizzazione. Per la Fondazione, questi finanziamenti si inseriscono in un più ampio progetto culturale e architettonico, AGO Modena Fabbriche Culturali, nato con l'obiettivo di mettere a sistema le istituzioni culturali cittadine, realizzando il nuovo polo culturale di Modena, con la riqualificazione del complesso dell'ex ospedale Sant'Agostino, che ospiterà la Fondazione Modena

Arti visive, i Musei universitari, il Teatro anatomico, parte delle collezioni librerie della Biblioteca Estense, il Centro interdipartimentale sulle digital humanities dell'Università di Modena e Reggio Emilia e il Future Education Modena.

Inoltre, un ulteriore progetto di ricerca coinvolgerà il laboratorio AlmageLab del Dipartimento di Ingegneria Informatica diretto da Rita Cucchiara, che utilizzerà le immagini digitalizzate per il *training* di alcune algoritmi di machine learning. Gli obiettivi sono molto ambiziosi: il riconoscimento automatico della grafia manoscritta di Ludovico Antonio Muratori e dei toponimi presenti sulle mappe cartografiche.

Teca Digitale

Lo scopo della *digital library* rimane quello di sfruttare appieno le potenzialità delle nuove tecnologie digitali e la vastità del web come mezzo sia di promozione che di valorizzazione dei beni culturali.

Il primo obiettivo è quello di consentire l'accesso al patrimonio della BEU digitalizzato nell'ambito di precedenti campagne di digitalizzazione,¹ al momento accessibile soltanto attraverso un terminale fisso collocato nella sala di lettura della BEU stessa.

L'attuale Teca Digitale - così chiamata - offre agli utenti la possibilità di visualizzare circa 330.000 immagini, a cui non sono associati metadati descrittivi, a parte il numero di segnatura presente nel nome del file, che costituisce quindi l'unica possibile chiave di accesso, dopo aver effettuato una ricerca manuale in diversi cataloghi.

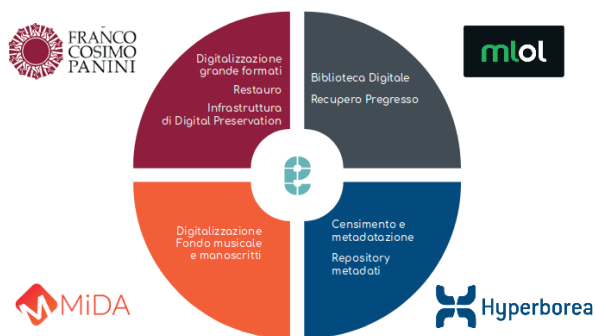
È da sottolineare infatti che la BEU manca di un catalogo unico, ma possiede diversi inventari e cataloghi parziali, realizzati in epoche diverse e con diverse metodologie e finalità. Sono cataloghi spesso incompleti, presenti solo in formato cartaceo o digitalizzati in formato immagine (es. in PDF senza OCR). Tale circostanza limita l'accesso a studiosi e ricercatori che abbiano una conoscenza approfondita dei fondi della Biblioteca. Questo impedisce che l'importante patrimonio culturale della biblioteca possa essere consultato da un'utenza non locale e/o non specialistica: il personale bibliotecario riceve continue richieste di digitalizzazioni di documenti da parte di studiosi di tutto il mondo.

In Teca Digitale è presente anche l'archivio Muratoria-

no, consistente di quasi 180.000 immagini raccolte in circa 6.700 cartelle, che è confluito, nella parte dei carteggi e documenti, sia in Internet Culturale² che in Europea, arricchito da metadati descrittivi a cura del personale bibliotecario. Le scansioni delle filze dalla 69 alla 98 sono però disponibili solo in Teca Digitale.

Obiettivo primario del progetto è dunque rendere facile l'accesso al patrimonio già digitalizzato, come ai nuclei che verranno digitalizzati ex novo, attraverso un'interfaccia semplice e un motore di ricerca che utilizzi anche i metadati (titolo, autore ecc.) recuperati dai cataloghi precedenti.

Le aziende



Extense è l'associazione temporanea di imprese che si è aggiudicata il bando e per i prossimi tre anni si occuperà della realizzazione del progetto.

Capofila è la modenese Franco Cosimo Panini Editore, che si occupa del restauro, della fornitura del sistema di archiviazione digitale e delle digitalizzazioni dei documenti di grande formato, come le mappe; Horizons Unlimited per la realizzazione e la gestione della piattaforma digitale della Biblioteca Estense mette a disposizione la piattaforma MLOL, che conta circa 6.000 biblioteche aderenti in 19 regioni italiane e 9 paesi stranieri con oltre 800.000 utenti registrati; Hyperborea si occupa del coordinamento del progetto, del censimento dei nuclei documentari e della metadatozione, fornendo il software di catalogazione Arianna; MiDA Informatica si occupa infine della maggior parte delle attività di digitalizzazione a esclusione dei grandi formati.

Altro elemento di novità è rappresentato dall'inclu-

sione nel progetto dell'iniziativa "Digitiamo", laboratorio di inserimento lavorativo per adolescenti e giovani con autismo, reso possibile dalla collaborazione con l'associazione Aut Aut, che vedrà coinvolti i ragazzi nelle attività di digitalizzazione, al fine di valorizzarne le attitudini e metterne a frutto i talenti.

Il processo di lavorazione

Nello specifico, i fondi da digitalizzare saranno:

- circa 800 mappe, in formati medio-grandi, del fondo cartografico;
- il fondo musicale, conservato in circa 1.000 cassette e consistente in oltre 20.000 oggetti.

Per questi documenti dovranno anche essere create schede descrittive con metadati minimi, recuperati da cataloghi e inventari esistenti, oltre che dal lavoro svolto dai ricercatori di Unimore.

Inoltre, dovranno essere create le schede per il patrimonio già digitalizzato, cioè:

- il fondo Muratori, contenente circa 180.000 immagini;
- la restante Teca Digitale, che ne contiene oltre 330.000.

Sono quindi stati definiti due processi, uno per il materiale da digitalizzare ex novo, un altro per la normalizzazione delle digitalizzazioni pregresse.

Il primo prevede che tutti i documenti siano esaminati da una restauratrice prima della digitalizzazione, al fine di valutarne lo stato di conservazione e garantire l'identificazione. Tutti i documenti che non passeranno la verifica verranno stabilizzati e digitalizzati solo dopo la messa in sicurezza, o saranno esclusi dal processo ove richiedano un intervento di restauro di eccessiva complessità.

Il secondo processo – sul patrimonio già digitalizzato – prevede invece:

- un'analisi e censimento della Teca Digitale;
- l'estrazione e normalizzazione delle signature presenti nei nomi dei file;
- il *matching* di tali signature con cataloghi e inventari esterni (Internet Culturale, MANUS ecc.) per il recupero dei metadati esistenti;
- il rilevamento di metadati minimi da parte di catalogatori specializzati.

Alla fine di entrambi i processi, le immagini digitali confluiranno nel DAM (Digital Asset Manager), infrastruttura di *long term preservation* e i metadati descrittivi a esse collegati nel software Arianna. Infine, tutto verrà pubblicato tramite la piattaforma MLOL.

Protocollo IIIF

Al centro del progetto, vi è l'utilizzo di IIIF (International Image Interoperability Framework), un protocollo open source per la presentazione, la ricerca e l'annotazione di immagini ad alta definizione. Nato alla fine del 2011 da un consorzio di importanti biblioteche,³ è divenuto in pochi anni uno standard de facto per le grandi istituzioni culturali di tutto il mondo. In Europa, si segnalano la BnF, la Bodleian Library di Oxford, gli archivi svizzeri e la Biblioteca Apostolica Vaticana, che a settembre 2018 ha anche firmato una dichiarazione congiunta con l'Istituto Centrale per il Catalogo e la Documentazione (ICCD):⁴ non ci risulta che biblioteche italiane l'abbiano ancora implementato per il loro patrimonio digitale. IIIF può essere considerato un analogo di HTTP dedicato alle immagini: permette lo scambio di digita-

lizzazioni fra istituzioni diverse, attraverso una URL che comunica sia l'immagine che i metadati associati. Attraverso viewer specifici come Mirador,⁵ sono disponibili diverse funzioni (Figura 1):

- lo zoom;
- la navigazione di una scansione multipagina;
- la comparazione di più immagini provenienti da archivi diversi;
- la manipolazione dell'immagine con strumenti basilari di photoediting (es. contrasto, colori ecc.);
- l'annotazione di porzioni dell'immagine, con tanto di commento a testo libero e tag del commento stesso. L'annotazione stessa diventa un oggetto digitale proprio, con URL e coordinate dell'immagine ritagliata.

Utilizzando il formato JSON-LD e i protocolli web standard elaborati dal W3C (come per esempio Web Annotation), IIIF rende semplice analizzare e condividere immagini, migrare attraverso sistemi tecnologici diversi, e offrire un accesso arricchito a ricercatori e umanisti.

Digitalizzazione grandi formati

Come appena detto, il protocollo IIIF è perfetto per la visualizzazione di immagini in altissima definizione. Per queste, sarà adottata una politica di digital preservation in linea con i programmi per la digitalizzazione Metamorfoze (Paesi Bassi) e FADGI (USA). Verrà inoltre mantenuto lo standard ISO 19264-1.⁶ I master file verranno archiviati utilizzando il formato aperto DNG per i singoli scatti, e il TIFF per le ricostruzioni finali: con questo approccio, le informa-

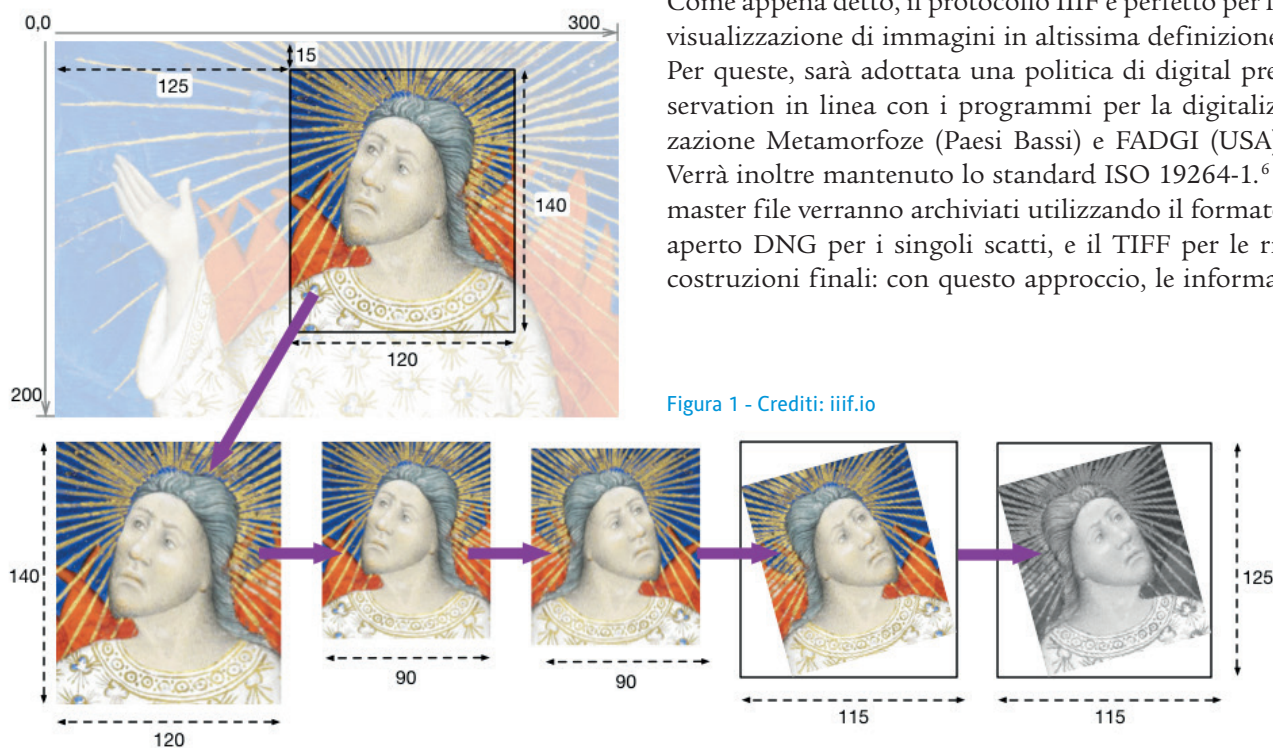


Figura 1 - Crediti: iiif.io



Figura 2

zioni vengono preservate, permettendo in futuro di ri-calibrare e sviluppare nuovamente i DNG per generare nuovamente i surrogati digitali.

I documenti di grandi dimensioni (appartenenti al fondo cartografico) verranno digitalizzati con una macchina modificata ad hoc per il progetto da Franco Cosimo Panini e Haltadefinizione. La macchina è stata utilizzata in precedenza per la scansione ad altissima definizione di dipinti come *La Nascita di Venere* di Botticelli o *Giuditta e Oloferne* di Caravaggio, che erano appesi in verticale. Modificando la struttura come in figura, sarà possibile ruotare la struttura in orizzontale e permettere la scansione delle mappe appoggiate su un carrello. La macchina è capace di una risoluzione a 4000 ppi, permettendo uno zoom 40X (un centimetro sulla mappa diventa grande 40 centimetri) (Figura 2).

Digitalizzazione altri formati

Mentre per la digitalizzazione dei grandi formati si pongono problemi principalmente di natura tecnologica (strumentazione da utilizzare, maneggevolezza dei supporti, dimensioni delle immagini prodotte ecc.), per i formati inferiori all'A2 gli aspetti delicati da affrontare sono quelli di natura logistica e organizzativa, ovvero la definizione di un processo e di modalità e tempistiche che consentano la lavorazione di una quantità decisamente maggiore di documenti, pur nel rispetto di tutte le misure di sicurezza da

adottare per la conservazione dei materiali originali.

Per quel che riguarda il fondo musicale – circa 21.000 documenti in 3.800 fascicoli, per una stima di circa 280.000 pagine – verrà usata una macchina Copybook V-shape, che può gestire documenti dal formato A5 all'A2. La macchina permette di aprire i volumi limitatamente, a 120°, e possiede una illuminazione “fredda”, senza raggi infrarossi o ultravioletti, che possono danneggiare il materiale

più fragile. La risoluzione sarà a 600 dpi per tutti i formati uguali e inferiori all'A4, e a 400 dpi per quelli più grandi. L'operazione verrà gestita da MIDA Informatica, ditta bergamasca che ha digitalizzato tra l'altro il *Codice Atlantico* di Leonardo per la Biblioteca Ambrosiana di Milano.

Il repository di metadati

Hyperborea, azienda con sede nel Polo scientifico e tecnologico di Navacchio (Pisa), ha in carico, oltre al coordinamento del progetto e alle attività di censimento, metadattazione e recupero dei metadati esistenti, la fornitura della nuova versione della suite Arianna, che sarà utilizzata per la metadattazione e descrizione dell'intero patrimonio delle Gallerie Estensi.

Arianna è un software per la descrizione catalografica multi-ambito di oggetti culturali, che consente di collegare la scheda descrittiva di ciascun oggetto alla sua riproduzione digitale, affiancando l'interfaccia di compilazione della scheda al viewer dell'oggetto digitale.

Supportando i differenti tracciati descrittivi utilizzati nei vari ambiti e consentendo esportazioni nei formati standard più diffusi, Arianna costituisce lo strumento ottimale per la descrizione di un patrimonio estremamente eterogeneo e complesso (manoscritti, libri antichi e moderni, fondi archivistici, cartografia, foto, stampe e disegni ecc.) come quello conservato dalla BEU e dalle Gallerie Estensi.

Arianna è un software per la descrizione catalografica

multi-ambito di oggetti culturali che, supportando i differenti tracciati descrittivi utilizzati nei vari ambiti e consentendo esportazioni nei formati standard più diffusi, costituisce lo strumento ottimale per la descrizione di un patrimonio estremamente eterogeneo e complesso (manoscritti, libri antichi e moderni, fondi archivistici, cartografia, foto, stampe e disegni ecc.) come quello conservato dalla BEU e dalle Gallerie Estensi. Arianna consente inoltre di descrivere il *contesto*, cioè l'insieme di persone, organizzazioni, luoghi, eventi che definiscono l'ambiente in cui gli oggetti culturali sono inseriti.

Al termine dello specifico intervento di digitalizzazione e metadattazione verrà svolta un'ulteriore attività di ricognizione dell'intero patrimonio – a livello di fondi e collezioni – realizzando le relative schede descrittive in Arianna e creando tutti i presupposti affinché futuri interventi di inventariazione e catalogazione di singoli fondi possano essere svolti su basi solide e andare ad alimentare un'unica coerente descrizione complessiva.

La piattaforma

La biblioteca digitale sarà basata sulla piattaforma MLOL. A discrezione della BEU, oltre alla collezione del patrimonio digitalizzato potranno essere attivate le classiche collezioni MLOL (in prestito digitale) e la collezione di Risorse Open.

La piattaforma MLOL offrirà dunque le funzioni di

- ricerca
- navigazione
- catalogo
- visualizzazione

delle risorse digitali, oltre che di gestione degli utenti. Il motore di ricerca di MLOL, recentemente reimplemenato con Apache SOLR, garantirà migliori performance, oltre che una gestione personalizzata dei *pesi* e dell'*information retrieval*, e la possibilità di gestire la ricerca – e i filtri/faccette – con schemi di metadati multipli basati sulle collezioni, andando oltre lo schema Dublin Core normalmente usato da MLOL.

MLOL infatti si interfacerà con il software Arianna per poter ricercare in tutti i metadati descrittivi presenti nella scheda di catalogo di quest'ultimo.

Su MLOL, ogni scheda sarà imperniata sul visualizza-

tore IIIF, con le sopra descritte funzionalità di navigazione, zoom, comparazione e annotazione.

Sarà dunque possibile comparare direttamente i documenti della Biblioteca Estense Universitaria con i documenti presenti nelle più grandi biblioteche del mondo: già da oggi, su tutti i portali MLOL sono presenti i *manifest IIIF* di una collezione crescente di manoscritti, mappe, spartiti provenienti da Bodleian Library, archivi svizzeri, Harvard Museum, NYPL, National Library of Wales ecc.

L'infrastruttura di annotazione potrà essere utilizzata per progetti di *crowdsourcing*, e andrà ad aiutare direttamente i progetti di ricerca sul *machine learning* del laboratorio AlmageLab.

ANDREA ZANNI

MLOL

andrea.zanni84@gmail.com

NOTE

¹ A questo proposito, un approfondimento è disponibile nell'articolo: ANDREA PALAZZI, *Il trattamento delle immagini digitali alla Biblioteca Estense Universitaria di Modena (1990-2010)*, "Digitalia", Vol. 2, 2010. URL: <http://digitalia.sbn.it/article/view/239>.

² L'archivio è disponibile all'indirizzo: http://www.internetculturale.it/it/41/collezioni-digitali/26196/archivio-lo-dovico-antonio-muratori_-carteggi-e-documenti.

³ Il *charter* iniziale comprendeva: Stanford University, British Library, Bodleian Libraries (Oxford University), Bibliothèque nationale de France, Nasjonalbiblioteket (National Library of Norway), Los Alamos National Library, e Cornell University. URL: <https://www.cni.org/topics/information-access-retrieval/international-image-interoperability-framework>.

⁴ Maggiori informazioni su: <http://www.iccd.beniculturali.it/it/159/eventi/4485/standard-per-la-digitalizzazione-del-patrimonio-culturale-e-interoperabilit>.

⁵ projectmirador.org.

⁶ <https://www.iso.org/standard/64221.html>.

DOI: 10.3302/0392-8586-201904-063-1