

Pianificare la digitalizzazione

GRUPPO DI LAVORO
DELLA BIBLIOTECA DIGITALE BEIC

Fondazione BEIC, Milano
customerservice@beic.it

Collezioni di libri rari e manoscritti nell'esperienza della Biblioteca digitale BEIC

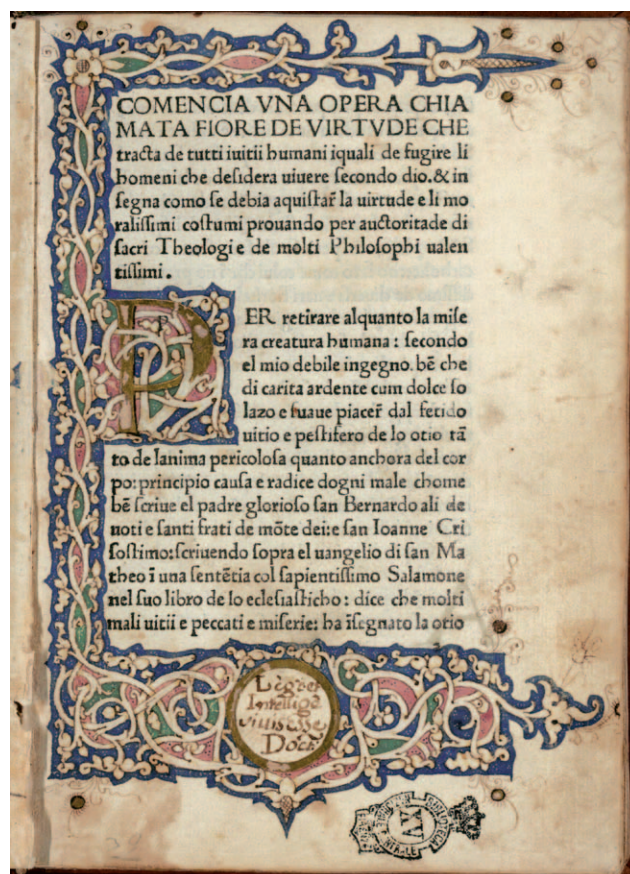
Le Linee guida per pianificare la digitalizzazione di collezioni di libri rari e manoscritti sono state tradotte nel 2015 dal gruppo di lavoro della Biblioteca digitale BEIC (Chiara Cauzzi, Chiara Consonni, Valeria De Francesca, Danilo Deana, Lisa Longhi, Federica Viazzi) con la supervisione di Angela Nuovo (IFLA Rare Books and Special Collections Section). Lo scopo della traduzione era anche quello di confrontare le procedure seguite all'interno della Biblioteca digitale BEIC¹ con quelle descritte nello standard.

Le Linee guida sono divise in otto capitoli: 1. Definizione del progetto; 2. Selezione degli originali; 3. Flusso di lavoro per la creazione della collezione; 4. Metadati; 5. Presentazione; 6. Diffusione, promozione e riutilizzo delle copie digitali; 7. Valutazione; 8. Conservazione a lungo termine delle collezioni digitali.

Si è scelto di riportare qui in corsivo uno o più paragrafi per ogni capitolo e di commentarli alla luce di quanto fatto nella realizzazione della Biblioteca digitale BEIC, così da mostrane tutte le implicazioni.

L'attento esame di una serie di questioni generali faciliterà la definizione e l'implementazione di progetti e contribuirà a garantirne il successo. Un progetto di digitalizzazione può avere come risultato differenti prodotti e servizi, per cui è necessario tenere in considerazione le risorse disponibili e definire chiaramente gli obiettivi e i risultati nella fase iniziale. Questo servirà per la promozione e la raccolta dei fondi e fornirà indicatori significativi per la valutazione finale.

Prima di avviare un progetto di digitalizzazione – qualunque sia l'impegno previsto in termini economici, professionali o di personale – è necessario individuare gli obiettivi che si intendono raggiungere. I testi da digitalizzare verranno così scelti in modo da soddisfare i



Fiore di virtù, [Venezia], Beretin Convento Ca' Grande, 1477, carta 1r

requisiti scientifici e tecnici e ai quali dovranno rispondere. La fase progettuale è di vitale importanza perché gli attori possano avere chiaro l'iter del progetto e analizzarne concretamente la fattibilità in termini economici, di strumentazioni e competenze. È bene definire, sempre durante la fase iniziale, anche le questioni burocratiche e amministrative, soprattutto quelle legate ai diritti d'uso, e una strategia di valutazione del proget-

to durante le fasi di realizzazione e sul lungo periodo. La Biblioteca digitale BEIC si pone obiettivi molto diversi rispetto a quelli che costituiscono le basi, per esempio, di Internet Archive. Quest'ultimo ha come scopo quello di fornire un accesso universale alla conoscenza, una sorta di versione *full text* di WorldCat, nata in contrapposizione a Google Books e guardando all'antica biblioteca di Alessandria. Invece tra le parole chiave usate per identificare la Biblioteca digitale BEIC – oltre a interdisciplinare, multimediale e interconnessa – c'è anche “selettiva”, e da questi punti cardine si è partiti per costruirla.

La selezione degli originali è un compito essenziale nello sviluppo di una collezione digitale e dipende principalmente dai bibliotecari specializzati in questo tipo di fondi. Le collezioni, le opere, le edizioni e gli esemplari devono essere studiati e verificati tenendo presente la portata della nuova collezione digitale. Sono da tenere in considerazione anche cronologia, geografia, autore, soggetto, formato, possessore ecc.

La Biblioteca digitale BEIC si è affidata a studiosi e a istituzioni di prestigio per la selezione degli esemplari da digitalizzare, allo scopo di coprire gradualmente i principali rami del sapere umanistico e scientifico. Preliminarmente, ha però risposto a quelle che secondo le *Linee guida* sono le domande che possono aiutare a individuare la portata del progetto e contribuire al processo di scelta. La prima domanda riguarda la volontà di individuare e digitalizzare collezioni già esistenti, oppure di crearne di nuove da collezioni reali appartenenti a diverse istituzioni, ma con elementi in comune (provenienza, possessore, lingua e così via).

In alcuni casi la Biblioteca digitale BEIC ha digitalizzato interi fondi o parte di essi, come il “Fondo Albrecht von Haller”, donato alla Biblioteca Nazionale Braidense di Milano dall'imperatore Giuseppe II d'Austria, che conserva 15.000 volumi e circa 150 manoscritti provenienti dalla Biblioteca di Albrecht von Haller, professore di anatomia, chirurgia e botanica all'Università di Gottinga nel XVIII secolo. Talvolta ha invece creato nuove collezioni virtuali digitalizzando esemplari provenienti da diverse istituzioni, com'è stato il caso della collezione “Incunaboli italiani in lingua volgare”, per cui è stata stipulata una convenzione con il Dipartimento di Italianistica e Spettacolo di Sapienza Università di Roma per la digitalizzazione di più di 1.600 incunaboli, risultanti da una selezione attuata sulla base dell'Incunabola Short Title Catalogue della British Library.

Il processo di creazione di una collezione digitale può essere diviso in varie fasi o stadi. Ogni istituzione può organizzare queste attività in maniera leggermente diversa, ma i processi di base rimangono gli stessi. Tutte le procedure dovrebbero essere adeguatamente documentate. Se il disegno generale del progetto e i compiti dei partecipanti sono definiti e conosciuti con chiarezza da tutti, possono essere prevenuti errori e fraintendimenti.

La prima fase comprende l'ispezione e la preparazione degli originali e la verifica dei metadati esistenti nel catalogo (associati ai documenti). La seconda fase riguarda la digitalizzazione vera e propria. Prima di procedere devono essere definite le caratteristiche dell'immagine digitale, in modo che sia il più possibile fedele all'originale (inclusione dei margini, scalatura, dimensioni delle immagini, parametri di risoluzione e colore ecc.). La terza fase riguarda l'elaborazione delle immagini: il controllo di qualità delle acquisizioni, l'implementazione di funzionalità aggiuntive che migliorino l'accesso e la fruizione (per esempio l'OCR o la georeferenziazione), infine il trasferimento delle immagini in un archivio per la loro conservazione a lungo termine.

La Biblioteca digitale BEIC ha una caratteristica che la distingue dalle istituzioni per cui sono pensate le *Linee guida*: i materiali rari e i manoscritti oggetto delle campagne di digitalizzazione non sono parte del patrimonio della Fondazione stessa, ma sono messi a disposizione dalle biblioteche o dagli enti che conservano gli originali a seguito di accordi specifici.

La fase successiva è affidata ad aziende specializzate; la scelta di esternalizzare l'acquisizione delle immagini è una delle possibili strade che si possono prendere avviando un progetto di digitalizzazione: per la Biblioteca digitale BEIC è sembrata la via migliore, viste le dimensioni della Fondazione e i costi di gestione che avrebbe avuto una sezione interna dedicata a questo compito.

La terza fase è invece completamente effettuata all'interno della Biblioteca digitale BEIC e prevede il controllo di qualità delle immagini. L'obiettivo è quello di rilevare eventuali lacune e verificare la correttezza dei metadati tecnici associati alle immagini. Per quanto riguarda le procedure aggiuntive per migliorare l'accesso alle risorse digitali, è prassi procedere al riconoscimento dei caratteri tutte le volte che il testo originale lo permette (quindi evitando, ad esempio, i testi in greco o in Fraktur).

L'ultimo passo è il trasferimento delle immagini e dei metadati associati in un archivio per il controllo periodico e la conservazione a lungo termine, nel caso della Biblioteca digitale BEIC l'Archivio delle collezioni digitali.

La comunità bibliotecaria ha storicamente usato i metadati bibliografici, ovvero la catalogazione, così come alcuni tipi di metadati strutturali e amministrativi per la gestione e l'accesso alle collezioni fisiche. Oggi i bibliotecari stanno creando modelli di metadati aggiornati che non solo favoriscano l'accesso agli oggetti digitali ma forniscano allo stesso tempo le informazioni necessarie per la conservazione a lungo termine e per l'accesso facilitato agli stessi oggetti digitali attraverso i sistemi di ricerca in rete.

Le *Linee guida* affermano chiaramente che una buona metadattazione dei materiali digitalizzati favorisce, anzi assicura, sia l'accesso alle collezioni digitali sia la loro conservazione e stabilità nel tempo, costituendo, per così dire, lo scheletro del progetto.

La Biblioteca digitale BEIC utilizza per la metadattazione lo standard METS (Metadata Encoding and Transmission Standard): si tratta di uno schema XML progettato dalla Library of Congress per codificare i metadati necessari alla gestione degli oggetti di una biblioteca digitale. La struttura dello schema è molto articolata, ma al suo interno si possono facilmente individuare le sezioni dedicate alle quattro tipologie di metadati di cui parlano le *Linee guida*: bibliografici, strutturali, tecnici e amministrativi.

Nella Biblioteca digitale BEIC i metadati bibliografici sono codificati in MARC 21. Le *Regole italiane di catalogazione* (REICAT), MARC 21 e ISBD costituiscono i riferimenti della procedura di catalogazione.

Rispetto alle pratiche descritte nelle *Linee guida* – dove l'iter consigliato è prima la metadattazione, poi la digitalizzazione e infine il collegamento dell'oggetto digitale al catalogo – nella Biblioteca digitale BEIC le prime due fasi sono invertite: prima c'è la ricezione delle immagini, poi si procede alla creazione dei metadati bibliografici. Questo si giustifica tenendo presente la particolare natura della Biblioteca stessa.

Per ciascun oggetto digitale viene redatta una scheda approfondita che non si limita alla descrizione tradizionale della risorsa, ma che fa in modo che la singola registrazione si inserisca nel contesto di catalogo organizzato, coerente ed amichevole. Il catalogo si estende quindi in due dimensioni: ampiezza e profondità. In ampiezza, poiché sfrutta tutti gli strumenti di raggruppamento delle registrazioni che permettono una navigazione più agile e risultati di ricerca più efficaci in un'ottica *FRBR-oriented* (le faccette e soprattutto l'uso estensivo del titolo uniforme per raggruppare le diverse manifestazioni ed espressioni delle opere); in profondità



Pietro Andrea Mattioli, *Commentarii in sex libros Pedacii Dioscoridis Anazarbei de medica materia*, Venezia, Vincenzo Valgrisi, 1565

tà, poiché per ogni opera digitalizzata si procede a una catalogazione di dettaglio, che evidenzia le opere contenute nella pubblicazione, dedicando a ciascuna una registrazione analitica, dalla quale si raggiunge il punto di inizio dell'opera nell'esemplare digitale.

La metadattazione descrittiva va di pari passo con la costruzione della mappa strutturale, cioè con la creazione della seconda tipologia di metadati trattati nelle *Linee guida*.

La mappa strutturale mira a restituire un'immagine chiara e il più possibile fedele all'originale dell'opera digitalizzata, sia dal punto di vista fisico sia per quanto riguarda le suddivisioni logiche. Sono evidenziati, per esempio, gli *ex libris*, i timbri e le note; la numerazione delle pagine è riportata fedelmente, incluse le pagine bianche (come prescrivono le regole di catalogazione del libro antico), gli errori di numerazione sono segnalati con l'indicazione del numero di pagina corretto

inserito tra parentesi quadre e preceduto dalla locuzione *id est* (i.e.). Nell'eventualità in cui l'intera opera non presenti numerazione, si procede all'indicazione della segnatura dei fascicoli o, come *extrema ratio*, all'attribuzione dell'intera sequenza numerica.

Un protocollo di inserimento dei metadati così meticoloso, che coinvolge le immagini di tutta l'opera digitalizzata, svolge una doppia funzione: la principale è quella di fornire informazioni il più possibile esaustive e accurate al lettore, ma costituisce anche una forma di controllo incrociato con i dati catalografici e le procedure di controllo delle immagini.

Una buona descrizione bibliografica e una solida mappa strutturale hanno bisogno, per esser il più possibile condivisibili con altre istituzioni culturali, di ulteriori metadati, non visibili dai lettori: quelli tecnici e amministrativi.

I metadati tecnici descrivono le caratteristiche delle immagini acquisite: dimensione in pixel e in byte, campionatura, compressione, risoluzione, strumentazione uti-

lizzata, data di acquisizione. Tutti questi metadati sono solitamente generati in automatico al momento della digitalizzazione e sono codificati secondo lo schema MIX (Metadata for Images in XML Standard).

I metadati amministrativi sono anch'essi uno strumento indispensabile per la gestione delle risorse digitali e vanno incontro alle esigenze di accesso degli utenti e degli amministratori per offrire maggiori informazioni sulla creazione dell'oggetto, la sua gestione ed eventuali restrizioni d'uso.

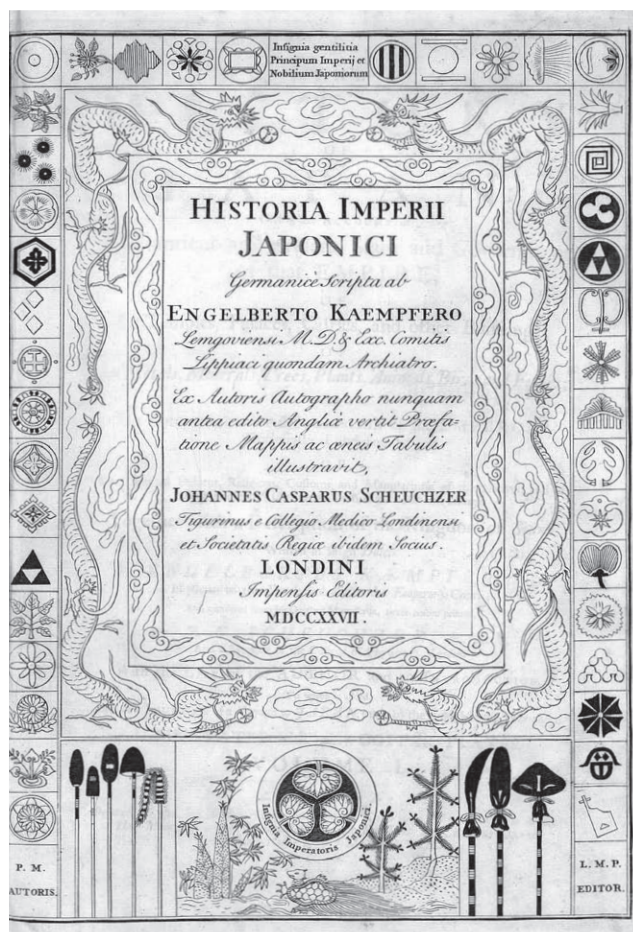
Ciò che gli utenti si aspettano da una copia digitale è l'accesso libero e gratuito, la facile localizzazione delle opere attraverso i comuni motori di ricerca, l'interazione semplice e la presentazione attraverso browser e plug-in di uso comune, diverse opzioni di visualizzazione (inclusa la presentazione a doppia pagina e lo zoom), l'esistenza di marcatori (principalmente per il successivo recupero del testo desiderato), la possibilità di fare annotazioni, stampare i documenti, scaricarli, riutilizzarli e combinarli. I formati di presentazione dovrebbero perciò essere facilmente consultabili, scaricabili e manipolabili.

La Biblioteca digitale BEIC ha dedicato molta attenzione alla messa a punto di un catalogo amichevole particolarmente curato nella presentazione dei dati e nella possibilità di raffinare la ricerca secondo filtri efficaci. Le registrazioni bibliografiche sono state inoltre collegate ad altri cataloghi e risorse nazionali e internazionali (SBN, ACNP, ISTC, GW). Si è infine pensato di predisporre uno spazio personale che consenta all'utente di salvare e organizzare i risultati individuati durante la sessione di lavoro.

Al catalogo sono stati affiancati i navigatori, ovvero una rappresentazione grafica dei dati creata per permettere all'utente di consultare e interpretare set complessi di dati in un formato facilmente comprensibile e manipolabile.

Per la visualizzazione dei materiali la Biblioteca digitale BEIC utilizza uno strumento specifico che ha la capacità di racchiudere in un'unica interfaccia funzionalità diverse: Mets Viewer permette infatti di visualizzare la mappa strutturale, ridimensionare la pagina a seconda delle necessità, mostrare due pagine affiancate per facilitare la lettura, vedere tutte le immagini del documento in un'anteprima per captare rapidamente i particolari visivamente più immediati (come la presenza di illustrazioni e miniature) e selezionare direttamente la pagina di interesse.

Secondo le *Linee guida* infatti la presentazione deve per-



Engelbert Kaempfer, *History of Japan*, London, Thomas Woodward, 1728, frontespizio calcografico

mettere di individuare facilmente in che punto del documento ci si trovi e di tornare altrettanto facilmente ai punti chiave (frontespizio, inizio dei capitoli, colophon ecc.), ricreando l'esperienza della lettura dell'originale o addirittura superandola grazie a funzionalità aggiuntive.

La maggior parte delle biblioteche e dei progetti sovvenzionati hanno l'obbligo di garantire l'accesso globale a tutte le collezioni digitalizzate, accesso che spesso comprende la distribuzione dei metadati, dei collegamenti e talvolta anche delle stesse immagini.

Come politica generale la Biblioteca digitale BEIC rilascia i suoi metadati descrittivi con licenza Public Domain Dedication (CC0 1.0 Universal), che prevede la rinuncia a tutti i diritti, nella misura consentita dalla legge; è quindi possibile copiare, modificare, distribuire e utilizzare, anche a fini commerciali, tutti i metadati senza alcuna limitazione. Questo ha fatto sì che le registrazioni bibliografiche della Biblioteca potessero essere raccolte da grandi aggregatori come Europeana ed essere interrogabili attraverso i principali motori di ricerca. La riproduzione digitale delle opere è invece solitamente distribuita con una licenza Creative Commons (CC BY-NC-ND).

Un modo per incrementare accesso e visibilità delle collezioni è l'uso di portali e altri aggregatori, che possono essere internazionali, nazionali o tematici.

Una biblioteca digitale senza accessi è destinata a morire e uno dei punti cardine per mantenere e incrementare le ricerche è rendere il più possibile visibili e accessibili le collezioni digitali. Le *Linee guida* raccomandano di sfruttare le potenzialità di aggregatori, portali tematici più o meno generalisti (inclusi i social network) e di arricchire le collezioni digitali di note, commenti e altre informazioni ad alto valore aggiunto.

Le strategie di diffusione delle informazioni e promozione delle collezioni possono essere molteplici; le strade intraprese dalla Biblioteca digitale BEIC sono essenzialmente due: la prima, vista la scientificità con la quale vengono trattate le informazioni, è quella di avviare delle *partnerships* con istituzioni specializzate, come il Consorzio delle Biblioteche Europee di Ricerca (CERL), la British Library e la Staatsbibliothek di Berlino; la seconda prevede invece una collaborazione con Wikimedia Italia, così da raccogliere i frutti dell'unione tra la capillarità di Wikipedia e la scientificità della Biblioteca digitale BEIC.



Jacques Ozanam, *Nouveaux elements d'algebre*, Amsterdam, George Gallet, 1702, antiporta

La promozione passa anche attraverso progetti paralleli e mostre virtuali, come il portale dedicato all'idraulica.

La maggior parte delle biblioteche desidera valutare l'esito dei metodi di digitalizzazione utilizzati, così come la diffusione e l'impatto dei progetti e dei programmi di digitalizzazione. Molte organizzazioni hanno sviluppato strumenti che facilitano il processo di valutazione. Le statistiche relative alla produzione e all'uso di copie digitali sono in grado di fornire un buon punto di partenza per la valutazione quantitativa che tenga conto di aspetti come il numero di libri/oggetti digitalizzati, il numero di visite alle pagine del portale, il numero di volte in cui un oggetto digitale è stato visualizzato o scaricato, il numero di volte in cui è stato citato o collegato ecc.

Per la Biblioteca digitale BEIC la valutazione è un elemento fondamentale. Essa infatti permette da un lato



Neu-eröffnetes Italien, Nürnberg, Johann Leonhard Buggel, 1703

di comprendere le dinamiche proprie di una biblioteca digitale, dall'altro di pianificare al meglio i progetti futuri. La politica in atto presso la Biblioteca è quella del *constant users involvement*, ovvero della ricerca costante di un feedback da parte degli utenti durante l'intero svolgimento del progetto.

Vengono anche utilizzati dati quantitativi raccolti periodicamente sull'accesso alle collezioni digitali, che seguono il principio dell'anonimizzazione dei dati, ovvero il mascheramento della fisionomia del singolo utente per proteggerne la privacy.

Le statistiche, relative per esempio ai termini più ricercati nel catalogo, permettono di orientare i progetti futuri, suggerendo altre possibili campagne di digitalizzazione o mettendo in luce le collezioni che hanno bisogno di maggiore promozione per essere apprezzate.

Le istituzioni dovrebbero sviluppare una strategia per la conservazione a lungo termine delle loro collezioni digitali, soprattutto alla luce del costo della digitalizzazione, dell'investimento di personale e del possibile danno causato a materiali unici e rari.

La Biblioteca digitale BEIC, con l'aiuto del Politecnico di Milano, ha realizzato un Archivio delle collezioni digitali il cui scopo è quello di assicurare la conservazione a lungo termine dei materiali digitalizzati (i documenti XML contenenti i metadati relativi al documento e alle immagini frutto della digitalizzazione e le immagini stesse) che, prima di essere archiviati, vengono esaminati così da assicurarne la coerenza e la completezza. L'Archivio, che sfrutta applicativi utilizzati in contesti che richiedono un'altissima affidabilità, è stato progettato sulla base del modello OAIS (Open Archival Information System), lo standard ISO 14721:2003 che definisce concetti, modelli e funzionalità degli archivi digitali.

La traduzione delle *Linee guida* ha permesso al gruppo di lavoro della Biblioteca digitale BEIC un confronto che da un lato ha consentito di riconoscere come alcune pratiche messe a punto con l'esperienza siano conformi a uno *standard* internazionale, dall'altro ha rappresentato l'occasione di migliorarne altre, nella convinzione che le biblioteche digitali, come quelle tradizionali, siano un organismo in continua evoluzione.

Si ringrazia la Fondazione BEIC per avere messo a disposizione il materiale iconografico che correde questo articolo.

NOTE

¹ <www.beic.it/articoli/biblioteca-digitale>.

DOI: 10.3302/0392-8586-201507-007-1

ABSTRACT

In 2015, Beic Digital Library working group translated the *Guidelines for Planning the Digitization of Rare Book and Manuscript Collections* into Italian. The article compares the policies in place at Beic Digital Library with those prescribed by the standard in order to acknowledge similarities and reflect upon dissimilarities.