



Il contesto dei metadati e i profili applicativi nella struttura globale dei cataloghi e nei progetti di digitalizzazione della Biblioteca Apostolica Vaticana

Paola Manoni

Il tema dell'interoperabilità tecnica si dispiega nella realtà informatica della Biblioteca Vaticana in un ambito che riguarda l'aggregazione dei dati bibliografici (provenienti da sistemi con strutture di metadati diversi) e l'interconnessione dei dati strutturati (prevista nella condivisione dei programmi cooperativi) in seno al recente progetto di digitalizzazione delle collezioni dei fondi manoscritti e di incunaboli. La relazione ha lo scopo di illustrare il contesto, i modelli applicativi e i sistemi attualmente in uso, puntando l'attenzione sull'intelaiatura dei metadati coinvolti nella struttura globale dei nuovi cataloghi della Vaticana, inaugurati nel web il 16 maggio 2012, e l'organizzazione della biblioteca digitale, di prossima pubblicazione nel web. Come discorso preliminare è opportuno spendere qualche parola sull'attuale stato degli sviluppi informatici e catalografici della Vaticana, al fine di poter collocare con maggiore dettaglio la realtà della Biblioteca nel più ampio scenario progettuale

dell'interoperabilità globale nel settore bibliotecario.

Partendo dal presente e andando a ritroso, la Biblioteca Apostolica Vaticana (BAV)¹ ha sentito l'esigenza di far interagire le diverse basi di dati bibliografiche, stipulando un metodo di aggregazione di dati e di sintassi eterogenee, per la descrizione delle diverse collezioni che compongono il patrimonio della Vaticana. È stato dunque pensato un sistema per stabilire relazioni tra i diversi cataloghi, al fine di consentire una ricerca integrata (e non federata) all'interno della piattaforma di un catalogo generale. Senza ripercorrere la storia dei cataloghi elettronici della Vaticana ma, solo per dare qualche informazione rapida, sono oggi disponibili alla consultazione le basi di dati relative a:

- stampati (monografie e periodici di testo): il primo catalogo elettronico costruito alla metà degli anni Ottanta dello scorso secolo su un primo nucleo di catalogazione corrente, proseguito con la completa immissione dei dati retrospettivi, sottoposti poi a revisione (con particolare riguardo ai punti di accesso);
- materiali grafici (stampe, disegni, fotografie): nel 1998 è iniziato il progetto "Stampe on line", che comprende la catalogazione analitica delle stampe così come dei disegni. Dal 2001 si procede alla scansione digitale delle immagini ed il loro legame ipertestuale con le schede bibliografiche;
- monete e medaglie: progetto in corso d'opera, avviato nel 2001, a tutt'oggi comprende descrizioni numismatiche fino al XVI secolo. Come per i materiali grafici, per la quasi totalità dei dati sono presenti immagini digitali: le riprese fotografiche del dritto e del rovescio, per ciascuna unità descrittiva;

¹<http://www.vaticanlibrary.va>.

- incunaboli: catalogo elettronico di recente costituzione (disponibile dal 2009) che assomma le descrizioni complete dell'inventario curato da P. William Sheehan, pubblicato nel 1997. È in corso realizzazione il progetto denominato BAVIC (Bibliothecae Apostolicae Vaticanae Incunabulorum Catalogus) che consiste nella catalogazione analitica degli esemplari.

I cataloghi sopra menzionati sono strutturati nella sintassi MARC21 mentre, per quanto riguarda l'allestimento dei cataloghi elettronici dei manoscritti e dei complessi documentari archivistici (questi ultimi presenti nella Sezione Archivi del Dipartimento dei Manoscritti), ci riferiamo ad un altro tipo di trattamento informatico e catalografico in relazione, sia al formato dei dati, sia all'architettura di sistema e, come diremo, loro interoperabilità.

- Manoscritti: a partire dal 2002 è operante un progetto di conversione retrospettiva dei cataloghi ed inventari cartacei dei fondi. Il progetto è stato articolato in diverse fasi. Preliminarmente, per la definizione dei criteri di elaborazione dei dati, è stata scelta la conformità agli standard TEI (del consorzio Text Encoding Initiative). La catalogazione elettronica si conforma pertanto agli elementi stabiliti nel formato di codifica TEI-MS – specifico per la descrizione dei manoscritti e dunque, da un punto di vista gestionale, è stato progettato un database in grado di gestire in linguaggio XML tale struttura. Il sistema, ideato dal Coordinamento dei servizi informatici della Vaticana è denominato InForMA (Informatics For Manuscripts and Archives) il quale gestisce la produzione e la ricerca dei dati sui manoscritti e sulle carte di archivio.
- Materiale archivistico: quest'ultimo, come dicevamo, è formalizzato nel medesimo linguaggio XML ma secondo standard EAD (Encoding Archival Description). Il sistema infatti può

gestire distinte collezioni di dati ovvero documenti che fanno riferimento a schemi di metadati diversi, consentendone al contempo sia una gestione separata (vari formati per fattispecie differenti) che correlata – come ad esempio nei casi della condivisione delle liste di voci di autorità (per nomi, titoli e descrittori semantici) e dei collegamenti fra schede bibliografiche e file di immagini digitali.

A partire dal 2011, si è avviato uno studio per l'adozione di un nuovo sistema di gestione di OPAC in grado di aggregare in un catalogo generale i contenuti dei cataloghi individuali ma anche di gestire strutture di indici separate e referenziate a ciascun tipo di catalogo e, ancora, di collegare ai record bibliografici le altre risorse informative disponibili in Biblioteca; in ultimo, di gestire informazioni utili per i ricercatori, a latere della mera interrogazione del catalogo. Lo strumento che è stato costruito consente l'alimentazione regolare dei dati provenienti dai diversi cataloghi in un repository nel quale i diversi formati di dati ovvero MARC21, TEI-MS e EAD sono conservati, ciascuno per la semantica di riferimento. Il sistema è conforme ai protocolli di interoperabilità quali OAI-PMH e OpenURL. Vorrei spendere qualche parola sull'impiego in BAV di questi due protocolli, con particolare riguardo al primo per l'esperienza di aggregazione dei dati per il catalogo generale. Come è noto, l'obiettivo principale del protocollo, è di consentire a due o più archivi di dati separati di scambiare metadati. OAI-PMH definisce il protocollo di comunicazione attraverso regole e metodi per il trasferimento dei metadati, costituito secondo le due classiche entità di base: il data provider e il service provider. Il data provider fornisce l'OAI-PMH in modo da pubblicare i propri metadati. Il service provider invia all'OAI-PMH le richieste al data provider via http e raccoglie i metadati. Il data provider risponde con un messaggio XML. Nelle problematiche dell'harvesting dei dati, più che sull'impiego di MARCXML per i dati

catalografici in MARC21, che evidentemente non ha posto problemi di sorta – data l'ampia sperimentazione su cui esso poggia, vale la pena soffermarsi sul trattamento dei metadati per archivi e manoscritti implicati nella gestione del catalogo generale della Vaticana. L'alimentazione regolare del catalogo generale per le descrizioni di manoscritti e archivi codificati nei formati EAD e TEI avviene secondo un formato di conversione studiato per mantenere la medesima organizzazione e granularità strutturale, rinunciando in tal senso all'interazione via OAI-PMH, stante la difficoltà a gestire crosswalk soddisfacenti, che evitino la riduzione o la perdita d'informazione nella rappresentazione dei dati. Si pensi in proposito all'organizzazione gerarchica delle unità descrittive nell'EAD, alla ricchezza degli attributi e alla complessità degli elementi annidati, in entrambe le codifiche EAD e TEI-MS. La letteratura sull'argomento abbonda di valutazioni sul Dublin Core (anche in versione qualificata) con cui il protocollo rappresenta i dati che aggrega, quale formato di mediazione troppo generico per le tipologie di metadati a cui ci stiamo riferendo. Nella costruzione del catalogo generale, è stata effettuata una scelta informatica di tipo ibrido poiché, accanto al funzionamento protocollare dell'harvesting per i cataloghi in MARC21, è stato realizzato un webservice secondo un modello di interazione peer-to-peer che dal sistema InForMA passa i dati EAD e TEI-MS al catalogo generale, garantendo anche in questo secondo caso l'interoperabilità fra sistemi eterogenei. Al di là della tecnica di harvesting, il sistema che gestisce i nuovi cataloghi, realizzato con le soluzioni tecnologiche offerte da Infor, offre un sistema d'indicizzazione che consente la rappresentazione unitaria dei risultati di una ricerca. In termini pratici significa che il framework di interoperabilità per la comunicazione tra i cataloghi, ha funzione di dispositivo per la raccolta di dati eterogenei, i quali confluiscono all'interno del medesimo sistema in cui, in tempo reale, vengono raccolte le descrizioni bibliografiche,

di diverso formato, provenienti dagli specifici cataloghi. In altre parole, l'interazione avviene all'interno di una base di dati, il cui modello concettuale si fonda sulle nozioni informatiche di oggetti, classi, attributi e metodi, dove le classi e gli attributi sono usati per descrivere gli aspetti strutturali (i metadati per ciascun dominio di applicazione) mentre i metodi sono usati per rappresentare le funzioni degli oggetti (unità informative trattate) desunti dalle diverse realtà catalografiche. Se nella problematica dell'interazione si volessero situare anche altre tipologie di aggregazione di informazioni, nella valutazione della struttura del front-end del nuovo sistema della Vaticana, si vede come, per ciascuna unità informativa dedicata ai diversi cataloghi, l'interrogazione dell'OPAC è offerta attraverso un elemento di interfaccia tra le diverse widgets proposte che offrono contenuti testuali, visivi, multimediali pertinenti ai distinti ambiti documentari. L'applicazione presenta al pubblico un funzionale complesso organizzativo, riconducibile al modello del portale, costituito da pagine web dedicate a ciascuna attività catalografica, OPAC interrogabili attraverso indici, distinti per ciascun catalogo e tipologia di materiale. La ricerca presenta anche la possibilità di effettuare selezioni post-coordinate come l'associazione di categorie e concetti analoghi, attraverso un reticolo automatico di collegamenti che propongono allo studioso i documenti disponibili del medesimo autore o del medesimo editore in altra lingua, altra edizione o altro supporto etc.; tecniche di scorrimento di liste di risultati mediante il raffinamento progressivo attraverso categorie, proprietà e attributi. E ancora, ricerche per logica fuzzy² che consente l'analisi del termine ricercato nelle sue distinte parti (radice, prefissi, suffissi, ortografie varianti, etc.) così da suggerire altri risultati affini allo studioso che compie una ricerca. La ricerca, a partire da un record bibliografico, è

²Logica polivalente (letteralmente 'sfumata'), estensione della logica booleana ovvero la categorizzazione in sotto ranghi di una variabile continua.

ulteriormente estendibile alla collezione delle banche dati elettroniche disponibili in BAV o altre risorse disponibili via web mediante un link resolver che si conforma allo standard openURL, protocollo per lo scambio di metadati, finalizzato alla gestione di servizi del cosiddetto “linking contestualizzato”, largamente impiegato nelle banche dati bibliografiche, nei sistemi citazionali accademici e nell’open archive. È infine possibile gestire meccanismi di interazione con RSS feed per ora non attivati, così come la funzione di word cloud. Entrambi saranno eventualmente consentiti per le esigenze dei singoli utenti, i quali potranno accedere al sistema con autenticazione (quando sarà stabilita l’attivazione del profilo). L’opzione per la consultazione dell’applicativo, mediante immissione delle credenziali dell’utente, consentirà anche la personalizzazione della visualizzazione delle informazioni e delle funzionalità degli OPAC (p.e. eliminazione di widget, ordinamento o riduzione degli indici di ricerca, etc.). Da un punto di vista dell’aggregazione dei dati e dell’interoperabilità tra sistemi in BAV, si potrebbe ulteriormente estendere la trattazione con l’interazione tra i cataloghi e l’implementazione RFID realizzata in Vaticana e questa con il sistema di sicurezza³ che traccia i movimenti dei libri e delle persone all’interno dei possibili percorsi in Biblioteca ma, per non abbandonare il tema centrale del nostro discorso, non mi spingerò oltre passando invece ad analizzare il tema dell’interoperabilità dal punto di vista dei progetti di digitalizzazione della BAV. Su questo argomento, la Biblioteca sta mettendo in pratica la digitalizzazione sia nell’ottica della conservazione a lungo termine, sia nella costituzione di una biblioteca digitale fruibile via web e attraverso i collegamenti con i cataloghi. Lasciando da parte il primo aspetto, la cui organizzazione o specifico impiego di metadati non ha evidentemente una diretta implicazione con l’argomento che stiamo oggi discutendo,

³Sistema ad uso interno, a tutela della privacy.

consideriamo invece l'assetto del secondo obiettivo. Devo subito aggiungere che, mentre il primo progetto (la conservazione a lungo termine) ha già preso il suo avvio, il secondo non ha ancora prodotto evidenza pubblica ma si prevede la presentazione nel web del primo nucleo di dati e oggetti digitali nel corso del prossimo anno. Senza alcun timore di smentita, si può affermare che curare una biblioteca di oggetti digitali significa gestirne, necessariamente, i relativi metadati. E le scelte che in tal senso si sono compiute in BAV vertono sulla gestione degli oggetti digitali secondo modello METS (Metadata Encoding and Transmission Standard) quale contenitore delle informazioni per codificare in XML i metadati necessari per la gestione degli oggetti della biblioteca digitale. Inoltre, a ciascun oggetto digitale fruibile attraverso profilo METS viene assegnato un URL persistente. Per i progetti nati in seno ad accordi cooperativi, stipulati con altre istituzioni a carattere internazionale, che vedono la costituzione di biblioteche digitali condivise con la BAV, gli oggetti digitali sono espressi nei grafi RDF quale approccio per l'interoperabilità strutturale, dovendo gestire un collegamento tra oggetti digitali appartenenti a collezioni in diverse biblioteche. Da questo si ricava una descrizione di oggetto digitale composito, formato da un insieme di risorse distribuite nel web indipendente dalla loro allocazione. Per tale attività ogni aggregazione sarà descritta in una *resource map* che esprime le relazioni semantiche esistenti tra le risorse dell'aggregazione. È previsto che ad ogni mappa sia associato un identificatore (URI) che quando verrà invocato (dereferenziato) con una richiesta http, ne fornirà la rappresentazione serializzata in base al formato standard RDF/XML. Tali specificazioni tecniche fanno riferimento, in prima istanza, a quanto previsto per l'importante opera di digitalizzazione della Vaticana in compartecipazione delle Bodleian Libraries di Oxford, resa possibile grazie al contributo della Polonsky Foundation, impegnata a sostenere le iniziative che

forniscono accesso e conoscenza dei beni culturali dell'umanità, conservati nelle grandi collezioni librarie nel mondo. In questo specifico piano di lavoro si prevede, in cinque anni, la digitalizzazione di un milione e mezzo di pagine complessive, fra i volumi scelti da ambedue le istituzioni relativi a manoscritti (greci ed ebraici) ed incunaboli. Sicché, tra i primi tesori che saranno connessi nel web tramite URI dereferenziabili saranno, tra gli incunaboli, i famosi *De Europa* di Pio II Piccolomini, la Bibbia latina delle 42 linee di Johann Gutenberg. Mentre per i manoscritti ebraici della Vaticana, una delle più importanti collezioni di codici ebraici esistenti anche se non fra le più ampie, sarà scelto il *Sifra* scritto fra la fine del IX e la metà del X secolo, probabilmente il più antico codice ebraico giunto sino a noi, un'intera Bibbia scritta in Italia intorno all'anno 1100 e ancora: commentari biblici, Halakhah e Kabbalah, commentari talmudici, e scritti di ambito liturgico, filosofico, medico e astronomico. I primi manoscritti greci che entreranno infine nella logica inferenziale dei linked data saranno importanti testimoni delle opere di Omero, Sofocle, Platone, Ippocrate, oltre a codici del Nuovo Testamento e dei Padri della Chiesa, molti dei quali riccamente decorati con miniature bizantine.

Ai fini di una corretta indicizzazione, si invitano i lettori a citare esclusivamente il testo in lingua inglese; l'unico, infatti, che presenta l'indicazione del numero di pagina, l'abstract, le keywords e le date del processo redazionale.

Manoni, P. "Metadata framework and application profiles in the global structure of catalogs and digitization projects of the Vatican Library". *JLIS.it*. Vol. 4, n. 1 (Gennaio/January 2013): Art: #5516. DOI: [10.4403/jlis.it-5516](https://doi.org/10.4403/jlis.it-5516). Web.

