

Dig *Italia*

Anno VI, Numero 1 - 2011

Rivista del digitale nei beni culturali

ICCU-ROMA

Dolly: un software per la gestione della Biblioteca Digitale Lombarda

Agnese Galeffi – Salvatore Vassallo¹

Università degli Studi di Pavia

Con la presente relazione gli autori intendono illustrare le diverse fasi e i risultati di un progetto finalizzato alla realizzazione di una serie di strumenti utili alla costituzione della Biblioteca Digitale della Lombardia. Partner scientifico della Regione Lombardia è stato il Dipartimento di Scienze Storiche e Geografiche “C.M. Cipolla” dell’Università di Pavia.

Il progetto per una biblioteca digitale lombarda

Il progetto Biblioteca Digitale della Lombardia si innesta su una serie di iniziative, avviate sin dal 2007, volte a potenziare e a integrare i siti web regionali settoriali², facendoli confluire in un portale, punto di accesso unificato ai beni culturali della Lombardia. La biblioteca digitale, che endemicamente si compone di materiale disomogeneo e sfugge le logiche dei limiti e vincoli fisici, si pone idealmente come primo esperimento di integrazione e sforzo comune.

La convenzione iniziale tra la Regione Lombardia e l’Università degli studi di Pavia prevedeva esclusivamente la creazione di documenti (*Tutorial e Linee guida per i progetti di digitalizzazione*) e indicazioni sugli standard al fine di normare e uniformare i progetti di digitalizzazione cofinanziati dalla Regione. I documenti prodotti e messi a disposizione sul web forniscono la formazione di base, necessaria al personale delle istituzioni culturali che, per la prima volta, si trova ad affrontare le tematiche della digitalizzazione.

Il progetto è stato successivamente ampliato, abbracciando l’idea di sviluppare un software *open source* che guidi il catalogatore nella descrizione e nella gestione degli oggetti digitali e si ponga, dunque, come dispositivo atto a mettere in pratica le scelte metodologiche, le buone pratiche e gli standard illustrati nel *Tutorial* e nelle *Linee guida*.

Per la stesura di questi ultimi documenti si è partiti da una valutazione di strumenti nazionali e internazionali³ analoghi, al fine di identificare gli argomenti ritenuti adeguati e necessari all’alfabetizzazione del bibliotecario digitale.

¹ Paul Gabriele Weston ha coordinato la ricerca che ha condotto all’elaborazione delle diverse componenti del progetto.

² Nello specifico: “Lombardia storica”, “Lombardia Beni culturali”, “Biblioteche” e “Archivio di Etnografia e Storia Sociale”(AESS).

³ Tra i quali: Athena, *Deliverables and documents*, <http://www.athenaeurope.org/index.php?en/>

Il *Tutorial* è un documento sintetico che, pur partendo da un approccio generalista (sin dall'*incipit* incentrato su cosa si intenda per conversione al digitale e sui metadati in generale), accenna ad alcune caratteristiche della Biblioteca Digitale della Lombardia, come le scelte effettuate in materia di metadati descrittivi. Il documento si focalizza, inoltre, sulla scelta dei materiali da digitalizzare, fornendo idee e *checklist* di supporto, che possano guidare il bibliotecario nella selezione.

Le *Linee Guida* sono invece un documento più ampio che tocca aspetti tecnici quali, ad esempio, le caratteristiche dell'oggetto digitale, la creazione e il trattamento delle immagini derivate, i formati per i file master e derivati, le strategie e i costi della conservazione digitale. In questo modo, all'operatore vengono prospettate tutte le informazioni necessarie per avviare progetti di digitalizzazione secondo il principio SOAP (Scan Once for All Purposes), per documentare le scelte effettuate, per utilizzare schemi di metadati e formati standard e per creare capitola-ti tecnici adeguati a vincolare i fornitori al rispetto di questi principi.

I metadati della biblioteca digitale lombarda: la scelta di METS

La scelta di utilizzare il formato METS⁴ (Metadata Encoding and Transmission Standard) per accogliere i diversi metadati che orbitano attorno agli oggetti digitali può apparire controversa in Italia che sin dai primi passi mossi nel mondo delle biblioteche digitali si è dotato di uno standard nazionale, quale i MAG⁵ (Metadati Amministrativi Gestionali).

La scelta di adottare METS, preventivamente discussa all'interno del gruppo di lavoro in un incontro con Giovanni Bergamin della Biblioteca Nazionale Centrale di Firenze⁶, è dettata da alcune esigenze (nel campo della manutenzione, flessibilità, interoperabilità, gestione di formati differenti) non pienamente soddisfatte dalle scelte architettoniche di MAG⁷.

149/athena-deliverables-and-documents; Australian Government - Department of Finance and Deregulation, *Better Practice Checklist - 18. Digitisation of Records*, <http://www.finance.gov.au/e-government/better-practice-and-collaboration/better-practice-checklists/digitisation.html>. Minerva Europe, *Good practices in digitisation*, <http://www.minervaeurope.org/bestpractices/listgoodpract.htm>.

⁴ <http://www.loc.gov/standards/mets/>.

⁵ Si veda: http://www.iccu.sbn.it/opencms/opencms/it/main/attivita/gruppilav_commissioni/pagina_99.html.

⁶ A tal proposito va segnalato come anche Magazzini Digitali ha scelto di non adottare MAG ma uno standard maggiormente flessibile come MPEG-21 (ISO/IEC 21000), si veda <http://www.rinascimento-digitale.itmagazzinidigitali.phtml>.

⁷ Del resto l'evoluzione dell'uso dei MAG sembra volta a superare questi limiti integrando i due formati, ossia con un utilizzo dei MAG all'interno di METS, come profilo applicativo specifico. A tal proposito, si vedano i lavori di *mapping* MAG-METS http://www.iccu.sbn.it/opencms/opencms/it/main/standard/pagina_372.html. Per un esempio concreto si consultino i materiali del progetto SIMBAD (Sistema Informativo di Metadati Bibliografici ad Architettura Digitale) della Biblioteca digitale provinciale "P. Albino" di Campobasso. <http://www.provincia.campobasso.it/biblioteca/digitale/progetti/simbad/index.html>.

Dal punto di vista della Regione Lombardia era essenziale dotarsi di una soluzione sostenibile nel tempo e che dunque non richiedesse costanti aggiornamenti e allineamenti a carico della Regione stessa: a tal proposito, va sottolineato come abbracciare uno standard internazionale, gestito ed aggiornato dalla Library of Congress, sollevi dalle difficoltà e, parzialmente, dai costi di tali attività. Inoltre permette di avvalersi di applicativi (proprietary o liberi) prodotti in tutto il mondo⁸ senza dover richiedere sviluppi onerosi per un singolo specifico formato.

Per quanto riguarda l'esigenza di flessibilità, bisogna immaginarsi MAG come un armadio intagliato in un unico blocco - seppur solido - rigido e sostanzialmente immutabile: la parte descrittiva avverrà necessariamente in Dublin Core⁹ e i metadati tecnici saranno espressi attraverso metadati proprietari¹⁰ (nel caso delle immagini parzialmente derivati da una versione dello standard ANSI/NISO Z39.87-2006 precedente a quella poi pubblicata).

Al contrario, METS può essere paragonato a un armadio componibile, articolabile a piacimento che preveda cassette di tipologie diverse a seconda delle necessità: la descrizione potrà dunque, ad esempio, essere redatta con l'impiego di Dublin Core, utilizzando i Semantic Elements di Europeana¹¹, con MODS¹² (Metadata Object Description Schema), con EAD¹³ (Encoded Archival Description), in RDF¹⁴ (Resource Description Framework), ma anche in modalità diverse in parallelo.

I vantaggi in termini di flessibilità di un simile approccio sono evidenti, soprattutto in un contesto come quello della Regione Lombardia, nel quale è necessaria la coesistenza di schemi descrittivi differenti (ad esempio il citato EAD per la descrizione dei materiali archivistici, componente centrale all'interno dei beni culturali regionali e, in futuro, estensione ideale del primo nucleo della Biblioteca Digitale Lombarda).

L'adozione di METS permette, quindi, di superare le difficoltà incontrate dalla comunità archivistica che, nel caso dei metadati MAG, si è trovata costretta ad esprimere descrizioni estremamente granulari comprimendole in un record Dublin Core. Al contrario, METS consente di utilizzare un formato nativo del mondo degli archi-

⁸ Si pensi, solo a titolo di esempio, a software per l'estrazione e la manipolazione di metadati tecnici come JHOVE (JSTOR Harvard Object Validation Environment)
<http://hul.harvard.edu/jhove/index.html>.

⁹ <http://dublincore.org/documents/dces/>.

¹⁰ I limiti di un simile approccio sono evidenti laddove si osservi che lo schema MAG attuale non consente l'inserimento di immagini in formato image/webp (formato sempre più utilizzato nella creazione di derivate) o non consente l'inserimento di risoluzioni video superiori a 720x576 (ben inferiori alla ormai diffusa 1080p, ossia 1920x1080).

¹¹ http://www.version1.europeana.eu/c/document_library/get_file?uuid=a830cb84-9e71-41d6-9ca3-cc36415d16f8&groupId=10602.

¹² <http://www.loc.gov/standards/mods/>.

¹³ <http://www.loc.gov/ead/>.

¹⁴ <http://www.w3.org/RDF/>.

vi come EAD, salvo poi convertirlo nei Semantic Elements di Europeana per necessità rinvenenti dal riversamento e dallo scambio dei dati¹⁵.

La flessibilità comporta, tuttavia, la definizione per ciascun progetto di alcuni paletti, ossia la messa a punto di un profilo applicativo METS che specifichi quali metadati (ossia che tipo di cassetto, per riprendere la metafora precedente) prevedere¹⁶. Nel caso della Biblioteca Digitale della Lombardia si è scelto che il progetto e la collezione vengano descritti dai metadati previsti all'interno del progetto MICHAEL¹⁷, mentre per gli oggetti digitali si fa riferimento ai Semantic Elements di Europeana. Per quanto riguarda i metadati tecnici, la Biblioteca Digitale della Lombardia utilizzerà per le immagini statiche MIX (Metadata for Images in Xml standard). La flessibilità di METS permetterà in futuro di appoggiarsi a metadati specifici come i Video Technical Metadata Extension Schema¹⁸ e gli Audio Technical Metadata Extension Schema¹⁹, anche essi sviluppati o mantenuti dalla Library of Congress, qualora ve ne fosse la necessità²⁰.

Il software Dolly

Proprio la scelta di avvalersi di METS, assai inusuale nel panorama italiano, ha spinto la Regione Lombardia ad affidare all'Università di Pavia l'incarico di progettare e sviluppare un software per la gestione delle immagini che, al contempo, guidasse e agevolasse i bibliotecari nella creazione dei metadati descrittivi, strutturali e tecnici conformi agli standard prescelti.

Tramite il software Dolly, la cui realizzazione tecnica è stata affidata alla cooperativa Codex²¹, viene messo a disposizione delle biblioteche uno strumento di semplice utilizzo, modulare ed estensibile, anche grazie alla scelta forte di optare per una licenza libera²².

La spina dorsale del software è il modulo descrittivo che, come si è accennato in precedenza, si snoda su più livelli, dedicati a progetti, collezioni, oggetti originari

¹⁵ A tal proposito si vedano il lavoro di armonizzazione e gli strumenti messi a disposizione dal progetto APENet

http://www.apenet.eu/index.php?option=com_content&view=article&id=94&Itemid=150.

¹⁶ A riguardo dei profili METS si veda a titolo di esempio Judith Pearce, *The Australian METS profile: a journey about metadata*, «D-Lib Magazine», 14 (2008), n. 3/4, <http://www.dlib.org/dlib/march08/pearce/03pearce.html>.

¹⁷ <http://www.michael-culture.org/>. In particolare si veda *Modello dati Michael*, http://www.michael-culture.eu/documents/MICHAELDataModel_ITA.pdf.

¹⁸ <http://memory.loc.gov/mets/Schemas/VMD.xsd>.

¹⁹ <http://memory.loc.gov/mets/Schemas/AMD.xsd>.

²⁰ La scelta è peraltro coerente con le indicazioni del Gruppo di lavoro metadati per l'archivio digitale del Sistema Archivistico Nazionale, che propone un profilo di METS che utilizzi per l'appunto per i metadati tecnici MIX, nel caso delle immagini, e i succitati schemi promossi dalla Library of Congress per audio e video.

²¹ <http://www.codexcoop.it/>.

²² Il software è disponibile all'indirizzo: <https://github.com/codexcoop/dolly> rilasciato sotto i termini della GPLv3, si veda: <http://www.gnu.org/licenses/gpl-3.0.txt>.

(anch'essi a loro volta a più livelli) e infine ad oggetti digitali (costituiti dall'insieme delle singole immagini).

La scheda progetto e la scheda collezione non sono altro che la trasposizione delle caratteristiche previste dal *data model* MICHAEL in un sistema di *data entry* (Figura 1), che tuttavia risulta assolutamente personalizzabile per tutto ciò che concerne i vocabolari controllati utilizzati (ad esempio i codici linguistici, gli ambiti tematici della collezione etc.).

Figura 1: Dolly – maschera di inserimento di un progetto

Le scelte e le implementazioni più innovative riguardano la descrizione dell'oggetto originario (che nelle prime sperimentazioni in corso coincide con l'oggetto analogico digitalizzato): è infatti possibile l'importazione via z39.50 o direttamente tramite un record Unimarc, ma, soprattutto, è possibile avviare importazioni a catena (Figura 2), caratteristica di estrema utilità nel caso di monografie a livelli. Questa peculiarità, utile in tutti i contesti di descrizioni multilivello, rende il software utilizzabile anche per descrivere fondi archivistici e materiale non prettamente librario.



Figura 3: Dolly – TOC Editor: l’associazione fra il sommario e l’immagine avviene tramite trascinamento della voce di indice sull’immagine o viceversa.

L’aspetto che ha richiesto uno sforzo maggiore è senz’altro la *TOC editor*, un modulo, che ricorda per semplicità e immediatezza la navigazione delle cartelle di un comune computer (Figura 3), con cui il catalogatore può dettagliare la struttura dell’oggetto digitale, evidenziandone le parti, i capitoli, i singoli articoli etc. La struttura così disegnata verrà poi utilizzata dal *viewer*²⁴ per creare l’indice e i segnalibri, consentendo la navigazione tra le singole immagini (si veda oltre la Figura 4).

Accanto ai moduli sin qui illustrati, Dolly prevede una serie di componenti accessorie, come la funzione di esportazione in METS e, in futuro, in MAG e il modulo amministrativo per la gestione degli utenti e degli accessi.

I risultati delle prime sperimentazioni

Il software, le scelte di descrizione e le *Linee Guida* sono stati infine sottoposti a test nell’ambito di quattro progetti pilota di digitalizzazione, cofinanziati dalla Regione Lombardia, che hanno visto coinvolte la Biblioteca civica “A. Mai” di Bergamo, la Biblioteca civica di Monza, la Biblioteca civica Ricottiana di Voghera e la Rete Bibliotecaria Bresciana.

L’intento è stato in primo luogo quello di stressare il sistema per verificarne la resilienza e la flessibilità nell’utilizzo; a tale scopo la tipologia di materiali digitalizzati

²⁴ Dolly utilizza come *viewer* una versione appositamente personalizzata dell’Internet Archive BookReader <http://openlibrary.org/dev/docs/bookreader>.



Figura 4: Visualizzazione di un periodico e della sua struttura

è stata selezionata in modo da includere nel test documenti di natura, struttura e ambito disciplinare diversi.

I progetti pilota vanno a costituire il primo nucleo della Biblioteca Digitale della Lombardia che, dopo una prima fase improntata ad una crescita secondo un moto browniano, dovrà procedere nella selezione dei documenti, secondo una carta delle collezioni digitali, che ne faccia emergere il profilo.

Questa prima fase ha invece dato luogo a circa 100.000 nuove immagini (a cui aggiungere le quasi 3000 immagini del «Giornale di Voghera» e dell'«Idea popolare», provenienti dal recupero del pregresso digitalizzato dalla biblioteca Ricottiana di Voghera) che spaziano dai libretti d'opera delle rappresentazioni teatrali tenutesi a Bergamo, a periodici come «Il Cittadino di Monza» o «L'eco della Provincia Iriense», fino a pubblicazioni di storia locale.

La varietà dei materiali ha permesso di verificare il software nelle sue diverse componenti: i libretti d'opera, ad esempio, non presentano grosse problematiche dal punto di vista della creazione delle derivate, considerando le dimensioni contenute dei file, ma sono stati fondamentali nel potenziare l'importazione Unimarc. Al contrario, i periodici in grande formato sono stati una sfida sia per ottenere derivate che fossero contemporaneamente leggibili e fruibili via web (e dunque con un peso circoscritto), sia nella costruzione del sommario per la navigazione tra le annate.

L'eterogeneità dei materiali ha portato anche a sviluppare un protocollo catalografico che sia, in prima battuta, uno strumento per uniformare le descrizioni dei diversi catalogatori, ma che in futuro possa affiancare il *Tutorial* e le *Linee Guida* per i futuri operatori della Biblioteca Digitale. Da un lato, si è trattato di definire l'oggetto digitalizzato e il livello al quale agganciare il suo surrogato digitale (si pensi,

nel caso dei periodici, alla scelta tra l'annata e il numero), dall'altro si è reso necessario stabilire indicazioni di massima per circoscrivere l'uso di campi semanticamente laschi, sulla granularità dei sommari e sulla scelta di vocabolari controllati.

Conclusioni e prospettive future

Il software Dolly e le prime sperimentazioni da parte delle biblioteche pilota non sono che il tassello iniziale nella composizione del quadro della Biblioteca Digitale della Lombardia: per una crescita futura del progetto la Regione dovrà dotarsi di meccanismi per l'identificazione univoca degli oggetti digitali e di un piano per la loro conservazione a lungo termine.

Per la Regione Lombardia risulta infatti strategico, in considerazione dei numerosi e variegati progetti che ruotano intorno alla galassia dei beni culturali, dotarsi di identificativi univoci: se, dal punto di vista della Biblioteca Digitale, gli identificativi NBN²⁵ (National Bibliography Numbers) possono sembrare la scelta naturale, va sottolineato come, visto l'attuale protrarsi della fase di sperimentazione in Italia e il fiorire all'estero di soluzioni alternative, esistano strade diverse da poter percorrere²⁶.

Contemporaneamente, Dolly dovrà progredire in modo tale da gestire materiale non omogeneo con quanto finora testato in fase di sperimentazione.

In particolare, nel futuro prossimo bisognerà verificare la possibilità di usare gli strumenti messi a disposizione dal progetto Biblioteca Digitale della Lombardia in ambiti diversi da quello bibliotecario (ad esempio per materiali archivistici o museali) e il software dovrà evolvere in maniera tale da gestire (nella creazione delle derivate, nell'estrazione dei metadati tecnici e per ciò che concerne l'indicizzazione) surrogati digitali che non siano esclusivamente immagini statiche.

Infine, dovrà essere implementato il previsto modulo *workflow*, una sorta di collettore delle diverse funzioni già messe a disposizione, un filo d'Arianna che guidi il bibliotecario nei meandri del labirinto della digitalizzazione. In questo modo sarà possibile stampare un ordine di digitalizzazione da allegare ai bandi, potendo al tempo stesso automatizzare parte dei controlli di qualità e verificare in tempo reale lo stato di avanzamento di un progetto

²⁵ Juha Hakala, *Using national bibliography numbers as uniform resource names*, RFC 3188 (Informational), ottobre 2001, <http://www.ietf.org/rfc/rfc3188.txt>.

²⁶ Sui limiti di sistemi di identificazione non completamente aperti si veda Ed Summers, *Geeks bearing gifts*, marzo 2011, <http://inkdroid.org/journal/2011/03/22/geeks-bearing-gifts/>, per una panoramica generale dei diversi sistemi di identificazione univoca si rimanda a Hans-Werner Hilse - Jochen Kothe, *Implementing persistent identifiers*, Consortium of European Research Libraries: London - European Commission on Preservation and Access: Amsterdam, 2006, <http://nbn-resolving.de/urn:nbn:de:gbv:7-isbn-90-6984-508-3-8>. Sullo stesso tema risultano di estrema importanza e attualità le linee guide e raccomandazione dell'ATHENA (Access to cultural heritage networks across Europe) project <http://www.athenaeurope.org>: raccolte in ATHENA WP3 Working Group "Identifying standards and developing recommendations", *Persistent identifiers (pids): recommendations for institutions*, ATHENA project, 2011, <http://www.athenaeurope.org/getFile.php?id=779>.