

# Interoperabilità

*Verso un modello aperto*

La situazione più frequente in rete è quella di una molteplicità di collezioni e risorse digitali, a cui si accede con diverse interfacce di ricerca. Il singolo utente deve quindi entrare e uscire da tante “biblioteche digitali” (in realtà molte collezioni digitali vengono chiamate “biblioteche” pur non avendone le caratteristiche), per riuscire a identificare quella che incontra i suoi bisogni di informazione. Anche nel caso d’uso del motore di ricerca, questo non riuscirà a identificare i contenuti interni a una collezione digitale. Non solo i produttori commerciali di informazione contribuiscono a creare questa situazione di difficoltà di accesso alle risorse, per ovvi motivi di profitto, ma anche le biblioteche digitali prodotte da istituzioni pubbliche, come biblioteche, archivi e musei. Questo accade perché la maggioranza delle collezioni digitali non è interoperabile.

Quando infatti le biblioteche muovono dalla gestione dell’OPAC, quello per cui si è soliti applicare l’automazione, verso la gestione dei depositi di contenuti digitali, si accorgono che ottenere l’interoperabilità non è banale.<sup>1</sup> L’interoperabilità richiede infatti un notevole numero di condizioni: dagli standard sui metadata (descrittivi, strutturali, amministrativi e per la preservazione), a modi per rendere questi metadata visibili ai motori di ricerca (che devono raccoglierci o fare harvesting), fino ad architetture co-

muni che possono rendere interoperabili i contenuti (ad esempio con gli open archive). L’interoperabilità, inoltre, non è un problema esclusivamente tecnico, legato solo all’applicazione di standard e protocolli: questi sono necessari ma non sufficienti. Per offrire il servizio che gli utenti si aspettano dalla biblioteca digitale è necessaria la cooperazione, intesa come una riorganizzazione delle attuali relazioni presenti sia all’interno di una singola istituzione, sia tra istituzioni culturali, in particolare archivi, biblioteche e musei interessati al fenomeno della convergenza. L’interoperabilità può quindi dirsi *tecnica*, ovvero la capacità per sistemi diversi di processare dati estratti da depositi distribuiti e di comunicare in tempo reale tramite interfacce condivise; *semantica*, che si riferisce a standard legati al contenuto dei dati, ad esempio agli schemi di metadata applicati; e infine *istituzionale*, dovuta alla collaborazione tra utenti e produttori di contenuti e legata al flusso del lavoro, alle partnership e alla necessità di oltrepassare le barriere istituzionali (nazionali, locali, internazionali), per una più efficiente soddisfazione delle necessità degli utenti.

In questo articolo affronteremo il tema dell’interoperabilità da due approcci: quello tecnologico e quello del cambiamento organizzativo, che sono da considerarsi due facce della stessa medaglia. Le domande a cui

vorremmo rispondere sono: quale modello di biblioteca digitale viene perseguito nelle attuali biblioteche digitali italiane? E quanto queste sono interoperabili?

## 1. I requisiti

Esistono diverse definizioni di interoperabilità, alcune più centrate sugli aspetti tecnici, altre più orientate verso l’accesso dell’utenza. Una delle definizioni migliori è quella di un gruppo di ingegneri presso la Cornell University:

Interoperability is defined as the ability of digital library components or services to be functionally and logically interchangeable by virtue of their having been implemented in accordance with a set of well-defined, publicly known interfaces.<sup>2</sup>

Il focus della definizione è sulle interfacce di ricerca. Queste sono molto importanti per gli utenti della rete, abituati a usare i browser web o i motori di ricerca per accedere alle risorse. L’architettura che si delinea è precisa: da una parte ci sono i contenuti, dall’altra i servizi. Si dà, cioè, per scontato che esistano dei depositi distribuiti in rete di contenuti digitali e di metadata, di cui un sottoinsieme viene reso disponibile ai motori di ricerca per essere accessibile agli utenti. Per evitare fraintendimenti, è bene indicare che i contenuti a cui gli autori di questa definizione di interoperabilità fanno riferimento non sono le pagine web, con cui spesso si confondono le biblioteche digitali. Gli autori si riferiscono ai primi archivi di e-print, che a cominciare

dalla comunità dei fisici erano già attivi nel 1999, e sono alle prese col problema di rendere interoperabili le interfacce di accesso ai contenuti di depositi aperti.

Il concetto di deposito, in particolare di deposito aperto (o archivio aperto), si è poi esteso alle biblioteche digitali create da istituzioni culturali, che per essere interoperabili devono quindi rispettare determinati requisiti standard, in particolare:

- l’identificazione della risorsa digitale e la sua nomenclatura;
- la creazione o la cattura di metadata basati su schemi riconosciuti;
- la memorizzazione e la gestione di archivi aperti;
- i protocolli per l’accesso aperto.

In questo articolo ci limiteremo al back office, cioè a considerare le problematiche di gestione organizzativa di una biblioteca digitale, lasciando la considerazione dei protocolli dell’accesso a un’altra puntata della rubrica.

### 1.1. Identificazione della risorsa digitale

Oltre ad essere descritte, le numerose risorse digitali hanno bisogno di essere identificate con un nome unico e persistente. Un particolare metadata è quindi l’identificatore unico di una risorsa digitale, e questo non è un concetto nuovo e non è specifico di Internet. Come per i libri è stata usata la stringa di numeri ISBN per denotare una particolare pubblicazione di un’opera, allo stesso modo le risorse digitali hanno bisogno di essere identificate univocamente. L’Internet Engineering Task Force (IETF) e il World Wide Web Consortium (W3C) usano il termine URI (Uniform Resource Identifier)

per definire un identificatore, che include sia l'URL (Uniform Resource Locator), cioè l'indirizzo della risorsa, sia l'URN (Uniform Resource Name), cioè il nome che la distingue da altre risorse. URI è ritenuto capace di identificare il contenuto intellettuale di un'opera, a prescindere dalle sue manifestazioni. Alcuni identificatori che sono stati applicati nel settore delle biblioteche digitali sono il Persistent Uniform Resource Locator (PURL) e il Digital Object Identifier (DOI). PURL è un servizio di intermediazione che consente, al cambio di indirizzo della risorsa, di mantenere inalterato l'identificatore. Il DOI, anch'esso unico e persistente, è un identificatore finalizzato non solo al recupero della risorsa ma anche alla gestione del pagamento dei diritti di accesso.

### 1.2. Metadata

La qualità delle risorse digitali contenute nei depositi e la loro accessibilità (o se vogliamo interoperabilità) sono determinate in modo preminente dalla qualità dei metadata che identificano le risorse stesse.

L'importanza dei metadata per l'interoperabilità è cresciuta nel tempo. Essi sono usati in tre modi diversi:

- come meta-tag, o etichette nelle intestazioni dei documenti HTML, incapsulati nei documenti stessi, come ad esempio il titolo sulla copertina;
- come un archivio separato di meta-informazione, ad esempio la scheda catalografica del libro o una lista di risorse digitali;
- come sistema informativo distribuito basato su un motore di ricerca (che può fare harvesting dei dati oppure

costruirsi degli indici) e collegato allo scarico del documento, se consentito dalle licenze di accesso.

I metadata non hanno quindi solo una funzione descrittiva, ma sono invece il supporto di una serie di operazioni automatizzate, e per essere efficaci devono consentire lo scambio tra sistemi diversi. È questa la principale esigenza di interoperabilità che rende importante applicare degli schemi standard. L'uso di schemi diversi di metadata è evidentemente un ostacolo all'interoperabilità delle risorse digitali, ma il problema è ineliminabile, in quanto comunità diverse hanno esigenze diverse. Per affrontare questo problema esistono due filosofie o scuole di pensiero opposte. Da una parte si trovano i sostenitori di un sistema unico che, con l'introduzione di diverse tipologie di oggetti o con l'applicazione estesa a diverse comunità di utenti, amplia il numero di informazioni fino a diventare però estremamente complesso e costoso da gestire. Questo, per esempio, è l'approccio di chi usa MARC. L'altro approccio, invece, tenta di impacchettare i diversi schemi di metadata in un'unica cornice di riferimento, il Resource Description Framework (RDF),<sup>3</sup> con cui tutti gli schemi possano essere comparabili. E questo è, per esempio, il sistema usato dal Dublin Core.<sup>4</sup>

Mentre i primi realizzatori di biblioteche digitali hanno cercato di estendere e adattare standard tradizionali, come il MARC, il set BIB-1 dello Z39.50 o SGML, gli standard emergenti nelle biblioteche sono Dublin Core e Encoded Archival Description (EAD). Il Dublin Core in particolare si pone come lo

standard di metadata generale, che non sostituisce altri metadata ma li affianca. Oltre alle biblioteche, altre istituzioni culturali stanno adoperandosi per una standardizzazione dei metadata. In ambito museale sono stati creati gli standard SPECTRUM (UK) e CDWA,<sup>5</sup> per supportare la gestione degli oggetti museali; il settore degli archivi ha sviluppato gli standard ISAD(G),<sup>6</sup> ISAAR (CPF)<sup>7</sup> e EAD<sup>8</sup> per far fronte alla gestione e al reperimento di documentazione d'archivio.

I sistemi di metadata sono quindi numerosi, e la loro interoperabilità si realizza attraverso i *crosswalks*, o tabelle di comparazione. Una ricerca dell'OCLC riguarda la realizzazione di un prototipo che traduce automaticamente da tre diversi schemi di metadata, MARC, TEI e URC, consentendo lo scambio dei dati. Il sistema, chiamato Spectrum,<sup>9</sup> consente allo stesso autore della risorsa di creare record in HTML, per costruire una banca dati. Un'interessante caratteristica del sistema Spectrum è la creazione di un sottosistema di record che rielabora l'informazione descrittiva immessa nella stringa di ricerca dell'utente, per generare l'URC, il TEI o il record MARC.

In Italia, nell'ambito del progetto ARSBN<sup>10</sup> della Biblioteca nazionale centrale di Firenze e successivamente della Biblioteca digitale italiana, è stato proposto lo schema MAG (Metadati amministrativi gestionali) promosso dall'ICCU.<sup>11</sup> Lo schema si basa sul modello OAIS,<sup>12</sup> focalizzando la funzione che tale modello definisce come SIP (Submission Information Protocol), cioè l'insieme di metadata che il fornitore dell'informazione

dà insieme al pacchetto contenente l'oggetto digitale, per la sua archiviazione e gestione. Una caratteristica dello schema MAG è quella di essere pensato per oggetti digitalizzati di cui le biblioteche hanno in catalogo l'originale analogico, così che è possibile estrarre dal catalogo SBN la descrizione bibliografica, aggiungendo ad essa i metadata amministrativi, legati soprattutto alle caratteristiche tecniche e di proprietà intellettuale dell'oggetto digitalizzato.

### 1.3. Formati di marcatura

Gli standard per i metadata e l'identificatore univoco non sono sufficienti per garantire l'accesso alla risorsa digitale e usare il suo contenuto. Ad esempio la descrizione di un articolo in un periodico elettronico non dice nulla di dove comincia e dove finisce l'abstract, dove sono le tavole e così via. Per un miglior accesso alla risorsa digitale è infatti necessario sapere in quali paragrafi o elementi costitutivi è stato strutturato il contenuto, con una granularità maggiore o minore a seconda delle scelte progettuali. Questo è il compito dei formati di marcatura delle risorse digitali che forniscono dei modelli di struttura delle risorse, o Document Type Definition (DTD), per definire il modo di presentazione della risorsa all'utente e anche le modalità di accesso.

Il formato di marcatura più diffuso è XML (EXtensible Markup Language), gestito dal consorzio W3. Tuttavia sempre più frequentemente le risorse digitali sono immagini digitali, equivalenti per forma e colori al documento originale. Questa distinzione è necessaria perché sta prendendo sempre

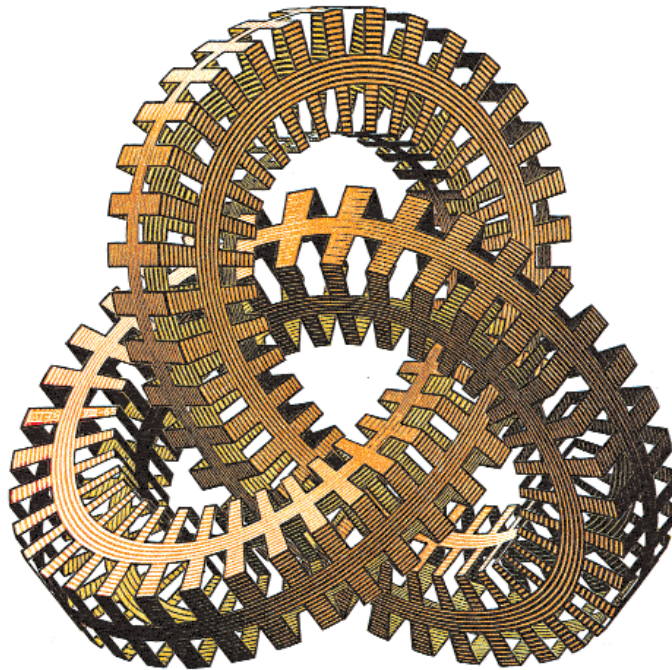
più campo la considerazione che la qualifica “digitale” si applichi alle sole “immagini” numeriche di un oggetto reale.

Tutti questi requisiti tecnici sono quindi necessari per garantire l'interoperabilità dei depositi di contenuti digitali. In altre parole, lo scopo di aderire a standard come il Dublin Core è quello di realizzare l'aggregazione di diverse risorse digitali con un'unica interfaccia, oppure il ri-uso delle risorse, nei limiti delle licenze di accesso, per le finalità specifiche di alcuni utenti. Gli standard, cioè, non sono mai un fine in sé, ma solo un mezzo per determinati obiettivi di accesso.

## 2. Il modello di biblioteca digitale in Italia

Gli standard di cui la biblioteca digitale ha bisogno per essere interoperabile possono tuttavia sembrare un'eccessiva rigidità imposta alle esigenze locali di una biblioteca digitale, oltre che un costo aggiuntivo. Non è un caso, ad esempio, che i costosi formati di marcatura granulari, come XML, vengano poco utilizzati dalle biblioteche, a favore del più diffuso formato immagine. Il problema dell'interoperabilità è infatti legato all'identificazione della comunità di riferimento e alle funzionalità che vengono offerte. Quest'ultima è circoscritta alla comunità locale o estesa all'uso di motori di ricerca e utenti remoti? È previsto il ri-uso della risorsa digitale da parte di altre applicazioni?

È anche difficile comprendere la prospettiva della biblioteca digitale inserita in un'architettura di rete distribuita, come proposto nella definizione di interoperabi-



lità posta all'inizio. Qual è il ruolo della biblioteca? Quello di gestire collezioni di risorse digitali, intese come contenuti e metadata, che sono localizzate in depositi della biblioteca? Oppure quello di realizzare l'aggregazione di collezioni di documenti digitali distribuiti, dati sensibili, informazione mobile, servizi di elaborazione pervasivi ecc.?

Il modello di biblioteca digitale aperta e interoperabile offre una notevole estensione del servizio tradizionale. Questo può essere indicato come un aumento del numero di utenti, maggiori contenuti resi accessibili, maggiori funzionalità che vengono offerte. L'evoluzione delle biblioteche in biblioteche digitali va verso la costruzione di un vero e proprio sistema informativo che integra diverse banche dati e oggetti digitali, testuali e non testuali, posseduti o aggregati in rete, e quindi rende possibile utilizzi diversi e innovativi delle risorse digitali per utenti, conosciuti o remoti, e anche per fornitori di servizi. Questa è

infatti la promessa dell'applicazione di standard per l'interoperabilità. Tuttavia non sarà possibile realizzare questo servizio esteso senza una cooperazione allargata, non solo ad altre biblioteche digitali, ma anche ad altre istituzioni culturali pubbliche e private.

La cooperazione in Italia ha una lunga tradizione e risultati di successo, come il progetto SBN. Questo tuttavia non ha evitato che la cooperazione sia stata spesso confusa e ostacolata da problematiche di centralizzazione *vs* decentramento, in cui c'è stata la difesa dell'autonomia da parte delle istituzioni locali che spesso si sono sentite minacciate dall'appesantimento amministrativo della centralizzazione. In ambito digitale, la cooperazione da un lato è necessaria, dall'altro deve adattarsi e usufruire delle tecnologie ora disponibili, e in particolare applicare i requisiti per l'interoperabilità, prima brevemente delineati. Una soluzione equilibrata sta nello stabilire quali attività possono essere oppor-

tunamente centralizzate in rete, con vantaggio dell'utente, oltre che con risparmi economici nell'architettura distribuita, coordinando a questo fine le biblioteche digitali in Italia. Il *Manifesto per le biblioteche digitali*,<sup>13</sup> ad esempio, indica alcuni obiettivi che necessitano di un approccio cooperativo:

- la preservazione;
- alcuni strumenti di servizio, come il Portale nazionale delle biblioteche digitali;
- l'adozione di software open source e standard aperti;
- la condivisione di linee guida e standard comuni.

Una recente indagine realizzata dalla Fondazione Rinascimento Digitale<sup>14</sup> ha ad esempio evidenziato alcuni importanti progetti che si muovono in tale direzione. Tra questi possiamo evidenziare, nell'ambito del progetto della Biblioteca digitale italiana, Internet culturale, il portale che si propone come punto di accesso per i vari settori dei beni culturali, compresi musica e cinema, e la pubblicazione di numerose linee guida e standard da parte del Progetto Minerva e dell'ICCU. Anche le biblioteche universitarie sono impegnate nella costruzione cooperativa della base di conoscenza di MetaLib, per realizzare una ricerca unica (o federata) di diverse banche dati e periodici elettronici.

## Conclusioni

Interoperabilità è quindi una visione o una realtà? Ciò che si può sperimentare oggi è la presenza di numerose collezioni digitali che non sono collegate tra di loro. L'utente ha ad esempio a disposizione le seguenti risorse:

- OPAC delle biblioteche;

- banche dati in linea;
- banche dati su cd-rom in rete;
- periodici elettronici;
- e-book;
- altre biblioteche digitali;
- risorse selezionate del web.

Possiamo dire, in conclusione, che occorrerà concentrare maggiori sforzi per l'interoperabilità delle biblioteche digitali e per migliorare l'accesso dell'utente. Questo comporta un'applicazione mirata sia di tecnologie standard sia di modelli organizzativi innovativi, verso un modello veramente aperto di biblioteca digitale.

### Note

<sup>1</sup> La problematica dell'interoperabilità è stata al centro di un seminario, svoltosi a Roma il 3 aprile 2007 su iniziativa del Progetto Minerva. I requisiti tecnici necessari per l'interoperabilità sono stati recentemente raccolti

nella versione italiana delle linee guida tecniche, a cura del Progetto europeo Minerva, a cui si rimanda per approfondire il tema.

<sup>2</sup> SANDRA PAYETTE – CHRISTOPHE BLANCHI – CARL LAGOZE – EDWARD A. OVERLY, *Interoperability for digital objects and repositories*, "D-LIB Magazine", 5 (1999), 5, <<http://dlib.org/dlib/may99/payette/05payette.html>>.

<sup>3</sup> RDF è un'infrastruttura che permette di codificare, scambiare e riutilizzare metadata strutturati. Una particolare applicazione dello standard RDF riguarda le ontologie e l'accesso semantico alle risorse digitali. Approfondimenti su RDF sono accessibili all'indirizzo: <<http://www.w3.org/RDF>>.

<sup>4</sup> Nel 1995 OCLC, insieme al National Center for Supercomputing Applications (NCSA), ha promosso un seminario sui metadata a Dublin, Ohio, riunendo informatici, editori ed esperti provenienti da varie specializzazioni disciplinari, con lo scopo di definire uno standard semplice e usabile per descrivere le risorse informative in rete. Questo seminario ha portato alla nascita di

uno dei più noti e diffusi metadata, chiamato, dal luogo dell'incontro, Dublin Core. Vedi: STUART WEIBEL – JEAN GODBY – ERIC MILLER – RON DANIEL, *OCLC/NCSA metadata workshop report*, Office of research, OCLC (Online Computer Library Center), 1995.

<sup>5</sup> GETTY RESEARCH INSTITUTE, *Categories for the Description of Works of Art (CDWA)*, <[http://www.getty.edu/research/conducting\\_research/standards/cdwa/](http://www.getty.edu/research/conducting_research/standards/cdwa/)>.

<sup>6</sup> International Standard for Archival Description (General) (ISAD(G)), 2. ed., <[http://www.ica.org/biblio/isad\\_g\\_2e.pdf](http://www.ica.org/biblio/isad_g_2e.pdf)>.

<sup>7</sup> International Standard Archival Authority Record for Corporate Bodies, Persons and Families (ISAAR(CPF)), <[http://www.ica.org/biblio/isaar\\_eng.pdf](http://www.ica.org/biblio/isaar_eng.pdf)>.

<sup>8</sup> Encoded Archival Description (EAD), <<http://www.loc.gov/ead/>>.

<sup>9</sup> <<http://www.oclc.org:5061/spectrum.html>>.

<sup>10</sup> ARSBNi è un progetto della Biblioteca nazionale di Firenze, che collega gli indici e le copertine dei volumi al record bibliografico della BNI.

<sup>11</sup> Lo schema MAG, ora alla ver-

sione 2.0, si è basato sulle indicazioni del Gruppo di studio sugli standard e le applicazioni di metadata nei beni culturali dell'ICCU. Un approfondimento in: CRISTINA MAGLIANO, *Metadata: il dibattito nazionale e internazionale*, in *Conferenza internazionale. "Futuro delle memorie digitali e patrimonio culturale"*, Firenze, 16-17 ottobre 2003, a cura di C. Castellani e V. Tola, Roma, ICCU, 2004, p. 183-200.

<sup>12</sup> Il modello OAIS (Open Archiving Information Systems), sviluppato dalla NASA per l'organizzazione dei documenti scientifici, è il modello concettuale di riferimento largamente usato per gestione, preservazione e accesso alle risorse digitali. <sup>13</sup> <<http://www.aib.it/aib/cg/gb/digd05a.htm3>>.

<sup>14</sup> La Fondazione Rinascimento Digitale ha presentato i risultati di questa indagine durante il convegno del 14-15 dicembre 2006 a Firenze. I risultati, frutto di un lavoro cooperativo, sono stati pubblicati nel rapporto *Biblioteche digitali in Italia*, all'indirizzo: <<http://www.rinascimento-digitale.it/index.php?SEZ=26>>.