

Dig *Italia*

Anno VII, Numero 1 - **2012**

Rivista del digitale nei beni culturali

ICCU-ROMA

Le biblioteche nella rete dei *linked data*

Fabio Di Giammarco

Biblioteca di storia moderna e contemporanea

Tra i paradigmi più interessanti riguardo l'utilizzo delle enormi quantità di risorse informative pubblicate sul web c'è quello dei *linked data*. Si tratta di una tecnologia che non solo permette l'aggregazione dei dati attraverso la rete, ma soprattutto consente di "riaccenderli" attraverso l'interoperabilità. Per i possessori di grandi quantità di dati rappresenta un'innovazione suscettibile di infinite applicazioni. Per il mondo della cultura può rappresentare una grande opportunità, in particolare per le biblioteche può essere l'occasione per superare i troppi formalismi uscendo così da una situazione – rispetto alla rete – di isolamento informativo¹.

Karen Coyle², che da più di trent'anni si occupa di standard, formati e cataloghi di nuova generazione, ha il merito di aver posto – ormai da diverso tempo – la questione del rapporto tra Internet e i cataloghi di biblioteca. Le sue riflessioni hanno preso le mosse da una semplice circostanza, da diversi anni sotto gli occhi di qualsiasi bibliotecario: sempre più utenti di biblioteca tendono ad iniziare una ricerca di informazioni utilizzando un motore

sul web piuttosto che il tradizionale catalogo. Questo perché – ha spiegato – il catalogo appare ormai uno strumento vecchio, chiuso in se stesso e soprattutto non comunicante con tutti gli altri tipi di risorse in rete, e quindi per niente adeguato all'interattività che caratterizza il web.

Insomma, le biblioteche, come le abbiamo conosciute fino ad oggi, si trovano davanti ad uno snodo epocale: restare nel loro mondo chiuso e autoreferenziale, oppure aprire i loro dati all'interconnessione con il web. Compito davvero arduo: una sorta di rivoluzione copernicana. Si tratta per le biblioteche di trasformare – anche piuttosto velocemente visti i tempi della rete – il proprio concetto di catalogo, cioè in pratica il caposaldo della loro identità. Ma, per fortuna – come ha anche mostrato il convegno di Firenze – non stiamo ancora all'anno zero: la situazione è già in movimento.

La parola d'ordine è: interoperabilità.

I cataloghi delle biblioteche sono fermi al XIX o al massimo al XX secolo. Per fare il salto fino al XXI secolo devono diventare globalmen-

¹ Nell'interessante convegno *Global interoperability and linked data in libraries* tenutosi il 18 e 19 giugno scorsi a Firenze si è discusso sull'opportunità rappresentata dai *linked data* per le biblioteche, in particolare sull'interoperabilità e apertura dei dati bibliografici verso il mondo esterno. Sono stati anche esaminati standard, esperienze e *best practices* nell'adozione di tecnologie del web semantico nel settore bibliotecario.

² Per vent'anni Karen Coyle ha lavorato presso la University of California al progetto California Digital Library. Inoltre, ha ricoperto incarichi in vari comitati di standard come lo standard MARC, il gruppo Machine-Readable Bibliographic Information (MARBI), NISO AX e il comitato per lo standard OpenURL. È stata anche rappresentante dell'American Library Association per lo sviluppo di standard per e-book che ha portato allo standard ePub.

te interoperabili secondo i principi del web. Ma come? Cominciando a trasformare i propri formati e contenuti. I tradizionali metadati delle biblioteche non sono funzionali nell'era di Internet, non sono "collegabili" con altri set di metadati provenienti da domini diversi. Vanno cambiati.

Per fortuna, il mutamento di paradigma che deve investire il catalogo bibliografico può contare sulla forza innovativa dei principi del web semantico. Tim Berners-Lee – l'ideatore del World Wide Web – descrive l'avvento della nuova rete semantica come: «A web of thing in the world, described by data on the web», vale a dire un web costituito da cose reali descritte tramite dati. E sono quindi i dati – compresi quelli bibliografici – che collegati tra loro creano il vantaggio semantico, il passaggio da un "web dei documenti" – lineare, poco relazionato, leggibile solo dagli uomini – ad un "web dei dati", sorta di gigantesco *database* relazionale con informazioni leggibili e interpretabili dalle macchine e poi utilizzabili al massimo delle loro potenzialità dagli uomini.

A rendere concretamente possibile tutto questo ci pensano i *linked data*: una tecnologia per il web semantico nata per pubblicare e collegare i dati ad uso delle macchine. Di conseguenza, risulta evidente che il cambiamento del catalogo delle biblioteche passa attraverso l'adozione dei *linked data*.

Ma il percorso non è facile; comporta per le biblioteche una vera e propria rivoluzione suddivisibile in quattro tappe, come indicato da un report del *Library of Congress Working Group on the Future of Bibliographic Control*³: trasformazione dei metadati testuali in set di dati processabili dalle macchine; necessità che i dati siano "univocamente" identificabili nel contesto informativo del web; necessità della compatibilità dei dati con le tecnologie e gli standard del web; opportunità di

utilizzare un linguaggio trasversale e interoperabile nell'ambito del web.

Un lavoro molto interessante rispetto a queste problematiche è quello portato avanti dal *W3C Library Linked Data Incubator Group*⁴, nato per supportare e favorire lo sviluppo e la crescita della interoperabilità dei dati di biblioteche, archivi e musei sul web. Il gruppo finora ha esplorato – in ambito biblioteconomico – modelli di metadati, standard e protocolli per la creazione di sistemi di interoperabilità tra sistemi bibliotecari e ambienti di rete, incoraggiando – nello stesso tempo – le biblioteche a condividere contenuti, ad aprirsi verso altre comunità ri-orientandosi verso il web. Uno dei principali sforzi che il gruppo persegue è quello di promuovere l'uniformazione condivisa all'interno della comunità bibliotecaria intorno agli standards del web semantico. A premessa del lavoro fin qui svolto c'è la consapevolezza che i dati delle biblioteche non sono ancora integrati nel web e che gli standard – come il formato MARC o il protocollo per il recupero delle informazioni Z39.50 – sono stati (e continuano ad essere) sviluppati solo in funzione dello specifico contesto biblioteconomico.

Al fine di effettuare un'analisi dei benefici che i *linked data* possono apportare alle biblioteche, l'*Incubator Group* ha monitorato tutta una serie di casi di studio. Tra questi, l'esame di come la tecnologia dei *linked data* possa aiutare una grande biblioteca nazionale come la BnF (Bibliothèque nationale de France) – che ha già reso disponibili differenti tipi di risorse sul web – a mettere insieme dati provenienti da fonti diverse, a migliorare la pubblicazione delle risorse nel catalogo *on line* nonché ad allinearsi e collegarsi ad altre fonti informative sul web. Oppure, per quel che riguarda le biblioteche digitali, ha verificato come nel caso di *Europeana*⁵ – servizio che col-

³ <http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf>.

⁴ <http://www.w3.org/2005/Incubator/ld/>.

⁵ <http://www.europeana.eu/portal/>.

lega archivi, biblioteche, musei e materiale audiovisivo in tutta Europa – i *linked data* possano svolgere un ruolo fondamentale nel migliorare l'interoperabilità semantica tra modelli di metadati, arricchire i metadati esistenti, ottimizzare l'*harvesting* di dati e link e anche potenziare le procedure di ricerca fornendo un più facile accesso a terze parti⁶. Infine, è stato preso in considerazione anche il caso di un grande catalogo bibliografico *on line* (circa 25 milioni di record), sorta di prototipo di quel catalogo del XXI secolo perfettamente interoperabile secondo gli standard web. Vale a dire l'*Open Library*, biblioteca virtuale che mette insieme metadati di diversa natura: bibliografici, provenienti da Amazon e da altri editori, e anche generati dagli stessi utenti di OL. I dati della *Open Library*, non applicando gli standard delle biblioteche, non interagiscono con ambienti strutturati e, proprio per questo, risultano esportabili nel web e di conseguenza di facile incorporazione negli attuali reticoli dei *linked data*.

Tuttavia, l'attuale vantaggio dovuto alla "leggerezza" strutturale e approssimazione del "modello *Open Library*", potrebbe rivelarsi nel prossimo futuro del web semantico un handicap. Infatti, se è vero che oggi le biblioteche hanno bisogno della rete, è altrettanto vero che la rete avrà bisogno delle biblioteche. La fiducia della veridicità e della qualità dei dati – come ci ricorda la *true economy* – sarà l'investimento più importante per il web del futuro.

In questo contesto, diverrà fondamentale la provenienza delle informazioni, dal momento che è la fonte a conferire autorità al dato. Di

conseguenza, le biblioteche ma anche musei e archivi che hanno tradizionalmente strutturato e organizzato informazioni poi convertite in dati di qualità, se sapranno combinare strutturazione e apertura, potranno – nei futuri reticoli di *Linking Data Cloud* – ritagliarsi spazi di prestigio. Anche perché saranno in grado di contribuire in maniera significativa ad incrementare una caratteristica vitale per i *linked data*: il riuso, pratica che necessita – per consolidarsi e propagarsi – dell'aggregazione del maggior numero possibile di data set selezionati e validati. Un'iniziativa che va in questa direzione è il progetto *Linked Heritage*⁷ – finanziato attraverso l'ICT Policy Support Programme e coordinato dall'ICCU – che vede la partecipazione di 20 nazioni europee e decine di istituzioni culturali dei diversi settori. Suo obiettivo è – attraverso l'implementazione di una piattaforma multilingue e mediante coinvolgimento degli editori privati – quello di mettere insieme milioni di dati afferenti al dominio culturale, utilizzabili – mediante il servizio *Europeana* – come *linked open data*.

Ma, a parte progetti e strategie tese al superamento dell'isolazionismo catalogo-centrico, le biblioteche – come sostiene Karen Coyle – «per integrarsi in fretta con la tecnologia che la gente utilizza», hanno nell'immediato anche bisogno di assimilare i nuovi linguaggi informatici, a cominciare dalla "grammatica" del web semantico: l'*RDF*⁸. E non si tratta di inutili tecnicismi, perché come ha spiegato Tom Baker⁹ nel suo intervento al *Global Interoperability and Linked Data in Libraries*: «l'idea di web semantico può essere affrontata come fenomeno linguistico. Il modo che

⁶ Sulla pubblicazione di *Linked Open Data* per la cultura da parte di *Europeana*, vedi Fabio Di Giammarco, *I linked open data? Opportunità per la cultura*, «Il Sole 24 ore», 11 marzo 2012, supplemento *Nòva*.

⁷ <http://www.linkedheritage.eu/>.

⁸ Resource Description Framework (RDF) è lo strumento base proposto da W3C per la codifica, lo scambio e il riutilizzo di metadati strutturati, che consente l'interoperabilità tra applicazioni che si scambiano informazioni sul Web.

⁹ Dublin Core Metadata Initiative (DCMI).

permette l'integrazione di dati diversi in modo coerente si presenta come una lingua per dati. Non si tratta di una lingua parlata. Si tratta di una lingua inventata per comunicare informazioni e pensieri fundamentalmente umani e per l'uso umano, una lingua fatta però per essere letta ed elaborata non da lettori o ascoltatori umani bensì da computer secondo algoritmi ben definiti».

Le frasi RDF con le quali i bibliotecari dovrebbero riscrivere le loro descrizioni bibliografiche – così da renderle disponibili nel reticolo *linked data* – non si compongono di parole, ma di URI (Uniform Resource Identifiers), sequenza di caratteri con il prefisso `http://`. Con le URI è possibile nominare qualsiasi cosa: persone, luoghi, epoche storiche, servizi, persino concetti astratti. La regola è che ogni frase deve esprimere una singola asserzione in base ad un modello (tripartito) composto da un soggetto (per esempio "Umberto Eco"), un predicato ("è autore") e un oggetto ("Il nome della rosa"). Secondo la grammatica RDF queste frasi tripartite possono descrivere o un rapporto tra due entità, cioè tra un libro e il suo autore, oppure un rapporto tra un'entità e delle annotazioni testuali che la descrivono, come "Il nome della rosa" e data di pubblicazione "1980".

La "forza" dei *linked data* sta nel fatto che ogni asserzione diventa a sua volta generatrice di nuova informazione, venendo a creare un reticolo di triple; questo è possibile grazie al principio dell'inferenza che è quel processo per cui da una proposizione accolta come vera si passa ad una seconda proposizione la cui verità è dedotta dalla prima e così via. Spetta poi agli indirizzi URI – che sotto questa forma veicolano i dati "Umberto Eco", "è autore", "Il nome della rosa" – ancorare i significati ai fili delle reti che nel loro insieme formano il web semantico.

Ma perché poi tutto questo sistema funzioni è infine necessario che la tecnologia consenta l'univoca identificazione dei concetti, in altre parole che si utilizzino software che riconoscano gli oggetti e realizzino associazioni ed equivalenze tra essi tramite riferimento a "ontologie", cioè rappresentazioni formali, condivise ed esplicite, di specifici domini della conoscenza.

A questo punto l'affidabilità delle biblioteche come istituzioni autorevoli si trasferisce ai tanti indirizzi URI da loro creati e relative ontologie del mondo bibliotecario. Ed è questo – per le biblioteche – il modo nuovo per candidarsi come nodi centrali di un nuovo genere di catalogo "disintegrato", ma ricomponibile e ricombinabile nella crescente rete dei *linked data*.

* Per tutti i siti web l'ultima consultazione è avvenuta il 30 giugno 2012.