

Mauro Guerrini
Tiziana Possemato

Linked data per biblioteche, archivi e musei

Milano, Editrice Bibliografica,
2015, 264 p.

Uno degli imperativi della biblioteconomia contemporanea è quello del continuo e necessario adattarsi a nuovi linguaggi per la trasmissione della conoscenza registrata, linguaggi che, sotto la spinta delle sollecitazioni delle tecnologie digitali, si evolvono costantemente e rapidamente. Le biblioteche e le altre istituzioni culturali come archivi e musei non possono rimanere indietro rispetto ai modi in cui l'informazione si manifesta, si costruisce e viene comunicata, anzi, in un web sempre più caotico e affollato, devono adoperarsi – molto più che in passato – per organizzare l'informazione e garantirne la reperibilità.

Quello dei *linked data* – è stato sottolineato più volte in diversi convegni e occasioni formative – è un nuovo linguaggio della comunicazione realmente “globale”, che per la prima volta coinvolge trasversalmente il mondo dell'informazione, di chi la gestisce e la usa, consentendo un intrecciarsi di saperi e di differenti prospettive e significati, una vera e propria rete semantica di dati tra loro collegati, attraverso la quale l'informazione può essere meglio gestita e organizzata.

Il libro di Mauro Guerrini e Tiziana Possemato raccoglie organicamente lo studio che i due autori hanno compiuto negli ultimi anni e che già si era concretizzato in saggi, contributi in riviste e convegni e pertanto si propone come un'utile guida

all'universo tecnologico dei *linked data* e alla loro applicazione a musei, archivi e biblioteche.

Ciò che, in primis, si apprezza del libro è innanzitutto la chiarezza espositiva, indispensabile per rendere più appetibile a ogni tipologia di lettore un argomento particolarmente tecnico come quello delle tecnologie del Web semantico e che manifesta – come è già avvenuto per il volume su RDA, uscito in precedenza nella stessa collana e scritto dallo stesso Guerrini con Carlo Bianchini – la volontà di ampliare la platea dei lettori a chiunque voglia approfondire l'argomento (studenti, archivisti, bibliotecari, operatori di musei e chiunque si occupi di strutturare l'informazione per la sua diffusione nel web). In realtà, il legame tra i due volumi non è solo di tipo stilistico, ma anche e soprattutto di tipo concettuale: come gli autori sottolineano giustamente nella premessa, questo volume costituisce la naturale prosecuzione di *Introduzione a RDA* (Milano, Editrice Bibliografica, 2014), poiché i *linked data* esprimono la dimensione tecnologica ideale per concretizzare RDA, le linee guida specificamente concepite per la creazione di dati bibliografici e la scoperta delle risorse nel web.

Per spiegare una materia così complessa, gli autori hanno articolato l'argomento in dieci capitoli, completati da un saggio conclusivo di Carlo Bianchini e corredati da un'appendice e da un utile glossario dei numerosi termini tecnici che appaiono nel volume. Idealmente, però, è possibile suddividerne i contenuti in una prima parte (cap. 1-7) riguardante le tecnologie semantiche e in una seconda (cap. 8-10 e saggio conclusivo) più centrata sull'applicazione dei *linked data* alle risorse di archivi, musei e biblioteche.

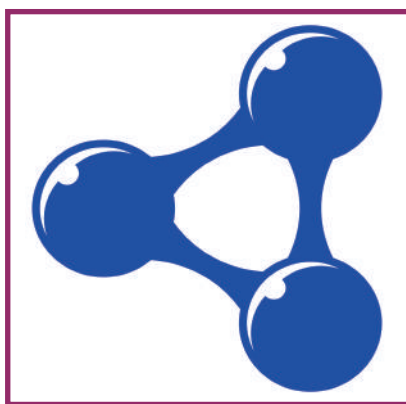
I primi capitoli hanno lo scopo di delineare il contesto tecnologico soffermandosi sui principali standard e tecnologie che caratterizzano il Web semantico e i *linked data* (cap. 1-4).

Si traccia così (cap. 1, *Web semantico e linked data*) lo scenario di innesto dei *linked data* nella strategia dell'informazione pensata da Tim Berners Lee con la formulazione dell'idea di Semantic web, un web di dati caratterizzato come estensione e naturale evoluzione dell'odierno web dei documenti, una “rete di cose del mondo descritta da dati nel web” e se ne esplorano le peculiarità che lo distinguono dall'attuale Rete (autoconsistenza del dato, libertà nell'utilizzo di vocabolari e ontologie, capacità di deduzione di nuova informazione mediante i processi inferenziali).

Il capitolo 2 (*Standard, linguaggi e tecnologie dei linked data*) ha lo scopo di introdurre gli standard, i linguaggi e i protocolli che costituiscono lo *stack* tecnologico messo a punto dal W3C per consentire a chiunque di pubblicare i dati nel web, costruire i vocabolari più adatti a rappresentare la terminologia più appropriata e le regole per gestire i dati sotto forma di *linked data*. Queste si basano su alcuni elementi essenziali: gli identificatori univoci per le risorse (URI), il modello per rappresentare e collegare i dati (Resource Description Framework, RDF) che consente di strutturare l'informazione in asserzioni semplici, i nuovi link RDF che permettono di collegare gli oggetti tramite relazioni qualificate. Attraverso queste semplici regole è possibile costruire set di dati (dataset) che, al pari delle risorse digitali, devono offrire garanzie di qualità, autenticità, affidabilità, integrità e accesso perpetuo ai contenuti (p. 69.) Successivamente (cap. 3) si introdu-

ce il terzo elemento fondante del Semantic web, rappresentato da vocabolari e ontologie, che consentono a ogni comunità di utilizzare la terminologia più appropriata per rappresentare i dati nel web. La scelta di vocabolari e ontologie appropriate riveste una grande importanza nella creazione di un dataset, in termini di struttura e funzionalità e di possibilità di riutilizzo e diffusione dello stesso. Dopo una breve introduzione ai principali linguaggi per la creazione di ontologie e vocabolari (SKOS, OWL RDFs), vengono così passati in rassegna i principali schemi di metadati per descrivere le risorse nel web e le più note ontologie per la descrizione bibliografica e archivistica e per la descrizione di persone, luoghi, diritti d'uso e metadati per la conservazione a lungo termine.

Altri due capitoli sono dedicati ad approfondire le tecnologie semantiche. Il capitolo 4 riassume brevemente le tecnologie del Web semantico esemplificate in varie rappresentazioni della pila tecnologica, attraverso la quale si può avere una visione complessiva delle tecnologie pensate da Tim Berners Lee e dai creatori del Semantic web. Interamente dedicato ai metodi di pubblicazione dei dati è il capitolo 5 (*Pubblicare linked data nel Web*) che si sofferma dapprima sulle principali prerogative per garantire la qualità dei dati immessi nel web in relazione all'informazione all'accesso e ai servizi ed espone in seguito le fasi che caratterizzano il ciclo di vita dei LOD, che vanno dalla scelta del dataset alla sua pubblicazione e corretta metadattazione. Si esamina la composizione della *lod cloud*, identificandone le origini nel "Linking open data project", iniziativa che ha condotto alla formazione di una grande nuvola di dataset di varia tipologia e provenienza, tra



loro correlati. Il collegamento dei dati del settore culturale alla nuvola riveste un enorme potenziale: implica infatti la più ampia fruibilità dei dati bibliografici, il potenziamento della semantica, la creazione di valore culturale ed economico dei dati stessi (p. 116).

Due brevi capitoletti (cap. 6-7), infine, si occupano rispettivamente delle licenze per i dati, prerogativa indispensabile per renderli davvero fruibili e riutilizzabili, e della normativa italiana ed europea in materia di apertura e trasparenza dei dati.

Più interessanti — e a mio avviso originali — i capitoli che compongono idealmente la seconda parte del volume, dove si entra nel merito delle applicazioni dei *linked data* al settore bibliotecario. Il passaggio dal record al dato segna un cambiamento epocale che porterà le biblioteche a produrre dati interoperabili e riutilizzabili nel web. "Perché l'informazione sia del Web e non solo nel Web", come recita il sottotitolo del libro, le biblioteche devono però compiere alcuni passi importanti: modificare la tradizionale struttura monolitica delle registrazioni bibliografiche, adottare formati di scambio più adatti al web, applicare nuove norme che corrispondano alle nuove funzioni che i cataloghi possono svolgere nel "web dei dati".

Il capitolo 8, *Biblioteche e Web Seman-*

tico, introduce i vantaggi che il mondo degli archivi, delle biblioteche e dei musei possono trarre dall'utilizzo dei *linked data*. I principali distributori dell'informazione possono essere pensati come un mondo esteso, disseminato, caratterizzato dalla ricchezza e al contempo dall'eterogeneità delle risorse, ma accomunato dalla necessità di strutturare i dati e proporre standard finalizzati alla registrazione dell'informazione. In questo senso si coglie l'interesse del Web semantico per questi *silos* informativi rimasti per molto tempo inaccessibili e non interoperabili né tra loro né con il resto del web e al contempo si rivela l'interesse degli istituti della memoria registrata per il Web semantico stesso, un interesse che le biblioteche, in particolare, hanno manifestato già con il rapporto *On the record* (2008), a partire dal quale si è avviato un processo irreversibile verso una riconsiderazione del record bibliografico e la sua trasformazione in un set di dati utilizzabili e processabili dalle macchine, aprendo così la strada dell'avvicinamento al Web semantico. In questo contesto il *metadato* dalla sua accezione tradizionale si trasforma in *dato significativo*, cioè processabile e interpretabile da una macchina. Nel Web semantico il record atomizzato e trasformato in RDF può così decostruirsi in un insieme di dati in cui ogni singolo elemento acquista significato componendosi con altri dati e generando asserzioni riusabili da comunità diverse che operano nel web. Ciò presuppone che si adottino nuovi modelli catalografici più granulari, che si esemplificano oggi nello standard RDA, giudicato dagli autori come schema perfetto per l'evoluzione delle norme e degli standard nella direzione dei *linked data*. È proprio dalla spe-

rimentazione di RDA che è nata e cresciuta la consapevolezza di come il tradizionale record in formato MARC fosse ormai totalmente inadeguato a esprimere la complessità delle relazioni previste dai modelli FR dell'FLA. Questa consapevolezza ha condotto, a partire dal 2011, al lancio della "Bibliographic framework transition initiative" che ha costituito l'avvio per lo studio di un nuovo ecosistema bibliografico caratterizzato da nuove regole per la descrizione e l'accesso alle risorse, nel quale vi è una puntuale identificazione del dato, una maggiore enfasi per le relazioni tra i dati e un impiego più diffuso di termini provenienti da vocabolari controllati. Il nuovo BIBFRAME propone dunque un modello di dati ispirato a FRBR, consentendo al contempo il riutilizzo dei dati bibliografici in contesti diversi, proponendosi come modello per lo scambio dei dati bibliografici nel Web semantico. Il passaggio dal record al dato dunque porterà alla creazione di ILS di nuova generazione che ci si auspica di vedere a breve. In questo senso vengono anticipate dagli autori alcune recenti implementazioni di sistemi orientati a RDA e BIBFRAME (8.6.1 e 8.6.2). Nel capitolo 9 sono recensiti i più grandi e significativi progetti di pubblicazione dei dati sotto forma di *linked open data* (LOD) da parte delle più importanti biblioteche italiane ed europee. Si tratta di progetti che, in attesa dell'adozione dei *linked data* come formato nativo per i record bibliografici che avverrà solamente con l'adesione a RDA e BIBFRAME, mirano a concretizzare il modello dei *linked data* nei cataloghi delle biblioteche e attraverso l'esposizione dei dati bibliografici. Oltre a progetti nazionali di un certo rilievo come camera.dat.it e SBN

in LOD, si richiamano altre importanti e note iniziative come quelle messe in atto dalle biblioteche nazionali di Francia, Regno Unito, Germania e Spagna. Queste istituzioni sono state pioniere nello sviluppo di piattaforme e servizi LOD e hanno avuto il merito di creare interfacce di ricerca innovative e di condividere dati bibliografici con la comunità del web attraverso appositi *linked data services*. Tra i nuovi progetti si accenna anche ad alcune piattaforme integrate che consentono l'intera gestione del *workflow* di trasformazione dei dati bibliografici in LOD come il progetto ALIADA (9.8). Il capitolo trova un suo completamento nell'appendice curata da Iryna Solodovnik, in cui sono presentati brevemente i principali schemi e ontologie "LOD ready" e alcune esperienze di *digital libraries* che hanno già accolto la pratica operativa dei *linked open data*. Infine (cap. 10) sono brevemente esposti i principali vantaggi per l'utenza che utilizzerà i dati sotto forma di *linked open data*. Se, infatti, la comunità del Web semantico ha avvertito l'importanza e il valore dei dati prodotti dalle istituzioni del patrimonio culturale - informazione ricca, strutturata, controllata e autorevole - altrettanto importanti saranno i benefici di chi fruisce di questi dati (tutti gli utenti del web) in termini di nuovi servizi e applicazioni, di accesso globale a risorse distribuite, di funzionalità di ricerca semantica in siti, applicazioni, OPAC e portali che sfrutteranno le tecnologie dei *linked data*. Di particolare interesse per i *library linked data* è poi il saggio conclusivo di Carlo Bianchini *Dai cataloghi alla navigazione semantica*, che ha il merito di chiudere il volume con una riflessione profonda sull'evoluzione

semantica del catalogo. L'autore si sofferma sulla parabola evolutiva del catalogo a partire dalla profonda riflessione teorica del modello logico apportata dal modello FRBR. Lo studio dell'IFLA ha avviato un ripensamento delle funzioni di questo strumento, evidenziando al contempo la necessità di nuove soluzioni tecnologiche che fossero in grado di evidenziare un modello logico più esplicito, e ha originato un profondo mutamento che ha cambiato la rotta della teoria catalografica: da questo studio si sono originati nuovi principi e norme catalografiche realmente innovative come RDA, vere e proprie linee guida per l'accesso e la scoperta di qualsiasi tipo di risorsa. Queste ultime sono uno standard di contenuti flessibile, modulare, adattabile a rappresentare ogni tipologia di dato e applicabili a differenti contesti tecnologici (più o meno complessi) e pertanto di rilevanza globale. In RDA l'approccio è atomizzato e granulare, le relazioni sono qualificate e l'attenzione si sposta dal record al singolo dato. In totale sintonia con l'approccio del Web semantico entreranno così a far parte del catalogo anche dati prodotti da altre comunità del web, e provenienti da altri archivi non bibliografici. Tra i vantaggi per l'utente vi sarà la navigazione semantica del catalogo, più centrata sul recupero dell'informazione nel suo contesto, resa possibile dalle nuove opportunità di condivisione e riutilizzo dei dati. Il concetto di riutilizzo dei dati e di interoperabilità tra sistemi differenti che dialogano tra loro per mezzo di standard e vocabolari condivisi porta con sé il nuovo concetto di responsabilità culturale che riguarda coloro che gestiscono dal punto di vista semantico, linguistico e concettuale i vocabolari e le ontologie messe in Rete; chiun-

que infatti potrà riutilizzare le voci di vocabolario create e queste potranno essere utilizzate in maniera automatica per creare inferenze e nuove deduzioni: pertanto la creazione di vocabolari e ontologie rappresenta una scelta strategica e culturale importante nel web dei dati per la comunicazione globale.

In conclusione, la lettura del volume di Guerrini e Possemato ci restituisce appieno un universo tecnologico complesso pieno di opportunità di sfide in cui biblioteche, archivi e musei si trovano ad affrontare una scelta strategica: abbracciare i *linked data* vuol dire non solo integrare fonti informative eterogenee in piattaforme che consentano di interrogare i dati del patrimonio culturale, ma anche prendere parte attiva allo sviluppo del web nell'ottica dell'integrazione dei saperi e delle conoscenze scientifiche.

ANTONELLA IACONO

Dottore di ricerca in Scienze Librarie
e documentarie
Università di Roma La Sapienza

DOI: [10.3302/0392-8586-201506-054-1](https://doi.org/10.3302/0392-8586-201506-054-1)