
Antonella Iacono

Linked data

Roma, AIB, 2014, p. 109, € 12,00
(Enciclopedia Tascabile, 32)

In questa collana di agevoli manuali arriva finalmente un libretto sui linked data, attuale frontiera del Web semantico. Il termine, coniato da Tim Berners-Lee già nel 2001, indica una concezione della rete che passa da un web di documenti, collegati tra loro attraverso link ipertestuali,

a un web di dati, dereferenziati, atomizzati, che si collegano tra in una nuvola in continua espansione.

I linked data sono la tecnologia che rende possibile e concretamente attuata questa concezione del web. Il passaggio è stato rappresentato dal Web 2.0: il web che si è aperto alle community e ai social network, creando così un primo livello di interazione tra documenti e pubblico, trasformatosi in *prosumer*, cioè sia produttore sia consumatore di informazione. Presto però ci si è accorti dei limiti di questa modalità, che non prevede nessun controllo bibliografico e della catena editoriale, che per secoli ha costituito la garanzia per un'informazione certificata e validata; nel Web 2.0 tutti noi possiamo scrivere, ma siamo autorizzati a scrivere anche banalità, falsità, sciocchezze, e tutto quello che non viene sottoposto a una validazione scientifica.

Da qui l'esigenza di una rete di collegamenti basata su strutture informative certe e documentabili; resta quindi la predominanza dell'utente, resta un'unica interfaccia amichevole, resta l'approccio a tecnologie facili, ma dietro c'è un lavoro di strutturazione ricco e complesso (in fondo i cataloghi di biblioteca questo hanno rappresentato da sempre: l'interfaccia amichevole di un addetto che ti prende il libro, e dietro un complesso lavoro di bibliotecari che strutturano il catalogo affinché quel libro possa essere trovato).

Il primo concetto base è open data: i dati devono essere aperti e accessibili, senza vincoli proprietari, solamente così la tecnologia dei linked data trova il suo senso; una tecnologia studiata per far dialogare le macchine e creare quella struttura inferenziale che nel catalogo è leggibile solamente dalla mente umana.

Alla base di tutto troviamo il model-

lo RDF,¹ costruito su semplici triple dove un soggetto e un oggetto sono legati da un predicato (Calvino è autore del *Barone rampante*); affinché questa asserzione possa essere letta dalla macchina occorre che ogni elemento diventi una "cosa" identificata da un URI,² che, a differenza dell'URL (il quale rimanda a una pagina web), identifica in maniera permanente quella "cosa" e solo quella "cosa", in tal modo la precedente asserzione diventa una tripla di URI tra loro collegati (URICalvino URIè autore del URIBarone rampante) leggibili dalla macchina attraverso la sintassi specifica RDF/XML.³

Affinché il tutto possa avere un significato semantico univoco, a monte devono essere strutturate ontologie e vocabolari che definiscano le diverse entità che compongono una tripla: nell'esempio precedente "Calvino" è descritto come persona nell'ontologia FOAF⁴ e come autore nell'ontologia VIAF⁵ che lega tra loro, come un authority file, tutte le forme che nel mondo designano un autore;⁶ "è autore di" sarà presente in un vocabolario che definisce i vari attributi; "*Il barone rampante*" nell'ontologia di opere, dove ognuna è legata alle diverse espressioni e manifestazioni e item⁷ che la rappresentano. Naturalmente anche ontologie e vocabolari sono dotati di un linguaggio macchina e sono certificati e resi disponibili nel registro CKAN.⁸

Il libro fornisce utili informazioni sia riguardo i diversi linguaggi sia riguardo ontologie e vocabolari già strutturati e utilizzabili per la formulazione di linked data. Ma ora viene il bello: come possono le biblioteche, che già sono in possesso di dati strutturati, ma chiusi negli OPAC consultabili solamente dall'interfaccia della singola biblioteca, aprirsi al web rendendo questi dati interoperabili⁹

con tutti gli altri presenti nella rete e visibili attraverso qualsiasi interfaccia? Il lavoro che si sta già facendo, soprattutto nelle biblioteche nazionali, per rendere i dati catalografici aperti e strutturati come linked data è notevole; l'esempio che in questo momento ci sembra meglio riuscito è quello della BNF (Bibliothèque Nationale de France),¹⁰ che presenta un'interfaccia ricca di informazioni e relazioni di facile consultazione; la Nazionale di Firenze ha strutturato il Nuovo Soggettario¹¹ come ontologia strutturata per i linked data; ma tutta la comunità internazionale dei bibliotecari è impegnata nella trasformazione dei propri strumenti di base (Dublin core, ISBD, CDD ecc.) in ontologie e vocabolari dereferenziati e interoperabili.

Ma il proprio dataset¹² va strutturato e pubblicato; per fare questo occorrono diversi passaggi: "identificazione delle fonti dalle quali estrarre i dati" (p. 57), forniti ognuno del proprio URI, "riutilizzo di vocabolari già esistenti [...] o la creazione di nuove ontologie" (p. 58), "generazione di file di dati RDF" (p. 58), operazione ormai facilitata da diversi applicativi anche *open source*, "creazione di collegamenti con altri dataset" (p. 58), disponibili e certificati, "pubblicazione del dataset nel Web semantico per renderlo disponibile al riutilizzo" (p. 59); operazione che comporta contestualmente la pubblicazione anche dei metadati, la registrazione del dataset in CKAN, e il dotarsi di un motore specifico secondo la sintassi SPARQL¹³ che sappia interrogare i linked data, ancora non rintracciabili attraverso i comuni motori di ricerca.¹⁴ Alla fine di questo processo bisogna dotarsi delle licenze che consentano ai dataset di essere liberamente consultabili e facciano dei linked data dei Linked open data (LOD), che si di-

stinguono in una gerarchia di 5 stelle: 1) l'informazione è disponibile sul web (qualsiasi formato) sotto una licenza aperta; 2) l'informazione è disponibile sotto forma di dati strutturati (es. Excel); 3) uso di formati non proprietari (es. CSV);¹⁵ 4) uso di standard W3C come RDF e SPARQL e di identificativo URI; 5) i dati sono collegati tra loro tramite RDF (p. 71).

Un esempio di come tutto questo debba incidere anche sul lavoro di catalogazione lo possiamo trovare in come cambia la descrizione in ISBD: ogni elemento della descrizione deve essere dereferenziato e atomizzato e dotato di un identificativo URI; sparisce il concetto di area, per es. quella della pubblicazione, e ogni elemento, luogo, editore o anno diviene singolo, dotato di URI, legato a un'ontologia per poter diventare soggetto a sua volta di un'altra tripla, e così via.

Nuove regole di catalogazione sono già state formulate, in particolare RDA,¹⁶ che fa proprio il modello concettuale di FRBR¹⁷ di entità, attributi e relazioni, e dotato di 76 vocabolari, ormai consente una catalogazione orientata ai linked data, attraverso anche la semplificazione operata dal modello BibFrame,¹⁸ che "consentirà di scambiare risorse bibliografiche in rete, garantendo che vengano prodotti dati bibliografici in cui siano chiaramente identificate le entità (autori, editori) coinvolte nella creazione di una risorsa e i concetti (soggetti) a essa associati" (p. 83); in quanto semplificazione del modello FRBR: qui "opera" comprende opera e espressione; "istanza" comprende manifestazione e item; "authority" comprende persone, luoghi, concetti; "annotazioni" comprende paratesto ed esotesto. Anche i concetti devono essere formulati in maniera univoca attraverso ontologie specifiche e dotati del proprio URI. A questo prov-

vedono il modello OWL¹⁹ per costruire ontologie di soggetti, e il modello SKOS²⁰ che "definisce classi e proprietà per esprimere la struttura di base e il contenuto di schemi concettuali come thesauri, schemi di classificazione, tassonomie" (p. 86). In questo ambiente di tecnologia avanzata, dove si consente alle macchine di elaborare anche entità come i concetti, trova nuovo significato il lavoro delle biblioteche, le quali, da sempre attente alla definizione concettuale dell'informazione, possono ora portare questa esperienza e tutti i loro dati già strutturati negli OPAC, all'intera comunità del web, in uno scambio che dia peso e valore all'informazione che non nasce dal web e per il web, ma che ormai nel web trova il suo maggior mezzo di diffusione. Quindi ben venga questa agevole "voce d'enciclopedia", che nella sua sintetica esposizione, peraltro corredata da una bibliografia di base e da un glossario dei termini tecnici e delle numerose sigle impiegate nella tecnologia dei linked data, costituisce il primo approccio, amichevole e chiarificatore, per quei bibliotecari che si vedono inquieti di fronte alla trasformazione del loro lavoro. Ma da quel lontano 1961 che vide la prima formulazione di *Principi di catalogazione* standardizzati, e ancor più dall'avvento pervasivo di Internet, la comunità dei bibliotecari si è sempre fatta trovare pronta a recepire le nuove necessità che nascevano nella società, ed è stata la prima a elaborare nuovi strumenti che aprissero gli ambienti "polverosi" delle biblioteche a un pubblico sempre più vasto, eterogeneo ed esigente.

ANGELO ARIEMMA

Università Sapienza, Roma
CDE "Altiero Spinelli"
angelo.ariemma@uniroma1.it

NOTE

¹ Resource Description Framework: standard model for data interchange on the Web, <<http://www.w3.org/RDF/>>.

² Uniform Resource Identifier: una stringa che identifica univocamente una risorsa.

³ <<http://www.w3.org/TR/rdf-syntax-grammar/>>.

⁴ <<http://xmlns.com/foaf/spec/>>.

⁵ Virtual International Authority File, <<http://viaf.org/>>.

⁶ Vengono così meno anche i problemi legati alle diverse lingue.

⁷ Per questa terminologia rimando alla mia recensione sul modello FRBR in "Biblioteche oggi", 2002, n. 6, p. 58-59.

⁸ Comprehensive Knowledge Archive Network, <<http://ckan.org/>>.

⁹ "Capacità di due o più sistemi, reti, mezzi, applicazioni o componenti, di scambiare informazioni tra loro e di essere poi in grado di utilizzarle" (Enciclopedia Treccani).

¹⁰ <<http://data.bnf.fr/>>.

¹¹ <<http://thes.bncf.firenze.sbn.it/ricerca.php>>.

¹² "Insieme di dati strutturati in forma relazionale" (Wikipedia).

¹³ SPARQL Protocol and RDF Query Language, <<http://www.w3.org/TR/rdf-sparql-query/>>.

¹⁴ Si veda l'esempio sul sito dei linked data del Senato, <<http://dati.senato.it/23>>.

¹⁵ "Comma-separated values è un formato di file basato su file di testo utilizzato per l'importazione di una tabella di dati" (Wikipedia).

¹⁶ Resource Description and Access, <<http://www.loc.gov/aba/rda/>>. Si veda CARLO BIANCHINI - MAURO GUERRINI, *Introduzione a RDA*, Milano, Editrice Bibliografica, 2014.

¹⁷ Dobbiamo pur dire, a vanto di chi, come noi, ne intuì subito il valore di cambio di paradigma della struttura catalografica, che il modello FRBR, elaborato dall'IFLA, precede anche le formulazioni di Tim Berners-Lee.

¹⁸ Bibliographic Framework Initiative, <<http://www.loc.gov/bibframe/>>.

¹⁹ Ontology Web Language, <<http://www.w3.org/TR/owl-features/>>.

²⁰ Simple Knowledge Organization System: common data model for sharing and linking knowledge organization systems via the Web, <<http://www.w3.org/2001/sw/wiki/SKOS>>.

DOI: 10.3302/0392-8586-201410-061-1