

ISBD nel web semantico

Lectio magistralis di Elena Escolano Rodríguez
all'Università di Firenze

Con "ISBD nel web semantico" di Elena Escolano Rodríguez giunge al quinto appuntamento la serie delle lectio magistralis in biblioteconomia proposte nell'ambito del Master biennale di secondo livello in archivistica, biblioteconomia e codicologia dell'Università di Firenze coordinato da Mauro Guerrini.¹ La serie, inaugurata nel 2008 da Barbara Tillett e proseguita con i contributi di Patrick Le Boeuf (2009), Rahmatollah Fattahi (2010) e Lynne C. Howarth (2011), è stata caratterizzata fino a ora da una particolare attenzione verso problemi e questioni che interessano insieme la teoria e la prassi catalografica contemporanea. Il filo rosso che l'attraversa è il confronto della disciplina con un universo bibliografico e un universo degli utenti in continua evoluzione; il quinto appuntamento non fa eccezione. La *lectio* di Elena Escolano, infatti, che ha avuto luogo il 6 marzo 2012 presso la Sala "Comparetti" della Facoltà di lettere e filosofia, prosegue nel solco tracciato dai precedenti incontri e ragiona sull'importanza della qualità delle informazioni in un ambiente come quello del web, e del web semantico più in particolare; qualità che, nel loro ambito, le biblioteche hanno da sempre contribuito a garantire con il con-



trollo bibliografico. Ora, in un contesto più ampio com'è quello del web semantico, composto da differenti domini, ognuno codificato nella propria semantica da ontologie formali, la questione della qualità e dell'affidabilità dei dati è cruciale. Su questa base dovrebbe fondarsi l'architettura del web semantico, in cui vengono realizzati i collegamenti tra dati (*linked data*) e create le relazioni tra ontologie di domini differenti. In un ambiente così interconnesso la scoperta di nuovi oggetti sarà più facile, sarà possibile accedere a una gran quantità d'informazioni e navigare verso quelle collegate, un po' come succede quando si passeggia tra i libri di una biblioteca a scaffale aperto o si naviga tra i record presenti in un OPAC. Appare evidente ancora una volta il parallelo tra le dinamiche relazionali del web e quelle bibliotecarie della cooperazione e della condivisione delle registrazioni – scopo del controllo bibliografico. Sottolinea quindi la relatrice come le biblioteche e, più in generale, tutte le istituzioni che hanno tradizionalmente selezionato, strutturato e organizzato l'informazione, convertendola in dati di qualità, soddisfino i requisiti per occupare un posto di rilievo nel web e partecipare attivamente all'elaborazione di onto-

logie, alla creazione di collegamenti tra questi dati di qualità e, più in generale, allo sviluppo del web semantico.

Evidentemente in tale ambiente la visione diventa ora più granulare, l'attenzione si sposta dal record per concentrarsi sul dato, che passa a essere la nuova moneta di scambio per la cooperazione con altre biblioteche e altre realtà. Tramite esso le biblioteche possono comporre il proprio record e continuare a perseguire i medesimi obiettivi: permettere d'identificare per poter selezionare e avere accesso alla risorsa. Inoltre, possono sfruttare le potenzialità dei *linked data* per suggerire o proporre informazioni relazionate.

Il paradigma dei *linked data* applicato alle biblioteche, da una parte offre loro nuovi vantaggi e soluzioni a problemi attuali, dall'altra sollecita la disciplina a ragionare intorno a problematiche nuove e a rinnovarsi per non rimanere estranea ai nuovi canali di circolazione dell'informazione e del sapere. I vantaggi per le biblioteche che rendono i propri dati disponibili in accesso aperto (*linked open data*), permettono il loro riutilizzo e li collegano tra loro sono molteplici: estensibilità e integrazione con dati di altre istituzioni per fornire all'utente il maggior numero di informazioni elaborate da comunità diverse, in un'ottica di economicità; omogeneità con le istanze di trasparenza della pubblica amministrazione, anche al fine di evitare la proliferazione di progetti tra loro non coordinati e il conseguente spreco di energie e risorse; comprensibilità della semantica e del linguaggio bibliotecari e possibilità di presentare i risultati delle ricerche nella lingua dell'utente grazie alla catalogazione cooperativa multilingue.²

Per partecipare allo sviluppo del web semantico non basta solamente creare i dati e collegarli, ovvero i *linked data* non sono sufficienti; occorrono altri quattro requisiti che sono stati individuati da Tim Berners-Lee: usare URI per identificare o riferirsi alle risorse; usare URI di HTTP, affinché l'utente possa ricercare e localizzare le risorse tramite essi; fornire informazioni utili sulla risorsa, usando standard; includere collegamenti ad altri URI, affinché si possano scoprire informazioni collegate. Il web semantico corrisponde a un'estensione del web attuale dove le informazioni e i servizi si basano su dati con un significato ben rappresentato per poter essere condivisi e usati dalle persone e dalle macchine.

In tal senso nell'ambito bibliotecario molto era già stato fatto con la definizione da parte dell'IFLA di modelli e standard diffusi e riconosciuti internazionalmente; il dominio bibliotecario si distingue per essere un terreno stabile e strutturato provvisto di una semantica già ben codificata. Il passo in più che, ultimamente, l'IFLA si è proposto di fare – e che l'autrice ha voluto trattare nella sua *lectio* – è stato inserire la dichiarazione dei propri modelli e standard in RDF (Resource Description Framework).³ Tale azione, da un lato andava nella direzione di un'operazione di *branding* per difendere la qualità della propria terminologia da possibili fraintendimenti e usi impropri, dall'altro si configurava come un'occasione per rinnovare la validità dei propri standard e rivederli in prospettiva di un web semantico multilingue. A tal fine è stato creato all'interno dell'IFLA il Gruppo sui namespace, che si è occupato di gestire la redazione delle proprie ontologie in uno spazio

apposito, il namespace,⁴ identificato dall'URI: <http://iflstandards.info/ns>. Il gruppo sta lavorando all'interno dell'Open Metadata Registry (OMR), un'infrastruttura tecnica realizzata dal W3C a supporto dell'interoperabilità dei metadata per il web semantico; in futuro si prevede il trasferimento delle disposizioni dell'IFLA da OMR a uno spazio appositamente dedicato sul proprio sito. In questo spazio troviamo set di elementi e vocabolari relativi ai modelli della famiglia FRBR (FRBR, FRAD, FRSAD) e a ISBD; ogni elemento e ogni concetto è identificato univocamente da un URI opaco, senza determinazione di lingua, in funzione del multilinguismo, e per ognuno di questi vengono forniti: nome del metadata, etichetta, definizione, note di ampliamento dell'informazione o della sua applicazione, tipologia, stato di accettazione, lingua ecc.

A partire da questi vocabolari sarà inoltre possibile generare mappature tra vocabolari simili di standard esterni all'IFLA (come ad esempio RDA Content Type e Media Type); a tale scopo e per supportare l'interoperabilità tra le diverse comunità esiste già uno strumento chiamato Vocabulary Mapping Framework.

È stato più volte accennato prima come la partecipazione dell'IFLA al web semantico con i propri modelli e standard vada intesa anche ai fini della difesa e dello sviluppo del multilinguismo, che è tra i suoi obiettivi strategici. Nella seconda parte della sua *lectio*, Elena Escolano approfondisce questo tema e affronta le questioni emerse in occasione della conversione da parte della Biblioteca Nacional de España del proprio catalogo in RDF per poterlo pubblicare e collegare in rete. Le problematiche incontrate hanno riguardato

lo stile delle etichette (uso delle maiuscole e addossamento delle parole, differenza di trattamento tra etichette di elementi in ISBD e quelle di elementi in FRBR), la fonte di riferimento per la traduzione delle disposizioni in RDF (etichette, definizioni e note), l'uso delle qualificazioni per disambiguare e chiarire l'orientamento della relazione, e la flessione grammaticale per il genere maschile e femminile degli aggettivi all'interno dei vocabolari. L'esperienza maturata, il *modus operandi* e le soluzioni adottati dal team che si è occupato della traduzione in spagnolo delle disposizioni dell'IFLA e della loro pubblicazione in rete rivestiranno un sicuro interesse per l'intera comunità delle biblioteche.

Quanto detto finora – rileva la relatrice – non c'è dubbio che abbia influito sui lavori dell'ISBD Review Group e sui cambiamenti avvenuti nella nuova edizione consolidata di ISBD, pubblicata nel luglio 2011. L'analisi degli elementi di ISBD si è fatta più granulare, alla ricerca di una maggiore coerenza e qualità dei dati, di una loro migliore portabilità nel web semantico, nonché di interoperabilità tra questi dati e quelli derivanti dall'applicazione di altri standard (A.1.2). Alla scelta del livello di granularità, che incide sulla coerenza della descrizione, va ricondotta anche la maggiore attenzione prestata alla corretta descrizione delle risorse monografiche multipartite. Inoltre, dal momento che, come si è detto, la fonte d'informazione determina i dati, è stata riformulata più chiaramente la differenza di redazione che sussiste tra fonte preferita e fonte prescritta. Nel caso di informazioni inferite dal catalogatore o provenienti da fonti esterne a quelle prescritte, si ribadisce l'importanza dell'uso

delle parentesi quadre per rendere riconoscibile il nostro contributo, anche dalla macchina. Ulteriori modifiche sono intervenute nella nuova edizione per renderla più funzionale all'interpretazione da parte del computer, che ancora non è in grado di inferire informazioni, come nell'Area A, dove ora è sempre obbligatorio informare sul tipo di supporto.

Si è detto in precedenza dell'importanza che rivestono gli standard per i *linked data* e il web semantico e della necessità di realizzare corrispondenze tra questi. Nell'ultima parte del suo intervento ci si è soffermati sulla riunione di armonizzazione ISBD-ISSN-RDA, tenutasi a Glasgow il 3 e 4 novembre 2011, che ha messo in evidenza importanti differenze e l'opportunità appunto di una loro armonizzazione. Le divergenze maggiori emerse nella riunione sono state relative ai criteri di scelta della fonte preferita e al loro ordine, all'esistenza di elementi con lo stesso nome ma con definizioni e note d'applicazione differenti, all'uso dei dispositivi (parentesi quadre, note relative alla fonte, segni di omissione) appositamente ideati per garantire l'attendibilità delle informazioni fornite. Queste divergenze generano purtroppo problemi di interoperabilità, in quanto, scegliendo di applicare una regola piuttosto che un'altra, si può dare luogo a diverse descrizioni.

A questi problemi si è cercato di trovare soluzioni condivise che contribuiscano a riavvicinare tra loro gli standard (come ad esempio la proposta di elaborare un profilo di applicazione per le agenzie bibliografiche nazionali).

In conclusione è stata sottolineata l'importanza che ricoprono per il successo del web semantico la veri-

dicità, la qualità e l'attendibilità del dato. In questa direzione stanno andando le azioni e gli sforzi fin qui compiuti dall'IFLA per la promozione e diffusione dei propri standard tramite la creazione e il continuo ampliamento del proprio namespace, al quale attingono diversi progetti in corso. Il riconoscimento del valore delle attività dell'IFLA e del lavoro dei suoi bibliotecari, impegnati nel favorire l'avanzamento della disciplina, sono stati riconosciuti dal W3C che ha accettato gli standard dell'IFLA come regole internazionali di catalogazione valide per il web semantico.

La *lectio* di Elena Escolano Rodríguez ha anticipato temi che verranno ampiamente discussi nel corso del seminario di studi dal titolo "Global interoperability and linked data in libraries" che si terrà a Firenze, nell'Aula Magna dell'Università, il 18 e 19 giugno prossimi. Per maggiori informazioni su questo seminario si visiti il sito web all'indirizzo: <<http://www.linkedheritage.org/linkeddataseminar/>>.

CLAUDIO VENTURATO

Dottorando di ricerca
in scienze bibliografiche
<claudio.venturato@gmail.com>

NOTE

¹ La serie delle "Lecture Magistrali in Biblioteconomia" è curata da Mauro Guerini e promossa dall'Università degli studi di Firenze (Dipartimento di Scienze dell'antichità, medioevo e rinascimento e Linguistica, Master biennale di secondo livello in archivistica, biblioteconomia e codicologia e Sistema bibliotecario di ateneo), insieme alla Regione Toscana, alla Biblioteca nazionale centrale di Firenze, all'Associazione italiana biblioteche, a "Biblioteche oggi", "JLIS.it" e

a Casalini Libri. Il testo della *lectio magistralis* di Elena Escolano Rodríguez è stato distribuito a tutti i partecipanti ed è disponibile in formato elettronico al permalink: <<http://digital.casalini.it/9788885297999>>. Tutti i numeri della collana *Lecture magistrali in biblioteconomia* sono editi da Casalini Libri e disponibili su <www.torrossa.it>.

² Questi sono i vantaggi emersi dalle relazioni e dai documenti prodotti dal Library Linked Data Incubator Group, creato nel maggio 2010, la cui *mission* è quella di favorire l'aumento dell'interoperabilità dei dati delle biblioteche nel web. Per maggiori informazioni su questo gruppo, si visiti la pagina dedicata sul sito della W3C all'indirizzo: <<http://www.w3.org/2005/Incubator/lld/>>.

³ RDF è un insieme di specifiche sviluppate dal World Wide Web Consortium (W3C) per modellare i metadati in una varietà di formati e sintassi. RDF è un modello standard per la codifica, lo scambio e il riutilizzo di dati nel web, ma la sua caratteristica fondamentale è quella di consentire la combinazione di dati che provengono da schemi diversi e l'interoperabilità tra applicazioni che si scambiano informazioni nel web. Il formato normalmente associato a RDF è Extensible Markup Language (XML); per questo motivo la sintassi RDF/XML è conosciuta anche semplicemente come RDF.

⁴ Il *namespace* è una sorta di contenitore astratto che ospita un insieme di termini ai quali vengono forniti un URI e un significato e un contesto precisi, che rendono possibile la disambiguazione di termini simili in uso in altri namespace.