

# Il Library Linked Data Incubator Group e il futuro delle biblioteche

di Carlo Bianchini

Abstract: In October, 2011, the Library Linked Data Incubator Group (LLD XG) created by World Wide Web Consortium (W3C) published its final report. This report suggest libraries how they should made their data available in the web (and not just on the web), to make them available also directly by search engines. To comply with requirements the report, libraries should publish their data in a more expressive form, in which also values of relationships among data must be registered. The paper illustrates main features and requirements of library linked data, as described in the report and in relation to cataloguing topic, explains issues relating to the present format of library data and discuss the opportunities offered by RDA and REICAT, the two codes of rules published in compliant with FRBR. This report could result as important for the library community as FRBR has been till now.

## 1. Introduzione

Le biblioteche sembrano assumere un ruolo sempre più di nicchia nella società dei social network e dei motori di ricerca e attraversano un periodo di crisi dovuto all'obsolescenza dei media che costituiscono le loro raccolte tradizionali (evidente nelle sempre più accese e preoccupate discussioni sul futuro del libro; Darnton 2011) e all'incertezza sul ruolo della professione (in particolare dei catalogatori; (Revelli 2011)).

Per superare questa fase di crisi e ritrovare la strada per definire un nuovo futuro per le biblioteche, è necessario cercare nella direzione dei valori fondanti delle biblioteche: individuare le specificità delle biblioteche e della professione che rimangono imprescindibili nel nuovo contesto e trovare il punto su cui fare leva, sono passaggi obbligati per superare la crisi e uscirne in modo positivo.

Una folgorante metafora di Ranganathan spiega i rapporti tra le componenti essenziali della biblioteca paragonandola a un elettromagnete: questo non è costituito dal ferro, dalle spire o dall'elettricità considerate separatamente, ma esiste solo nel momento in cui l'elettricità, passando attraverso le spire del nucleo, crea il campo magnetico. Altrettanto una biblioteca

*inizia ad esistere solo se lettori, libri e personale funzionano insieme. Lettori, libri e personale formano una triade in una biblioteca. (Ranganathan 1963)<sup>1</sup>*

Alla luce di questa metafora, l'attuale situazione delle biblioteche rivela un profondo squilibrio: le biblioteche di oggi non funzionano più come dovrebbero perché non c'è proporzione tra desideri e aspettative dei lettori e servizi realmente offerti, tra universo bibliografico tradizionale e digitale e raccolte delle biblioteche, tra i modelli di approvvigionamento di informazioni seguiti dai lettori e gli strumenti messi a disposizione dalla professione.

La disponibilità sempre maggiore di risorse informative non bibliocentriche ha modificato i desideri e i costumi degli utenti in modo molto più rapido di quanto le biblioteche siano state in grado di modificare e aggiornare i propri strumenti di accesso alle informazioni. A parte qualche eccezione, i dati delle biblioteche sono chiusi in silos ermetici, impenetrabili e invisibili ai motori di ricerca, che sono invece il *Deus ex machina* dei nostri utenti.<sup>2</sup>

---

<sup>1</sup> S.R. RANGANATHAN, *The organization of libraries*, 3. ed., London, Oxford University Press, 1963, § 160.

<sup>2</sup> A torto se si considerano – oltre alle potenzialità – anche i rischi e i limiti posti dai motori di ricerca a un accesso efficace, libero, neutrale e tutelato alle informazioni del web (Maurer et al. 2007).

È possibile ripristinare questo equilibrio? Ci sono le condizioni per superare questa crisi? Come possono le biblioteche tentare di riprendere un ruolo centrale nella risposta ai bisogni informativi dei lettori?

## 2. I dati delle biblioteche

Millioni di dati bibliografici conservati negli OPAC delle biblioteche di tutto il mondo non sono raggiungibili dai motori di ricerca; essi rimangono invisibili alla maggior parte degli utenti (86%) che cercano informazioni sulla rete a partire dai motori di ricerca.

I motivi di questa situazione sono molti, ma si possono ricondurre a una ragione fondamentale: l'automazione delle biblioteche e la nascita del web sono stati due eventi contemporanei, paralleli ma fino a oggi estranei l'uno all'altro.

Al centro della crisi del ruolo delle biblioteche – anche digitali – e dei bibliotecari, c'è la constatazione che i lettori non percepiscono più il catalogo come il punto di partenza delle loro ricerche: i lettori sono e rimangono altrove, *sul web*, e cercano *nel web*. Le nostre biblioteche, anche le biblioteche 2.0, compaiono *sul web* ma non sono *nel web*.

Chi interroga un motore di ricerca per trovare un'opera, ottiene in risposta le pagine web che parlano di quell'opera, l'eventuale scansione digitale in Google Books e i dati dei cataloghi di vendita di libri in commercio e di antiquariato, ma *non* dagli opac, i cataloghi delle biblioteche. Alcuni opac sono visibili, come per esempio WorldCat o la National Library of Australia e, a livello nazionale, le Biblioteche Civiche Torinesi e la Braidense, ma rimangono eccezioni. Paradossalmente i *discovery tools*, l'ultimo ritrovato delle aziende fornitrici di software per il mondo bibliotecario, garantiscono l'esatto contrario di ciò di cui le biblioteche hanno bisogno: gli opac sono aperti verso il web e quando un utente non trova nel catalogo ciò che cerca, viene indirizzato sul web aperto. Inseguiamo il web imitando i motori di ricerca con i *next generation catalogues* e i *discovery tools*, ma questo approccio continua a tenere diviso il mondo dei dati bibliografici dal mondo dei dati del web.

Se i cataloghi non diventano parte del web, e i dati bibliografici non si integrano nel web, i lettori continueranno a non trovarli nei loro primi approcci alla ricerca e, con tutta probabilità, continueranno a ignorarli.

Lo sviluppo del web non ha lasciato indietro solo il mondo delle biblioteche; tutte le applicazioni che hanno sviluppato e prodotto banche dati con formati specifici (archivi anagrafici, geografici, statistici, giuridici ecc.) sono rimaste escluse dalla rete, perché il contenuto di centinaia di migliaia di archivi con strutture e formati di dati diversi tra loro costituisce una selva impenetrabile per i motori di ricerca, che continuano a ignorarla. Il web dei documenti si è già mosso in questo senso, cercando il modo migliore per rendere disponibili nel web i propri dati (Palmisano 2010).

È necessario rendere i dati interoperabili: non tra istituzioni di natura bibliotecaria o documentaria (musei, archivi, editori ecc.), come è stato il principale obiettivo dell'IFLA nei decenni scorsi, ma tra i silos culturali e l'esterno, il web aperto. È necessario registrare i dati bibliografici con modalità che consentano agli *elaboratori* – e soprattutto ai robot dei motori di ricerca – di riconoscere che sono di natura bibliografica (piuttosto che geografica, anagrafica, economica ecc.), di comprendere che tra due dati esiste una relazione e di stabilire quale tipo di relazione sia.

Il tema non può non interessare la comunità bibliotecaria nel suo complesso; essa ha perso molte occasioni perché la preoccupazione maggiore era costituita dai milioni di dati bibliografici già creati e registrati in un formato oggi totalmente superato. Il MARC è di fatto stato abbandonato (Marcum 2011), ma non sono molti gli sforzi a livello internazionale per cercare nuove soluzioni.

Malgrado molti detrattori, il catalogo – inteso come lo strumento di accesso e di gestione delle risorse – rimane il *core business* delle biblioteche: senza i dati non funzionano la ricerca, la consultazione, la circolazione, il reference, le acquisizioni, la conservazione, l'aggiornamento. Senza i dati è tutto inutile e chi non ne è convinto, provi a lavorare un giorno – un solo giorno – senza opac. La dimostrazione più incontrovertibile è il comportamento dei lettori che, non trovando i nostri dati con i motori di ricerca, ignorano semplicemente le biblioteche e *tutti i loro servizi*.

La rigidità imposta dal formato dei dati ha bloccato il processo di evoluzione delle biblioteche innescato dall'automazione, in modo forse banale, ma non del tutto imprevedibile.<sup>3</sup>

Mentre negli anni '70 e '80 del secolo scorso le biblioteche avviarono il loro processo di automazione malgrado la consapevolezza della necessità di un enorme lavoro di riconversione dei dati pregressi, negli anni successivi i bibliotecari hanno rinunciato a godere dei vantaggi dell'evoluzione tecnologica per il timore di non riuscire a conservare i dati del catalogo elettronico.

In modo apparentemente altrettanto paradossale, se il mondo delle biblioteche riuscirà a rimettersi al passo con i tempi sarà grazie alle nuove regole di catalogazione (RDA e REICAT) – che sono coraggiosamente andate oltre il presente (Tillett 2011) e sono in attesa che il mondo delle biblioteche si riallinei e sia in grado di applicarle – e ai lavori di un organismo internazionale che non si occupa di biblioteche: il WWW Consortium, con il suo Library Linked Data Incubator Group.

Per essere utili nel web, i dati bibliografici devono diventare capaci di svolgere nuove funzioni oltre a quelle che hanno svolto finora: essere leggibili, modulari e utilizzabili anche in ambienti non bibliografico-documentari (cioè nel web aperto) e devono rendere esplicito (cioè registrare separatamente) il valore delle relazioni tra i dati.

Le trasformazioni tecnologiche del catalogo e dei suoi metadati sono state sempre determinate dalla necessità di svolgere nuove funzioni. Nel tempo, per realizzare nuove funzioni si è lentamente e profondamente modificata la struttura del catalogo e l'organizzazione dei metadati: a ogni aumento di funzioni è corrisposta l'adozione di nuovi modelli logici, nuovi strumenti e nuovi formati di dati (vedi Appendice A. *Breve storia dei metadati*).

Dal momento che il web richiede nuove funzioni ai dati bibliografici – in modo che i dati bibliografici siano reperibili direttamente dai motori di ricerca – c'è da chiedersi: come cambierà il catalogo? quale sarà la sua struttura? come dovranno essere i nuovi metadati?

### 3. Nuove funzioni per entrare nel web

I nuovi codici di catalogazione derivati da modello di FRBR, come REICAT e RDA, e il rapporto del Library Linked Data Incubator Group, pubblicato nell'ottobre 2011, confermano la necessità di adottare modelli di registrazione e di rappresentazione dei dati molto più articolati di quelli attuali.

Per garantire l'uso effettivo dei dati bibliografici nel web aperto, è necessario che i metadati siano caratterizzati da:

- 1) maggiore granularità<sup>4</sup>, ovvero un insieme maggiore di elementi di dati da riconoscere e da registrare (nella terminologia RDA un nuovo *element set*, per Dublin Core un nuovo *description set* e per il W3C un nuovo *metadata element set*);
- 2) maggiore complessità nella rappresentazione delle entità e delle relazioni, ovvero del *data model*, perché sono necessari più legami per collegare un numero maggiore di dati pertinenti alla stessa rappresentazione; tale complessità potrà essere gestita anche mediante valori prestabiliti (come per esempio nelle Appendici C e D di REICAT e nei *Vocabularies* di RDA)
- 3) impiego di nuovi formati e nuovi strumenti tecnologici, che consentano la corretta interpretazione dei dati, del loro valore e del loro impiego da parte degli elaboratori (ma in ultima istanza da parte dei motori di ricerca), ovvero il ricorso ai Linked Data e a RDF.

---

<sup>3</sup> È da almeno dieci anni che è stato segnalato l'anacronismo del MARC e la necessità di trovare formati alternativi per i dati bibliografici. Si vedano per esempio Roy Tennant, *MARC must die*, "Library Journal", 15 ottobre 2002, <http://www.libraryjournal.com/article/CA250046.html> e David Flanders, 'Applying XML to the Bibliographic Description,' *Cataloging & Classification*, 33 (2) 2001, p. 17-28.

<sup>4</sup> Per *granularità* si intende il livello di risoluzione raggiunto da un sistema di archiviazione nella registrazione logica e fisica dei dati.

REICAT e RDA sono intervenute aumentando il numero dei metadati, elaborando un modello di dati più granulare e rendendo possibile una maggiore complessità nella rappresentazione delle entità bibliografiche, secondo uno scenario tecnologico più nuovo e articolato.

#### 4.1 Lo scenario delle nuove regole

Nello Scenario 1 – che corrisponde sostanzialmente a quello tracciato da REICAT e, in modo più articolato e dettagliato, da RDA – ciò che oggi immaginiamo come un'unica descrizione bibliografica (per esempio una descrizione che si sviluppa dall'area 1 all'area 8 dell'ISBD) deve essere vista come il risultato dell'aggregazione di un insieme di "pacchetti informativi" o, in altri termini (Dublin Core) come un 'insieme descrittivo' (*description set*) che corrisponde alle 8 aree standard della descrizione. In questo scenario, la descrizione bibliografica è l'insieme dei dati sufficienti e necessari a rappresentare la manifestazione, scelti tra gli attributi delle diverse entità coinvolte nella descrizione e assemblati secondo un ordine prestabilito.

In questo modo la rappresentazione di una entità bibliografica risulta essere la sintesi o l'estrazione e la presentazione ordinata di un insieme di dati registrati in archivi diversi sotto forma di attributi di entità (archivio delle opere o delle espressioni o delle manifestazioni, archivio delle persone o degli enti) e selezionati grazie alle relazioni definite tra i diversi archivi.

Le entità che è possibile descrivere sono definite dalle regole catalografiche che si applicano. L'insieme delle entità previste da RDA, per esempio, è definito da appositi registri, che si chiamano *RDA element sets*.<sup>5</sup> In RDA sono previsti 14 elementi per il registro delle *FRBR entities for RDA* (che corrispondono a opera, espressione, manifestazione, item, persona, famiglia, ente, concetto ecc.), 469 elementi per il *RDA Group 1*, 62 elementi per il *RDA Group 2* e 19 elementi per il *RDA Group 3*.

Anche ciascuna relazione assume un valore: per esempio nella tripla

(a) Opera – creata da – Persona

oltre a stabilire il valore assunto da Opera (per esempio, definita dalla tripla >Opera – Ha per titolo – "Lettere familiari"<) e da Persona (per esempio, definita dalla tripla >Persona – Ha per nome – "Francesco Marino"<) è necessario oggi esplicitare quale sia il valore della relazione "creata da"; per esempio, è necessario indicare se il creatore ha un ruolo di autore, attore, regista, fotografo, cantante ecc., cioè i tipi di responsabilità (cfr. per esempio REICAT 14.2.4).

La lettura delle seguenti due stringhe:

(b) MILL, JOHN S.. On liberty; 3d Ed. Boston; 1864, 16mo.

(c) Utilitarismo  
MILL, JOHN S.. Utilitarianism. London; 1863, 8vo.

è sufficiente per garantire la corretta comprensione dei dati da parte di un essere umano; al contrario, le due stringhe sono del tutto incomprensibili al software di un motore di ricerca, che non è in grado di distinguere tra un ruolo di autore, di traduttore o di regista (né di associare i dati alle rispettive entità).

Per questo motivo, i valori delle funzioni svolte (i tipi di responsabilità) devono essere resi espliciti: in RDA sono stati individuati 508 elementi (cioè possibili valori) nelle *RDA Relationships for Works, Expressions, Manifestations, Items*. Per esempio, in RDA per i legami tra opera e opera sono stati previsti i valori: "basato su", "adattamento di", "adattamento radio di", "commento di" ecc.).

È necessario distinguere concettualmente tra *Elements sets* e *Vocabularies*: gli insiemi di elementi indicano le entità e gli attributi rilevanti per quelle entità. Prendiamo per esempio le due triple seguenti:

(a) Opera – creata da – Persona

---

<sup>5</sup> <http://rdvocab.info/>

#### (d) Manifestazione – registrata su – Tipo di Supporto

Nella tripla (a) l'entità persona e l'entità opera possono assumere qualsiasi valore per il loro rispettivi attributi "Nome" e "Titolo", nel senso che chiunque può creare un'opera intitolata "Lettere familiari" e che un'opera può ricevere qualsiasi titolo. Perciò il valore assunto dagli elementi Persona e Opera è aperto, imprevedibile, e il dato può essere assegnato solo a posteriori. Le entità Persona e Opera sono elementi costitutivi dell'*Elements set* di RDA.

Nella tripla (d), i valori dell'elemento *Tipo di supporto (Carrier type)* – uno degli elementi elencati nel registro RDA Group 1 Element – sono prefissati in un elenco chiuso ed esaustivo (un *RDA vocabulary*), all'interno del quale si possono scegliere solo tipi di supporto esistenti ed esplicitamente previsti. Per esempio, e banalizzando, gli effetti di questa differenza potrebbero essere descritti così: se ci trovassimo a catalogare una determinata risorsa con una interfaccia utente, nell'inserire un dato relativo a un elemento preso dagli RDA element set avremmo a disposizione un campo libero, mentre nell'inserire un valore preso da un RDA vocabulary avremmo a disposizione solo un menù a tendina per scegliere tra valori prefissati.

I vocabolari chiusi previsti da RDA sono 70, tra i quali per esempio Tipo di contenuto (*Content Type*), Tipo di media (*Media Type*) e Modalità di emissione (*Mode of Issuance*).

Questa distinzione è vantaggiosa, perché nei vocabolari chiusi è possibile mantenere un maggiore controllo sulla coerenza e sulla affidabilità dei dati; lo svantaggio dei vocabolari chiusi è la necessità di aggiornamento costante; per esempio, l'Appendice D delle REICAT – che possiamo considerare un vocabolario chiuso relativo alle Designazioni specifiche del materiale – non ha previsto tra i tipi di supporto la chiavetta USB e ciò la rende già obsoleta.

## 4.2 Lo scenario dei library linked data

Nello scenario tracciato dai library linked data (LLD), le entità tra le quali si costituiscono le relazioni non sono sempre e soltanto le entità definite nel modello FRBR e facenti parte dell'universo bibliografico tradizionale delle biblioteche; oltre alla rappresentazione di tutte le relazioni possibili tra le entità di FRBR, e per esempio tra un'opera e le opere da essa derivate, lo scenario dei LLD prevede in più che si sviluppino collegamenti tra i dati prodotti all'interno della biblioteca e informazioni prodotte all'esterno e indipendentemente, da agenzie di natura non bibliografica.

L'idea di fondo di questo scenario è che non esiste più la distinzione – per quanto riguarda la forma e il luogo di registrazione – tra "dati bibliografici" e "altri dati" di natura diversa; nel web esistono solo dati: condivisibili, modulari e riutilizzabili.

La prima conseguenza di questo ampliamento di orizzonte è che i dati bibliografici devono essere riconoscibili fra molti altri tipi di dati non bibliografici: tecnicamente è possibile se la codifica dei dati diventa più complessa.<sup>6</sup>

Una seconda conseguenza, legata alla necessità che i dati prodotti siano condivisibili e riutilizzabili, è la produzione di molti archivi diversi e utilizzabili da parte di chiunque.

Il concetto catalografico che rappresenta meglio questo tipo di dati e questo nuovo modello di dati è l'archivio di autorità. Per esempio: quali potrebbero essere i vantaggi se, nel lavoro di catalogazione di tutti i giorni, anziché creare nuove intestazioni o collegare intestazioni già presenti – create in base alle REICAT (ICCU 2009) – da un archivio di autorità interno al catalogo, si potesse legare una descrizione direttamente alla forma di autorità presente nel VIAF, che comprende la forma REICAT, la forma RICA, la forma della Library of Congress, la forma RDA, la forma delle RAK ecc.?

---

<sup>6</sup> Un esempio semplice per cogliere questo aumento di complessità è quello offerto dal numero identificativo ISBN. Fino a quando la sua funzione si svolgeva nell'ambito dei dati bibliografici, era sufficiente una codifica di 10 cifre. Quando i libri hanno iniziato a essere venduti al di fuori del circuito librario (per esempio al supermercato), è diventata indispensabile una codifica di 13 cifre, con un prefisso di 3 cifre (978) per indicare che – tra tutti i tipi di prodotti del supermercato – si tratta di un libro.

Le conseguenze più immediate che saltano agli occhi sono sul lato biblioteche e sul lato utente: la gestione e la crescita dell'archivio VIAF potrebbe essere il risultato del lavoro cooperativo di catalogatori di tutto il mondo: ciascuno potrebbe utilizzare i dati registrati nell'archivio di autorità per identificare meglio l'entità relativa e per aggiungere le proprie informazioni (la forma secondo le regole nazionali o locali).

Dal lato utente, un utente di qualsiasi parte del mondo e di qualsiasi madrelingua, potrebbe cercare il proprio autore nella forma preferita o conosciuta *da lui*, e trovare *tutte le registrazioni* relative (non solo alle manifestazioni, ma a tutte le entità del Gruppo 1 di FRBR e degli altri Gruppi di entità di FRBR, FRAR e FR SAR). Non è forse il catalogo descritto da Cutter? Non è il desiderio di tutti i bibliotecari e soprattutto di tutti gli utenti?

Il modello dell'archivio di autorità applicato all'entità persona si estende, nello scenario dei LLD a quanti più archivi di autorità possibile: per esempio, un archivio degli editori, un archivio dei luoghi, un archivio dei formati, un archivio dei supporti, un archivio delle opere, un archivio delle espressioni, un archivio delle copie, un archivio dei soggetti (per esempio creati nella forma prevista da specifiche e diverse norme e collegati in una struttura simile a quella attuale del VIAF per le intestazioni d'autore), un archivio delle classificazioni ecc.

Per creare un legame tra una qualsiasi entità e un luogo (per esempio "Feletto Umberto"), è necessario sapere dove si trova e poterlo identificare. Invece di svolgere le necessarie ricerche e scoprire che si trova in provincia di Udine, è più facile e veloce individuarlo in un archivio d'autorità geografico, per esempio GeoNames (<http://www.geonames.org/>), dove la località è già identificata.

Il lavoro di catalogazione delle località geografiche è già stato fatto: non importa se è stato fatto da catalogatori o a fini bibliografici: i dati esistono già e se hanno le garanzie e le caratteristiche necessarie, possono e devono essere riutilizzati a fini bibliografici. Questo approccio in teoria non ha un limite: potrebbero essere utilizzate come archivi di autorità anche anagrafi di natura completamente diversa: anagrafi cittadine, banche dati delle società proprietarie dei diritti d'autore, sistemi informativi territoriali, anagrafi bancarie, archivi di istituzioni pubbliche, banche dati di beni museali, archivistici, artistici ecc.

Quali altri espedienti tecnici sono necessari per realizzare questo obiettivo?

## 5. Il web semantico e i linked data

Alla fine di ottobre 2011 è stato pubblicato il rapporto finale del gruppo di lavoro creato dal World Wide Web Consortium per lo studio di fattibilità dei library linked data: il W3C Library Linked Data Incubator Group (LLD XG). Questo rapporto è forse destinato ad avere sul mondo delle biblioteche un impatto altrettanto ampio, se non maggiore, di quanto non abbia avuto fino ad oggi il rapporto FRBR.

Tra i due rapporti esistono alcune importanti differenze, che è utile evidenziare:

1) FRBR è stato prodotto nell'ambito delle biblioteche, in seno all'IFLA, mentre i lavori del LLD XG sono stati condotti in un ambiente culturale estraneo alle biblioteche, anche se con la consulenza di bibliotecari esperti;<sup>7</sup>

2) la seconda differenza importante è il diverso punto di partenza dell'analisi: nel caso di FRBR il lavoro è iniziato dall'analisi degli standard di descrizione bibliografica esistenti (in particolare l'ISBD) allo scopo di ridurre i costi di catalogazione; nel caso del LLD XG Final Report il punto di avvio dello studio è stato il web

---

<sup>7</sup> Fanno parte dell'Incubator Group (XG): Thomas Baker, Dublin Core Metadata Initiative, US (W3C Invited Expert), Emmanuelle Bermès, Centre Pompidou, France (W3C Invited Expert), Karen Coyle, Consultant, US (W3C Invited Expert), Gordon Dunsire, Consultant, UK (W3C Invited Expert), Antoine Isaac, Europeana and Vrije Universiteit Amsterdam, Netherlands, Peter Murray, LYRIS, US (W3C Invited Expert), Michael Panzer, OCLC Online Computer Library Center, Inc., US, Jodi Schneider, DERI Galway at the National University of Ireland, Galway, Ireland, Ross Singer, Talis Group Ltd, UK, Ed Summers, Library of Congress, US, William Waites, University of Edinburgh (School of Informatics), UK, Jeff Young, OCLC Online Computer Library Center, Inc., US, Marcia Zeng, Kent State University, US (W3C Invited Expert).

semantico e l'analisi dei requisiti necessari ai dati bibliografici per diventare dati interconnessi (Linked data), cioè riutilizzabili, e aperti (Open Data), cioè ad accesso libero.<sup>8</sup>

3) nel rapporto del LLD XG si nota una quasi totale assenza di riferimenti impliciti ed espliciti relativi a principi, standard e codici di catalogazione; i riferimenti che sono stati inseriti in realtà servono prevalentemente per evidenziare i problemi e le difficoltà dei vari strumenti. Il punto è che la realizzazione di dati interconnessi (linked data) relativi alle biblioteche è indipendente dal tipo di norme che si adottano per produrre i dati: per esempio, la stesura di un testo prevede norme grammaticali, stilistiche, redazionali, citazionali ecc. che sono completamente indipendenti dal formato di registrazione del programma di elaborazione testi.

L'espressione *Linked Data* si usa per indicare le modalità migliori per pubblicare, condividere e collegare singoli dati, informazioni e conoscenze sul web semantico usando gli URI (Uniform Resource Identifier) e il formato RDF (Resource Description Framework).<sup>9</sup>

L'URI è un identificatore che identifica una risorsa nel web: è l'equivalente dell'ISBN nel mondo dell'editoria, ma è molto più efficace in termini di identificazione (Kumar 2010). Anche se l'URI non è privo di difetti (Bozzarelli and Cassella 2011), il concetto di uno strumento di identificazione univoca è ben conosciuto nel mondo delle biblioteche e non ha bisogno di particolari approfondimenti.

Con l'espressione Library Linked Data quindi si indicano dati pubblicati come linked data e il cui contenuto abbia rilevanza bibliografica.

## 5.1 Resource Description Framework

La sigla RDF indica il formato dei dati per i dati interconnessi (linked data): ogni dato che rispetta i requisiti dei dati interconnessi deve essere in formato RDF. Il formato RDF facilita l'uso di dati anche se i modelli logici con i quali sono prodotti sono diversi e ha la caratteristica specifica di permettere l'evoluzione dei modelli logici nel tempo senza la necessità di modificare gli strumenti necessari a utilizzare i dati registrati.

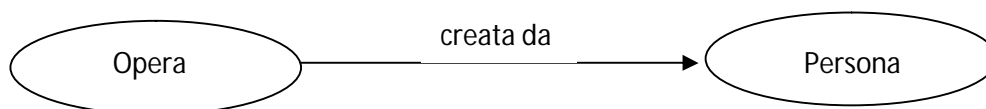
RDF è un'estensione della struttura di connessioni del web e consente di utilizzare gli URI per identificare il *tipo di relazioni* tra oggetti (oltre agli oggetti stessi).

L'essenza dell'RDF è descrivere le relazioni tra gli oggetti sotto forma di triple, cioè rappresentare concetti molto più semplici di quanto non siano il contenuto di una pagina web o di una descrizione bibliografica.

Il formato HTML consente di creare pagine (documenti) adatte a essere visualizzate sul web e di connetterle tra loro (tipicamente tramite uno o più link). In termini di metadati il documento è visto come un blocco indivisibile e potrebbe essere paragonato, in termini funzionali, a una descrizione ISBD completa dall'area 0 all'area 8.

Per creare un web basato sui dati invece che sui documenti, è necessario un formato che consenta di creare legami tra semplici dati, invece che tra interi documenti: RDF lo consente, perché crea relazioni tra oggetti, tipicamente sotto forma di triple, come quelle degli esempi (a) e (b) del par. 4, rappresentate mediante un *grafo*. La visualizzazione a grafo è il modello mentale più semplice possibile per l'RDF. Nella rappresentazione mediante un grafo la tripla (a) diventerebbe (figura 0):

Figura 0



<sup>8</sup> Gli aggettivi *linked* e *open* si riferiscono rispettivamente a una caratteristica tecnica e a una condizione giuridica dei dati.

<sup>9</sup> <http://linkeddata.org/>

nel quale gli oggetti (o le entità, in termini FRBR) opera e persona sono chiamati *nodi* e la relazione "creata da" è chiamata *arco*.

L'essenza di una tripla è una dichiarazione, cioè un'affermazione nella quale si stabilisce che una proprietà di un certo soggetto assume un determinato valore. Per esempio: il titolo (proprietà, o *attributo* in FRBR) di un libro (soggetto) è "La guerra gallica".

Una seconda caratteristica fondamentale del formato RDF è che tutti i nodi e tutte le relazioni sono identificate con un proprio URI; sicché in RDF l'identità di due cose o di due relazioni non si stabilisce sulla base dell'identità del valore assunto ("La guerra gallica", "Rossi, Mario" o "Svevo, Italo") ma dall'identità dell'URI. Se nel linguaggio naturale "Rossi, Mario" è uguale a "Rossi, Mario", per un elaboratore le due stringhe registrate sul web saranno uguali solo se saranno identificate con lo stesso URI (altrimenti sono due Mario Rossi diversi). Viceversa, "Schmidt, Ettore" e "Svevo, Italo" sono stringhe equivalenti per un elaboratore se, avendo due diversi URI, esiste una relazione di equivalenza tra i due soggetti.

La rete di relazioni tra oggetti univocamente identificati e descritti tramite relazioni così articolate, consente di evitare la costruzione di complesse matrici di conversione tra formati, che richiedono anni di lavoro e rischiano di non essere del tutto soddisfacenti. I rinvii tra dati e tra metadati sono registrate direttamente nei dati (e non esternamente).

L'uso di un modello logico così semplice consente di mescolare, visualizzare e condividere dati strutturati e semi strutturati attraverso diverse applicazioni.

Mediante una concatenazione di triple è possibile esprimere dati molto complessi, ovvero descrivere, rappresentare entità complesse come quelle bibliografiche; nel web semantico e nel mondo dei dati interconnessi non bisogna pensare più in termini di registrazioni (bibliografiche o di autorità), ma di triple espresse mediante l'RDF. Per esempio, posso definire un oggetto mediante molte relazioni con le sue proprietà (figura 1) e mediante relazioni con altri oggetti, definiti a loro volta mediante relazioni proprie (figura 2) (tratte da Davis and Heath 2009).

Figura 1

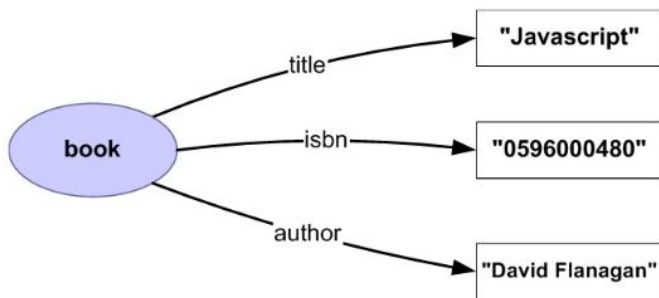
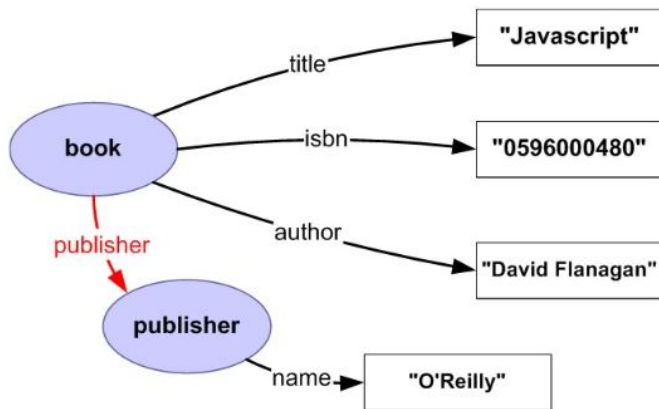


Figura 2





Il vantaggio di questo strumento tecnologico diventa più evidente quando si esamina come possono essere assemblati in modo semplice e rapido i dati che interessano un utente che cerca informazioni. Per ipotesi, due agenzie informative diverse hanno fornito dati diversi su uno stesso oggetto (per esempio un libro), secondo i grafi rappresentati in figura 3 e figura 4

Figura 3

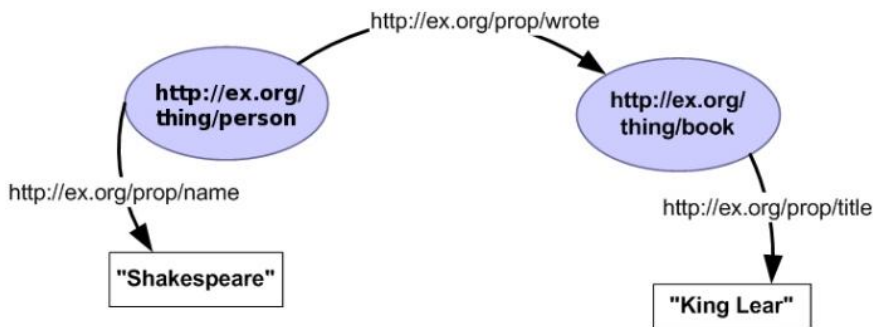
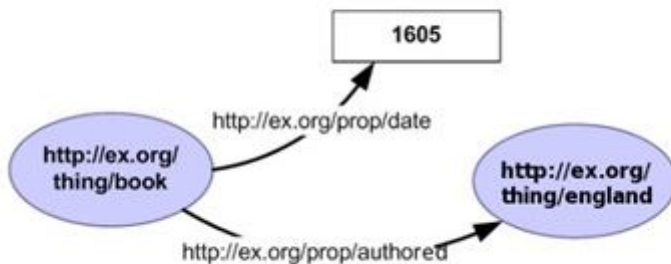


Figura 4



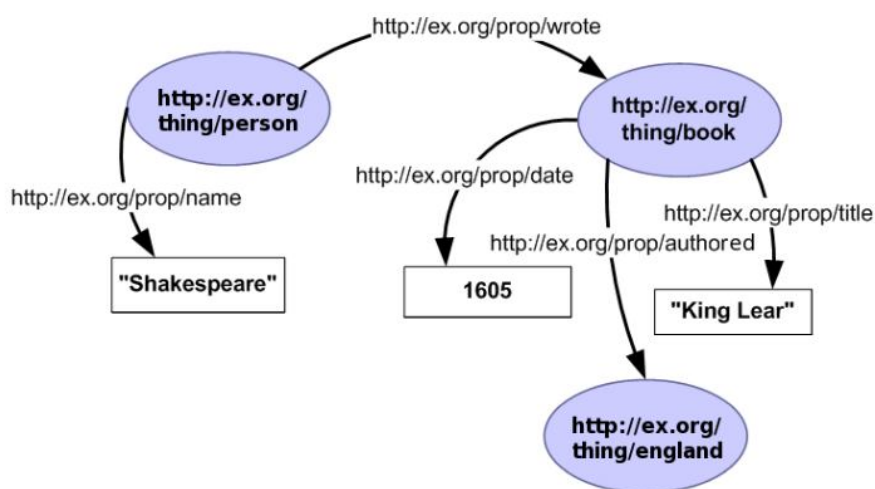
Nelle due figure, l'oggetto libro identificato tramite un URI (<http://ex.org/thing/book>), viene descritto con relazioni diverse dall'agenzia A: in figura 3 si dichiara che l'oggetto "libro" è stato scritto da una persona

(che a sua volta, mediante un'altra tripla, è definita come "Shakespeare") e che il medesimo oggetto "ha come titolo" (una relazione identificata con un apposito URI), il valore testuale "Re Lear".

Nella figura 4 vengono rappresentate altre due relazioni del medesimo oggetto, definite e messe a disposizione sul web da un'altra agenzia informativa B. In questo caso, al medesimo oggetto sono associati una data, che ha come valore numerico "1605", e un luogo di produzione, che è un oggetto identificato l'URI: <http://ex.org/thing/england>.

Nel momento in cui un motore di ricerca effettua una ricerca nel web semantico, individua e recupera tutte le relazioni esistenti con uno o più dati che corrispondono a un certo URI, proprio come oggi cerca tutte le ricorrenze di un testo nelle pagine del web. Nel caso delle figure 3 e 4, sarà in grado di recuperare sia le triple definite nella figura 3 dall'agenzia A che le triple rappresentate in figura 4 e create dall'agenzia B. Per l'elaboratore la differenza nella provenienza delle informazioni non sarà un problema e il grafo finale, che rappresenta il risultato completo della ricerca, potrà corrispondere a quello della figura 5.

Figura 5



## 6. Il Library Linked Data Incubator Group

Nel maggio 2011 il W3C ha incaricato un gruppo di lavoro (Incubator Group, abbreviato in XG) di "favorire l'aumento della interoperabilità globale dei dati bibliografici sul web, facendo collaborare persone coinvolte nelle attività del web semantico – con l'accento sui linked data – nella comunità bibliotecaria e oltre, lavorando a partire da iniziative in corso e identificando futuri percorsi di collaborazione" (Library Linked Data Incubator Group 2011).

Nel corso dei suoi lavori, l'XG ha analizzato come si possono utilizzare gli standard del web semantico e i principi dei linked data per rendere più visibili e riusabili – al di fuori del contesto di produzione originario, nel web più ampio – le notevoli risorse create, curate e prodotte in ambito bibliotecario (come dati bibliografici, dati di autorità e schemi concettuali).

Il punto di partenza del lavoro è stata l'analisi delle iniziative di applicazione di linked data in atto, includendo progetti di ogni scala, da locale a nazionale e raccogliendoli in una appendice. Questi casi sono stati la base dell'analisi dei vantaggi dei LLD, delle iniziative esistenti sui LLD, e dei problemi connessi alla gestione dei diritti dei LLD.

Il rapporto offre anche una sintesi dei risultati di un'indagine sulle attuali tecnologie per i LD e un elenco delle risorse disponibili al momento per i LLD.

Il rapporto finale non è l'unico risultato dell'XG, che ha pubblicato anche altri due documenti: la citata appendice – intitolata *Library Linked Data Incubator Group: Use Cases* – che descrive le applicazioni bibliografiche e mostra i vantaggi dell'adozione degli standard del web semantico e dei principi dei dati interconnessi nelle biblioteche pubbliche e il documento *Datasets, Value vocabularies, and Metadata Element Sets* che fornisce il quadro delle principali risorse disponibili per creare oggi dati interconnessi.

Il rapporto finale del LLD XG si articola in quattro parti: nella prima si definisce l'obiettivo del gruppo di lavoro, nella seconda si descrivono i possibili vantaggi derivanti della pubblicazione dei dati bibliografici come dati interconnessi, nella terza si delinea l'attuale situazione della gestione dei dati bibliografici e si evidenziano i problemi dei dati bibliografici tradizionali e nella quarta si formulano raccomandazioni immediatamente applicabili per i responsabili di biblioteche, per creatori di dati e di sistema, bibliotecari e archivisti, e responsabili di biblioteche.<sup>10</sup>

Il documento è corredato da tre Appendici: l'Appendice A è un elenco di risorse LLD esistenti, l'Appendice B è dedicata alle soluzioni tecnologiche richieste per la realizzazione dei LLD (URI, RDF, SKOS ecc.) e l'Appendice C tratta dell'Allineamento semantico, ovvero la creazione di relazioni di equivalenza tra oggetti appartenenti a diversi vocabolari chiusi (*value vocabularies*), insiemi di metadati (*metadata element sets*) o insiemi di dati (*datasets*).

## 7. I vantaggi dei LLD

L'approccio Linked data offre numerosi vantaggi rispetto all'attuale gestione dei dati bibliografici: tra gli aspetti più interessanti ci sono la possibilità di produrre dati condivisibili, modulari e riutilizzabili e il superamento dei limiti posti dalle barriere linguistiche (un obiettivo di ISBD) con il ricorso a URI, cioè a identificatori non linguistici.

La condivisione dei dati è un patrimonio acquisito in ambito bibliotecario, dalla realizzazione dei cataloghi cooperativi in poi. La modularità e il riutilizzo dei dati acquistano un significato nuovo e diverso se visti nel web, e non solo in ambito bibliografico. Modularità e riutilizzo significano uso effettivo e immediata applicazione di dati prodotti da altre agenzie (per esempio i dati di pubblicazione forniti dall'editore, come l'ISBN, la copertina, la presentazione editoriale ecc.) e, viceversa, l'impiego diretto da parte di altri dei dati bibliografici di alta qualità prodotti in ambito bibliotecario.

La condivisione dei dati da parte delle biblioteche e il collegamento a dati affidabili prodotti da altre agenzie può aumentare il valore dei dati ben oltre la somma dei dati delle singole biblioteche. Le biblioteche e le istituzioni culturali in generale si trovano in una posizione di vantaggio per pubblicare nel web metadati autorevoli per le risorse di rilevanza culturale duratura.

L'approccio linked data consente di promuovere la creazione dei dati in forma più semplice e più granulare rispetto alla registrazione bibliografica completa: i dati minimi relativi a una entità sono costituiti da una tripla e la registrazione bibliografica completa è costruita sulla base della sintesi di un consistente numero di triple. Per esprimere il concetto "King Lear / Shakespeare", relativamente banale nel linguaggio ISBD, sono necessarie almeno tre triple (vedi figura 5). Queste triple però possono essere state create da agenzie diverse, indipendentemente l'una dall'altra. Un'agenzia bibliografica che creasse la nuova tripla

(e) Re Lear --- è traduzione di --- King Lear

disporrebbe immediatamente delle altre tre triple della figura, perché questa tripla include un'entità (King Lear) identificata univocamente e caratterizzata dalle altre triple.

In un sistema di linked data quindi non esiste in termini di dati un contributo così piccolo da non essere utilizzabile o vantaggioso. In questo contesto, la diffusione dei dati bibliografici di qualità su nomi e soggetti prodotti e posseduti dalle biblioteche è di fondamentale importanza per il web e la creazione delle relazioni tra dati bibliografici di autorità e altri dati come Wikipedia, GeoNames, MusicBrainz ecc. (Library Linked Data Incubator Group 2011, § 2.1). Non è necessario ricreare i dati bibliografici, né ricreare il web in modo

---

<sup>10</sup> <http://www.w3.org/blog/SW/2011/10/27/w3c-library-linked-data-xg-final-report-published/>

diverso: è sufficiente migliorare i dati esistenti fornendoli di metadati adeguati per il web (Library Linked Data Incubator Group 2011, § 2.1).

L'XG sottolinea che qualunque progetto sviluppato per adeguare i dati bibliografici esistenti al formato richiesto dal web deve anche prevedere la pubblicazione sotto forma di linked data di una descrizione dettagliata del modello utilizzato per produrre i dati; ciò consente di condividere i dati relativi alle sperimentazioni che divengono linked data, cioè modulari e riutilizzabili da parte di altre agenzie bibliografiche.

Secondo il rapporto finale del LLD XG, l'approccio alla creazione dei Library linked data non può essere di tipo tradizionale, ovvero prevedere l'impegno di grandi istituzioni bibliotecarie che investono ingenti somme per produrre archivi di tipo nuovo (o convertire in massa i dati già esistenti): questo approccio è troppo costoso ed esiste il rischio reale che la mancanza di finanziamenti possa bloccare la strada verso la creazione del web dei dati bibliografici. Per questo motivo l'approccio deve essere decentrato e il carico di lavoro distribuito su molte agenzie che creano dati.

Gli attuali programmi di gestione dei dati bibliografici (ILS) usano soluzioni tecnologiche valide solo all'interno del mondo delle biblioteche: le biblioteche sono legate ai produttori di questi software di nicchia e non sono in grado di avvantaggiarsi delle eventuali soluzioni più economiche e più avanzate sviluppate nell'ambiente web. Anche i produttori di software sono costretti a lavorare su formati e standard in uso solo nel settore delle biblioteche (MARC, Z39.50, ISO 2709 ecc.) (§ 2.4).

Secondo il rapporto del LLD XG "i formati tradizionali dei dati bibliografici mescolano il significato dei dati e la codifica strutturata di essi in un blocco unico. L'inseparabilità tra il significato e la codifica nei dati tradizionali è la causa di una minore flessibilità nel trarre vantaggio dagli investimenti fatti sui dati. Fin dall'introduzione dei formati MARC negli anni 60 del secolo scorso, i dati digitali delle biblioteche sono stati gestiti prevalentemente sotto forma di 'registrazioni' che sono insieme definiti di informazioni registrate in file con una struttura rigidamente definita" (Library Linked Data Incubator Group 2011, § 4.2.2).

Il limite all'evoluzione delle tecnologie imposto a suo tempo dal MARC ora si è esteso agli ILS creati su misura per le biblioteche, che spesso sono concettualmente e tecnicamente obsoleti rispetto alle soluzioni tecniche in uso nel mondo del web (è il caso per esempio dei *Next Generation Catalogs* e dei *Discovery Tools* citati in precedenza). Secondo l'XG, lo spostamento dei dati verso un ambiente Linked data consentirebbe alle biblioteche di "avvalersi e interagire con un bacino di sviluppatori di software più ampio." (Library Linked Data Incubator Group 2011, § 2.2)

Per cogliere i potenziali vantaggi offerti dal sistema dei linked data in termini di catalogazione, si può paragonare il sistema alla situazione nella quale si troverebbe oggi un catalogatore che, per fare catalogazione derivata, avesse a disposizione un unico enorme catalogo collettivo costituito con il contributo di tutte le maggiori biblioteche del mondo. L'esempio è per difetto perché qualunque dato creato sotto forma di tripla e associato a un oggetto mediante un URI sarà potenzialmente utilizzabile, in modo trasparente, dalle biblioteche e dai suoi utenti. Il concetto fondamentale dei linked data è che non esistono più 'silos' di dati accessibili solo dall'apposita pagina di ricerca: nei sistemi linked data i dati sono disponibili direttamente per i motori di ricerca.

L'approccio linked data offre un considerevole vantaggio anche rispetto all'obsolescenza dei formati e dei metadati: nella creazione dei LLD è possibile mantenere distinto il valore semantico di un dato dalla sua struttura (sintassi o formato). Ciò implica che i linked data conservano pienamente il loro significato anche quando cambiano formato (Library Linked Data Incubator Group 2011, § 2.3); dal punto di vista pratico i linked data consentono di conservare correttamente le relazioni attualmente esistenti tra una descrizione e gli archivi di autorità e quelle che esisteranno tra le entità descritte e gli oggetti del web aperto.

Oggi le risorse di interesse bibliografico disponibili sotto forma di linked data non sono molte; si tratta soprattutto di insiemi di elementi di metadati, come DCMI e FRBR, e di vocabolari controllati come LCSH e DDC. Tra gli archivi bibliografici spiccano la pionieristica pubblicazione della British National Bibliography e i progetti delle biblioteche nazionali di Svezia, Ungheria, Germania, Francia, della Library of Congress, della British Library, della FAO e di OCLC (Library Linked Data Incubator Group 2011, § 3.2.2).

Il processo di pubblicazione dei linked data promette di ridurre notevolmente la ridondanza e gli sprechi nella duplicazione dei dati, ma ha già evidenziato alcuni punti critici: il livello relativamente basso di manutenzione dei vocabolari, lo scarso coordinamento tra i creatori dei vocabolari e la mancanza di

strumenti consolidati per la pubblicazione a basso costo delle enormi quantità di link semantici necessarie per i linked data (Library Linked Data Incubator Group 2011, § 3.2.3).

Come nel mondo dell'open access, non mancano problemi di diritti sui dati. Le banche dati bibliografiche prodotte dagli editori internazionali non saranno evidentemente rese accessibili a titolo gratuito. Lo stesso limite si pone per esempio con la pubblicazione di dati prodotti grazie alla catalogazione derivata: molti degli attuali cataloghi collettivi, realizzati su scala locale o nazionale, sono costituiti di dati derivati, modificati e migliorati dai catalogatori per i propri utenti e stabilire un diritto di proprietà o di uso è difficile (Library Linked Data Incubator Group 2011, § 3.3.1).

## 8. Le raccomandazioni del LLD XG

Il rapporto dell'Incubator Group si conclude con alcune raccomandazioni indirizzate ai responsabili delle biblioteche, agli enti di standardizzazione, ai creatori di sistemi e di dati e ai bibliotecari.

L'indicazione generale del rapporto riguarda il ruolo futuro delle biblioteche nel web delle informazioni: si auspica che le biblioteche integrino i propri dati con le altre risorse del web, garantendosi una maggiore visibilità e portando i loro servizi a chi cerca le informazioni. Le biblioteche, inserendosi nel web, hanno la possibilità di riassumere la leadership che si fonda nelle loro attività tradizionali: la gestione delle risorse per l'uso e per la conservazione, la descrizione delle risorse sulla base di regole condivise e la capacità di rispondere ai bisogni di chi cerca informazioni (Library Linked Data Incubator Group 2011, § 4).

Il primo passo da fare, secondo le raccomandazioni del LLD XG, è individuare progetti di LD a basso costo e ad alta priorità: i linked data, per loro natura, facilitano un approccio per gradi nella pubblicazione dei dati sul web, mentre il tentativo di pubblicare direttamente la totalità dei dati bibliografici nel loro complesso risulterebbe probabilmente fallimentare.

Nel web semantico è possibile creare dati in formato RDF, ma è noto che gli SKOS (Simple Knowledge Organization System), gli standard attualmente usati per pubblicare i sistemi di organizzazione della conoscenza, non sono in grado di rappresentare tutte le relazioni necessarie. In questo campo la ricerca è aperta e il mondo delle biblioteche ha molto da offrire, purché ciò avvenga in collaborazione e in accordo con la comunità del web semantico e con modalità che consentano di riutilizzare le soluzioni trovate anche al di fuori della comunità bibliotecaria (Library Linked Data Incubator Group 2011, § 4.2.1).

Anche il passaggio dagli attuali standard descrittivi ai linked data richiede la conoscenza dei nuovi principi per la progettazione dei dati ed è necessaria la massima diffusione delle buone pratiche e delle indicazioni per favorire i progetti di costruzione di ontologie e vocabolari strutturati per i dati bibliografici. C'è enorme bisogno di indicazioni per la progettazione di strumenti di descrizione dedicati ai dati bibliografici, a partire dai molti studi già disponibili per la progettazione dei linked data, per esempio *Linked Data: Evolving the Web into a Global Data Space* e da *Linked Data Patterns* (Heath and Bizer 2011; Dodds and Davis 2011).

Un ruolo fondamentale che spetta alle biblioteche nella creazione del web dei dati è la creazione degli URI, gli identificatori per il web, delle specifiche risorse bibliografiche e degli standard. Dal momento che qualsiasi oggetto è identificato nel web tramite un URI, è indispensabile dotare di questo identificatore tutti gli oggetti di interesse bibliografico e tutti gli standard di natura bibliotecaria. È un impegno particolarmente urgente perché la comunità del web si muove in modo molto rapido ed esiste il rischio reale che gli sviluppatori di software e di standard, per non subire ritardi nel loro lavoro, assegnino presto (e male) gli identificatori sostituendosi ai proprietari dei dati. Per esempio, esiste la possibilità che un programmatore abbia bisogno di un dato relativo a un libro schedato su SBN; anziché attendere che SBN renda esplicito il modo per creare l'URI relativo a quel documento, è probabile che adotti un sistema proprio per costruire l'URI e lo utilizzi immediatamente; per la natura dei linked data, è facile che tale sistema si diffonda e diventi uno standard di fatto. A questo punto, l'eventuale documento ufficiale prodotto in seguito dall'ICCU – certamente migliore sotto il profilo della qualità e della profondità dell'analisi – sarebbe verosimilmente ignorato, perché 'inutile' in quanto tardivo. Da ciò nasce l'urgenza di creare URI per le risorse e per gli standard bibliografici. Nell'esempio citato, è auspicabile che l'ICCU stabilisca una propria politica per il namespace di SBN: il termine namespace (spazio dei nomi) indica uno spazio, individuato da un nome (per esempio: *namespace <sbns>*), nel quale i nomi utilizzati per identificare

un'entità sono univoci; l'entità autore può ricorrere una sola volta nel namespace SBN, mentre ricorre molte volte nel web. Il namespace SBN garantisce che "autore" sia usato e interpretato in modo corretto all'*interno* e all'*esterno* del dominio SBN.

Le politiche per la creazione di namespace favoriscono "un approccio univoco, coerente e stabile, che migliora l'efficienza e l'efficacia e garantisce la qualità agli utenti degli URI e dei namespace".

Una politica di creazione dei namespace deve prevedere (§ 4.3.3):

- schemi utilizzati per creare gli URI, possibilmente sulla base di linee guida sulle migliori pratiche;
- impegno istituzionale per garantire la persistenza degli URI;
- controllo della versione di un vocabolario e dei suoi termini;
- impiego di URI basati sull'HTTP, che ricorrono all'Hypertext Transfer Protocol utilizzato universalmente dai browser web, e il rinvio a qualsiasi pagina web o rappresentazione leggibile da una macchina che documenti il significato degli URI;
- estensibilità del vocabolario da parte di altre organizzazioni;
- traduzione delle etichette e delle altre annotazioni in altre lingue.

Tra gli obiettivi prioritari, ci dovrebbe essere anche la creazione di collegamenti tra i dati bibliografici e i dati del web, con la descrizione dei dati bibliografici nei termini dei linked data – ovvero proprietà, classi e istanze (dove la proprietà corrisponde all'attributo, la classe all'entità e l'istanza all'item) – che hanno precise corrispondenze con quelli usati nel più ampio spazio dei linked data.

Il collegamento tra i dati bibliografici e i linked data può avvenire in due modi: utilizzando direttamente i vocabolari esistenti nel web (come nel caso si usi GeoNames per indicare i luoghi di pubblicazione) o creando un collegamento esplicito (o allineamento) tra l'elemento "luogo di pubblicazione" dei dati bibliografici e l'elemento "luogo" in GeoNames.

Un ruolo fondamentale della comunità bibliotecaria e archivistica è garantire anche in futuro l'accesso agli insiemi di metadati (element sets) e ai vocabolari chiusi, indispensabili per la comprensione e l'impiego dei linked data.

Il rapporto del LLD XG riconosce e assegna alle biblioteche il ruolo del controllo di qualità, coerenza e affidabilità sui linked data; quelli prodotti fino a oggi sono il risultato di conversioni singole e create ad hoc e non sono in genere sottoposti a controlli e a operazioni di aggiornamento e di verifica. In questo le biblioteche possono svolgere un ruolo unico e riconosciuto da tutta la comunità del web (Library Linked Data Incubator Group 2011, § 4.4.2).

## 9. Domande aperte

Se si tenta di immaginare l'infinita catena di dati collegati che possono risultare dal tentativo di rappresentare una porzione anche piccola di universo bibliografico, la prima questione che si pone è come si possa costruire una 'scheda', cioè come sia possibile resecare l'insieme dei dati rilevanti e necessari per descrivere una data entità bibliografica (una manifestazione, o un'espressione o un'opera, per usare termini di FRBR). Una risposta d'impulso suggerisce che gli standard descrittivi (come ISBD) saranno ancora più necessari, anche se non saranno necessariamente standard di natura o origine strettamente bibliografici (come risulta dalla posizione che RDA assegna all'ISBD; Bianchini e Guerrini 2009).

Una 'scheda' sarà l'equivalente di un testo (di una sequenza di informazioni e dati corredati di un significato non ambiguo e associati secondo precise regole sintattiche) creato a partire dal web dei dati.

Una seconda questione riguarda il problema di come si potrà garantire la qualità dei dati in un ambiente aperto? In altre parole, quale diventerà la funzione dell'authority control nel web dei dati? È plausibile che in futuro una parte dei dati che costituiranno le 'schede' saranno estranei al 'catalogo' prodotto *in house* da una biblioteca o da un sistema bibliotecario. Nella logica della distribuzione del lavoro di costruzione dei dati collegati, parte dei dati *di interesse bibliografico* non avranno più *origine bibliografica*, ma saranno

prodotti altrove: non solo da agenzie del ciclo del libro, come editori, librai ecc. ma anche da agenzie con obiettivi completamente diversi, come per esempio i dati geografici o i dati di identificazione degli organi amministrativi di uno stato. Quale ruolo avranno le biblioteche in questo contesto? Probabilmente avranno il compito di individuare, tra le molte fonti di dati disponibili, quelle più autorevoli: è una competenza che i bibliotecari hanno da sempre, anche se di norma è applicata nel servizio delle acquisizioni o nel servizio di reference per aiutare l'utente a reperire la risorsa che gli è più congeniale. Su questo punto la comunità professionale bibliotecaria dovrà svolgere un ruolo fondamentale, perché è la sola che possiede oggi le competenze necessarie per garantire integrità e univocità ai dati.

I momenti di crisi possono essere felici, perché aiutano a riconoscere e a puntare sui valori di lunga durata, cioè sugli aspetti che mantengono il loro ruolo essenziale anche quando tutto il resto sta cambiando.

I LLD non risolvono automaticamente i problemi della catalogazione: ne aprono certamente molti altri. Ma è ciò che la comunità del web chiede ed è l'unica soluzione per riprendere il passo dello sviluppo tecnologico del mondo non bibliotecario. Dobbiamo evitare il rischio che la comunità bibliotecaria faccia resistenza passiva, come succede in parte della comunità accademica mondiale, che non comprende, non accoglie e non sostiene il movimento open access.

La pubblicità è l'anima del commercio: se vogliamo le biblioteche piene di utenti, non è sufficiente che i dati siano a disposizione: devono essere reperiti dove gli utenti iniziano le loro ricerche.

## Bibliografia

Bianchini, Carlo, and Mauro Guerrini. 2009. "Quis custodiet ipsos custodes? Osservazioni sulle relazioni fra FRBR, ICP, ISBD e RDA." *Il Bibliotecario*, III S. 2009 (1, gennaio-aprile) (September): 59-85. <http://eprints.unifi.it/archive/00001743/>.

Bozzarelli, Oriana, and Maria Cassella. 2011. "Gli identificativi persistenti. A Torino un seminario di presentazione." *Biblioteche Oggi* XXIX (8): 66-70. <http://eprints.rclis.org/handle/10760/16344>.

Coyle, Karen. 2010. *Nuove direzioni in metadata*. <http://www.archive.org/details/NuoveDirezioniInMetadata>.

Darnton, Robert. 2011. *Il futuro del libro*. Milano: Adelphi.

Davis, Ian, and Tom Heath. 2009. The thirty minute guide to RDF and Linked Data. <http://www.slideshare.net/iandavis/30-minute-guide-to-rdf-and-linked-data>.

Dodds, Leigh, and Ian Davis. 2011. *Linked Data Patterns*. <http://patterns.dataincubator.org/book/>.

Heath, Tom, and Christian Bizer. 2011. *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool. <http://linkeddatabook.com/editions/1.0/>.

ICCU. 2009. *Regole italiane di catalogazione*. REICAT. Roma: ICCU.

Kumar, Vinit. 2010. Linked Data. A best practice for better knowledge transaction. In *Knowledge Transactions*. Documentation Research and Training Centre, ISI.

Library Linked Data Incubator Group. 2011. Library Linked Data Incubator Group Final Report. <http://www.w3.org/>. <http://www.w3.org/2005/Incubator/Ild/XGR-Ild-20111025/>.

- Marcum, Deanna. 2011. A Bibliographic Framework for the Digital Age (October 31, 2011): Bibliographic Framework Transition Initiative (Library of Congress). *31 october 2011*.  
<http://www.loc.gov/marc/transition/news/framework-103111.html>.
- Maurer, Hermann, Tilo Balke, F Kappe, Narayanan Kulathuramaiyer, S. Weber, and B. Zaka. 2007. "Report on dangers and opportunities posed by large search engines, particularly Google." *Institute for Information Systems and Computer Media, Graz University of Technology, Graz* (September).  
<http://www.google-watch.org/gpower.pdf>.
- Palmisano, Davide. 2010. From the Semantic Web to the Web of Data. Ten years of linking up. Lugano: Fondazione Bruno Kessler. <http://www.slideshare.net/dpalmisano/from-the-semantic-web-to-the-web-of-data-ten-years-of-linking-up>.
- Ranganathan, Shiyali Ramamrita. 1963. *The organization of libraries*. 3rd ed. London: Oxford University Press.
- Revelli, Carlo. 2011. "Il catalogo (e il catalogare): vaso di coccio tra vasi di ferro?" *Biblioteche oggi* XXIX (10): 8-12.
- Tillett, Barbara B. 2011. "Mantenere il ruolo delle biblioteche nel web semantico tramite RDA: Resource Description and Access." *JLIS.it* 2 (2): 1-19. doi:10.4403/jlis.it-4760.  
<http://leo.cilea.it/index.php/jlis/article/view/4760>.