

Dig *Italia*

Anno IX, Numero 1 - **2014**

ISSN 1972-6201

Rivista del digitale nei beni culturali

ICCU-ROMA



ICCU

Istituto centrale per il catalogo unico
delle biblioteche italiane e per le informazioni bibliografiche
<http://www.iccu.sbn.it>

Copyright © ICCU - Roma

La riproduzione totale o parziale del contenuto della rivista
è ammessa con obbligo di citazione

Un *Bibliographic Framework* per un web dei dati.

Il *Linked Data Model* e i servizi di supporto

Traduzione di Iolanda Cristaldi e di Roberto Morelato
3 gennaio 2014

Introduzione

La Library of Congress ha promosso a maggio 2011 la *Bibliographic Framework Initiative*. L'iniziativa mira a ri-definire e, nel lungo periodo, a concretizzare, un nuovo contesto bibliotecario per le realtà del settore che vedono nella "rete" un elemento centrale, al fine di realizzare una effettiva interconnessione. La forte motivazione da parte della comunità bibliotecaria di adottare nuove norme catalografiche e di rivedere il proprio modello di dati è parte fondante di questa iniziativa.

Il documento si propone, rispetto a questo obiettivo, come modello di riferimento per la comunità bibliotecaria, allo scopo di favorire la valutazione e la discussione; nel contempo è da considerare altresì in un contesto più ampio, in quanto si rivolge a una cerchia di interesse più estesa rispetto a quella bibliotecaria. Le biblioteche hanno una lunga e ricca storia nell'utilizzo della tecnologia finalizzata a realizzare economie di scala. La Library of Congress si è impegnata a sviluppare ed implementare il formato catalografico leggibile dalla macchina (comunemente conosciuto come MARC) negli anni '60, per condividere le informazioni catalografiche in forma elettronica. Ciò ha consentito di condividere, tra molti, risorse che altrimenti sarebbero state descritte più e più volte da centinaia di catalogatori.

Così nel corso del tempo il formato è rimasto sostanzialmente invariato, mentre metodi di interscambio e manipolazione dei dati hanno continuato ad evolversi e ciò ha garantito alla comunità bibliotecaria un periodo di effettiva stabilità per quanto riguarda il modello/formato dei dati.

Una storia di successo come poche altre.

Questa nuova tecnologia, ancora in fase sperimentale, ha preparato il terreno per lo sviluppo di standard, ha portato a una collaborazione diffusa e duratura tra le biblioteche e, cosa fondamentale, ha portato alla riduzione dei costi.

È ora di farlo di nuovo.

Il nuovo modello proposto è chiamato semplicemente BIBFRAME, diminutivo per *Bibliographic Framework*. È molto di più che una semplice sostituzione dell'attuale modello MARC. È la base per il futuro della descrizione bibliografica rispetto a quanto avviene su e dentro al web e il mondo interconnesso in cui viviamo. È stato progettato per integrarsi e interagire con un sistema di gestione delle informazioni più esteso rispetto a quello gestito dalle biblioteche o istituzioni analoghe.

Le linee guida sono le seguenti:

1. Distinguere chiaramente tra contenuto concettuale e manifestazione fisica (ad esempio, opere e istanze).

2. Puntare l'attenzione su una identificazione non ambigua delle entità (ad esempio, intestazioni di autorità).
3. Mettere in evidenza le relazioni tra le diverse entità.

In un mondo interconnesso su scala mondiale, è indispensabile essere in grado di citare i dati bibliografici in maniera che differenzino non solo l'opera concettuale (titolo e autore) dai dettagli fisici della manifestazione dell'opera (numero di pagine, illustrazioni), ma anche chiaramente identifichino le entità coinvolte nella creazione di una risorsa (autori, editori) e i relativi concetti (soggetti) associati alla risorsa stessa.

Normali pratiche di descrizione bibliografica almeno fino ad ora utilizzate, si sono concentrate sulla creazione di record catalografici autonomamente comprensibili, che aggregano informazioni relative al lavoro concettuale e al loro supporto fisico e facendo un ampio utilizzo di stringhe lessicali in qualità di identificatori, come, ad esempio, il nome dell'autore. Il modello BIBFRAME incoraggia invece la creazione di entità definite in maniera precisa e l'utilizzo di identificativi comprensibili dalla macchina che si prestano, a loro volta, a interpretazioni automatizzate per tali entità.

Tutto questo è in sinergia con le più attuali metodologie di gestione di dati, specialmente con quelle che sono integrate con il web. Come il web si sta evolvendo da una semplice rete di documenti collegati tra loro a una rete di documenti collegati ai loro relativi metadati, così è sempre più evidente che sono proprio questi gli elementi che favoriranno gli utenti nel reperire le informazioni sul web. Questa in definitiva, è la stessa idea che sta alla base di Schema.org, un'iniziativa lanciata da Google, Bing, Yandex e Yahoo! per creare un vocabolario universale per *web designers*, finalizzata a descrivere al meglio le informazioni contenute all'interno delle tradizionali pagine web.

Quando gli utenti iniziano le loro ricerche con

un motore di ricerca o un *social network*, il cui obiettivo è quello di aiutarli a individuare le informazioni, le organizzazioni culturali devono fornire il loro supporto nel dare le risposte, soprattutto quelle che sono contenute nelle biblioteche. Il modello BIBFRAME è stato progettato appositamente per coordinare e connettere fra loro catalogazione e metadati, creati dalle biblioteche a questo scopo. In breve, il modello BIBFRAME rappresenta il punto di accesso dell'intera comunità bibliotecaria per divenire parte di un più ampia rete di dati. Quando le biblioteche diventeranno parte di questa grande rete di dati, sfruttando l'uso di marcatori standard per differenti entità, allora l'attenzione si sposterà dal catturare e descrivere le risorse, verso il nuovo obiettivo di identificare e stabilire nuove relazioni tra le risorse stesse.

Sono comprese tutte le risorse reperibili in rete e in particolare quelle che si trovano al di fuori dei tradizionali confini dell'universo bibliotecario.

Sono queste relazioni – questi *links* – a guidare il web, trasformando il mondo delle informazioni da tanti 'silos' indipendenti a una rete di grafi che si dirama in ogni direzione. Sono ancora esse, le relazioni, a supportare i motori di ricerca ed altri servizi nell'incrementare la rilevanza nei risultati di ricerca, e cosa più importante, ad aiutare gli utenti nel trovare le informazioni che stanno cercando.

Le relazioni sono fondamentali per la concezione del web moderno, come esemplificato dal *Knowledge Graph* di Google. Tale sistema si avvale, da un lato, delle ricerche svolte dagli stessi utenti per suggerire loro altre informazioni verso le quali potrebbero orientarsi, dall'altro, attinge ai rapporti di Amazon tra le preferenze di acquisto ed i titoli dei libri, per consigliare una lista di prodotti aggiuntivi ai quali un cliente potrebbe essere interessato.

I bibliotecari sono esperti nell'identificare e catturare le relazioni bibliografiche, ed il modello BIBFRAME mette in campo una base per evidenziare questo tipo di conoscenza. Dal mo-

mento che nostri utenti cercano nel web informazioni rilevanti, tali relazioni bibliografiche contribuiranno materialmente ad arricchirlo e contribuiranno altresì ad assistere chi ricerca ulteriori informazioni.

Così, come è rilevante comprendere l'importanza della *Bibliographic Framework Initiative* ed il relativo modello in un contesto più ampio, è altresì importante considerare questo documento all'interno della comunità bibliotecaria come punto di partenza sul quale la comunità stessa continuerà a crescere.

Leggendo il modello BIBFRAME, è evidente che bisogna fare ancora molto.

È importante ricordare che anche questo modello, come MARC, deve essere in grado di adattarsi ad ogni possibile implementazione e consentire contestualmente lo scambio dei dati tra le biblioteche.

Deve essere in grado di supportare nuove regole e standard futuri di metadati, tra cui il più recente RDA (*Resource Description & Access*).

Il modello BIBFRAME deve, contemporanea-

mente, ampliare e restringere il set di formati per lo scambio dei dati bibliografici.

Anche se il modello BIBFRAME è solo una bozza, e quindi soggetto a modifiche, la Library of Congress vuole tuttavia poterlo condividere pubblicamente, non solo per informare sui progressi compiuti ma anche per stimolare un dibattito e un *feedback* costruttivo.

È la Library of Congress che conduce questa iniziativa, ma è importante che tutta la comunità bibliotecaria collabori in uno sforzo comune per creare un ambiente per la descrizione bibliografica e l'interscambio di dati e che riconosca e sfrutti le risorse di una rete di dati globalizzata.

Tutto ciò deve essere realizzato tenendo conto delle risorse e delle peculiarità delle biblioteche.

La Library of Congress attende e accoglie *feedback* sul nuovo modello BIBFRAME qui proposto.

Library of Congress

19 novembre 2012

Bibliographic Framework come modello per i Linked Data

Il concetto di *Linked Data*, seppur declinato in forme diverse, si ispira alle teorie dell'ideatore del web, Tim Berners Lee.

Nel suo articolo "*Giant Global Graph*"¹ ha espresso l'evoluzione del web con alcune considerazioni:

"La realizzazione del progetto [la creazione di internet] consisteva nel fatto che "non sono le connessioni a renderlo interessante, ma i computer". La rete è stata progettata per consentire ai computer di essere visti senza dover vedere le connessioni. Ora [il World Wide Web] aumenta ancora di più la sua forza. "Non è il computer, ma i documenti che sono interessanti". Ora si può navigare in un mare di documenti senza doversi preoccupare su quale computer sono memorizzati.

Ora la gente sta attuando una nuova evoluzione mentale, e cioè, "non sono i documenti ad essere interessanti, bensì gli argomenti di cui trattano".

Berners-Lee ricorda che le tappe di questa evoluzione si riferiscono ad un'idea di web fatta di collegamenti: web di computer (sebbene sia preferibile il termine "rete" piuttosto che "web"), web di documenti (ciò che genericamente viene chiamato "il web"), ed infine web di tutte le cose che vogliamo condividere. Sostiene che dovremmo estendere i principi di base del web in modo più diretto ai dati (ad esempio, il contenuto di basi di dati tradizionali), e che non dobbiamo essere timorosi nel fare collegamenti a risorse anche non strettamente digitali, come ad esempio, persone, oggetti tangibili o astratti, luoghi, e così via.

Questa visione espansiva di collegamenti è chiamata "Web dei dati" e costituisce il fondamento dei così detti *Linked Data*.

L'obiettivo di questa prima bozza è quello di fornire uno schema in grado di sostenere future modellazioni, nonché la codifica di beni bibliografici tradizionalmente effettuata con MARC 21.

Questo modello, sfruttando il web quale architettura di base, si prefigge di raggiungere due obiettivi:

- 1) favorire una maggiore integrazione delle risorse bibliografiche esistenti e
- 2) creare una mappa per raffinare, riqualificare e sviluppare approcci alternativi.

(Ulteriori approfondimenti su Linked Data e Linked Open Data a pp. 24)

La comunità bibliotecaria è da sempre stata pioniera nella creazione, nell'organizzazione e nella gestione anche collaborativa, di un flusso di dati in ambito culturale.

MARC 21 rappresenta l'ultima evoluzione di un formato di interscambio all'interno di un settore alquanto specialistico.

Le biblioteche producono e mantengono una quantità enorme di dati di alta qualità ben oltre i confini tradizionali della biblioteca. Se noi estendiamo il formato MARC 21 verso un modello di *Linked Data*, espandiamo anche l'utilità e il valore di questi dati proprio nella maniera in cui servono alla comunità bibliotecaria e alle istituzioni culturali.

Nel modellare MARC 21 come un web dei dati, è importante sia scomporre che ricostruire il suo stesso patrimonio informativo.

MARC ha una ricca storia nel saper supportare le esigenze in continua evoluzione della comunità bibliotecaria, e tale capacità la si può riassumerle in tre funzioni principali²:

- Dati riferiti all'aspetto intellettuale dell'opera,

¹ <http://dig.csail.mit.edu/breadcrumbs/node/215>

² Delsey, T. (2002) "Functional Analysis of the MARC 21 Bibliographic and Holdings Format." www.loc.gov/marc/frbr/functional-analysis.html

- Dati riferiti all'istanza effettiva dell'opera, cioè quanto può essere tangibile, accessibile attraverso una rete di dati, etc.
- Record di metadati, ad esempio numeri di controllo, record di *holding* e altre annotazioni

Dopo aver così scomposto la struttura informativa dei dati, si può cominciare a ricostruire i concetti impliciti racchiusi all'interno del set di dati come una risorsa linkabile di informazioni. Questi dati intesi come "risorsa MARC" possono quindi venire ri-assemblati in una nuova coerente architettura che funge da base per un nuovo quadro bibliografico e consente così la catalogazione cooperativa a un livello più granulare (persone, luoghi, soggetti, organizzazioni, ecc.). Poi, come per il web che si struttura in un'architettura di informazioni, così ogni volta che verranno aggiornate queste risorse MARC (ad esempio qualcuno aggiunge nuove informazioni su una persona, nuove mappature relative a un oggetto, etc.) ci potranno essere delle notifiche automatiche per aggiornare automaticamente il sistema.

E in più, tutti questi flussi informativi possono essere sfruttati efficacemente in collezioni locali o speciali ed essere facilmente sfruttati in ambito di catalogazione partecipata.

Questo documento fornisce una bozza per un modello di *Linked Data* per il nuovo *Bibliographic Framework* e identifica una serie di ulteriori spunti di discussione che aiuteranno a plasmare il futuro di questo lavoro.ù

Il *Bibliographic Framework* (BIBFRAME) *Linked Data Model* è concepito per essere una base di discussione e dialogo per la comunità bibliotecaria.

Certamente non è completo, né, se esso si verificherà efficace, sarà mai "completo" perché potrà e dovrà essere ampliato per accogliere un *range* di nuove applicazioni e risorse descrittive che non possiamo al momento nemmeno immaginare.

Il BIBFRAME *Linked Data Model* vuole armonizzare i seguenti elementi:

- flessibilità nell'accogliere futuri ambiti catalografici e scenari dell'informazione totalmente nuovi;
- il web come un modello architettonico per esprimere e connettere informazioni decentralizzate;
- adozione di un modello sociale e tecnologico al di fuori della comunità bibliotecaria;
- sviluppo sociale e tecnologico all'interno della stessa comunità bibliotecaria;
- precedenti tentativi di esprimere i materiali bibliografici come *Linked Data*;
- utilizzo di apparecchi tecnologici per espletare compiti meccanici, mentre si riconosce l'esperto in materia (il bibliotecario) come l'esplicito cervello dietro la macchina;
- riconoscere tutti gli sforzi già fatti nell'ambito della diffusione delle informazioni svolti dall'intera comunità bibliotecaria, editoriale, archivistica e museale;
- riconoscere la solida e proficua tradizione nella metodologia comune di trasmissione delle informazioni bibliografiche.

IL modello BIBFRAME

Il modello BIBFRAME è composto dalle seguenti classi:

- **OPERA CREATIVA:** un'entità che rappresenta l'essenza a livello concettuale dell'oggetto da catalogare;
- **ISTANZA:** un'entità che rimanda ad una manifestazione concreta e individuale dell'Opera;
- **AUTHORITY:** un'entità che rimanda alle nozioni chiave di intestazioni di autorità, che hanno precedentemente definito le relazioni rappresentate nell'Opera e nell'Istanza.

Esempi di *authority* possono essere persone, luoghi, argomenti, organizzazioni etc.

- **ANNOTAZIONI:** una risorsa che amplia una risorsa BIBFRAME con ulteriori informazioni aggiuntive. Un esempio può es-

sere quello che viene fornito dalle informazioni relative alla collocazione, alla copertina o alle recensioni.

La rappresentazione grafica illustrata nella figura 1, mostra un esempio di relazione complessa modellata con BIBFRAME.

Per quanto riguarda i modelli di relazione tra entità, incluso FRBR, il modello BIBFRAME riconosce entità, attributi e relazioni tra diverse entità. Esso sfrutta il modello proposto dal *World Wide Web Consortium's Resource Description Framework (RDF)* in grado di identificare univocamente ogni entità (risorsa), attributo, e relazione tra entità (proprietà) come risorsa Web.

Ciò consente di attivare eventuali annotazioni (come ad esempio l'implementazione di un vocabolario aggiuntivo o l'ampliamento di altre comunità bibliografiche) qualora fosse necessario.

BIBFRAME, L'Opera creativa

L'Opera creativa BIBFRAME, che d'ora in poi chiameremo semplicemente Opera, è, sul piano catalografico, un'entità puramente concettuale. L'Opera infatti rappresenta un'entità astratta in quanto non può essere circoscritta ad un singolo oggetto materiale. Essa va concepita come un elemento gestito dal Web, che non solo rimanda a contenuti

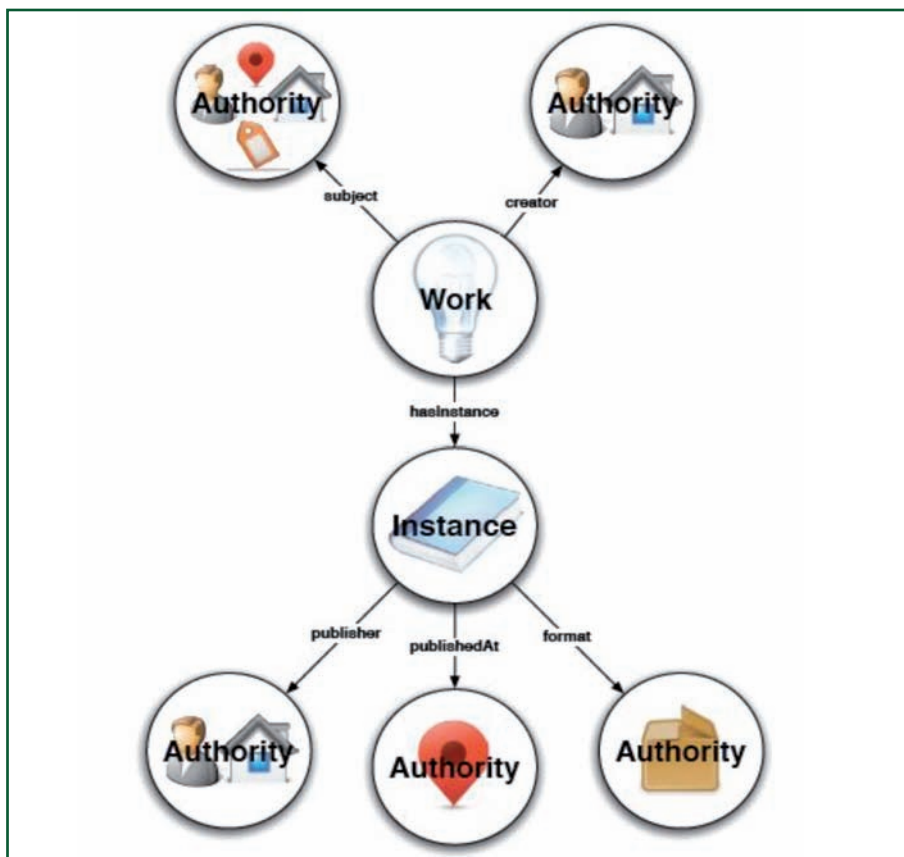


Figura 1: Una rappresentazione grafica del Linked Data Model BIBFRAME che identifica la relazione tra Opera e Istanza e la loro contestualizzazione rispetto ad una Authority quale risorsa raggiungibile via Web.

condivisi da varie altre Istanze associate ad essa, ma che si pone anche come punto di riferimento per altre Opere. Tutte le proprietà in comune alle Opere includono così relazioni di tipo concettuale con le *Authorities* BIBFRAME sia nell'ambito della soggettazione (argomento, persona, luogo geografico, ecc.) sia in riferimento a entità (persona fisica, ente, congresso, ecc.), sia in riferimento a entità responsabili della loro creazione. Le Opere, ad esempio, possono essere messe in relazione tra loro secondo un rapporto *parte/tutto*.

BIBFRAME, Le Istanze

L'Istanza BIBFRAME rappresenta la concretizzazione individuale, fisica o digitale, di un'Opera. Essa esiste in quanto elemento base gestito dal web e include sia proprietà specifiche da materializzare che relazioni contestuali con *Authorities* BIBFRAME, relative alla pubblicazione, alla produzione e alla distribuzione della risorsa materiale.

Ogni Istanza BIBFRAME è rappresentativa di una sola Opera BIBFRAME.

BIBFRAME, Le authorities

Le *authorities* BIBFRAME rimandano alle nozioni chiave di intestazioni di autorità in quanto punto d'arrivo delle relazioni rappresentate nell'Opera e nell'Istanza.

Esempi di *authorities* BIBFRAME possono essere persone, luoghi, argomenti, organizzazioni ecc.

Dal punto di vista catalografico, le *authorities* forniscono gli strumenti necessari per la disambiguazione e contemporaneamente, sincronizzano tra loro tutte le informazioni. Per quanto riguarda gli utenti, invece, le *authorities* costituiscono un'efficace strumento di controllo da utilizzare quale aiuto nella navigazione e nella contestualizzazione di Opere e

Istanze relazionate tra di loro.

Le *authorities* BIBFRAME non sono pensate per sostituire o competere con i risultati fino ad oggi ottenuti; esse mirano piuttosto a fornire un modello di astrazione comune che possa proporsi come anello di congiunzione tra essi, così da renderli ancor più efficaci.

L'Annotazione nel modello BIBFRAME

Le Biblioteche producono, mantengono e valorizzano una quantità enorme di dati di alta qualità che va ben oltre i confini tradizionali della biblioteca. Alla luce di questo, l'iniziativa *Bibliographic Framework* si pone l'obiettivo di "accogliere e differenziare metadati tecnici e computerizzati, comprese le annotazioni (recensioni e commenti) e dati di utilizzo³."

Anziché predefinire e limitare il potenziale d'utilizzo di tali dati, il modello BIBFRAME fornisce l'impalcatura necessaria per far sì che questi possano essere facilmente integrati sia dalle biblioteche che da altri utenti.

La figura 2 è una rappresentazione grafica del BIBFRAME *Linked Data Model* e riguarda un *framework* visto all'interno di un contesto rispetto al quale l'integrazione delle informazioni deriva da svariate fonti. In questo esempio il depositario dell'informazione (cioè chi è il depositario dell'Istanza e dove questa può essere reperita) non è una caratteristica insita nell'Istanza stessa quanto piuttosto come una dichiarazione, da parte di una biblioteca specifica, sull'Istanza in questione. Ciò consente alle biblioteche, registrandola come annotazione, di stabilire metadati descrittivi dell'Istanza in questione più dettagliati e che includono annotazioni relative a revisione, fruitori specifici, utilizzo, regolamentazione, ecc.

Le 'recensioni' attendibili, come rappresentato nella figura 2 dall'Annotazione all'Opera, costituiscono un ulteriore e autorevole esempio

³ A Bibliographic Framework for the Digital Age (October 31, 2011)
<http://www.loc.gov/marc/transition/news/framework-103111.html>

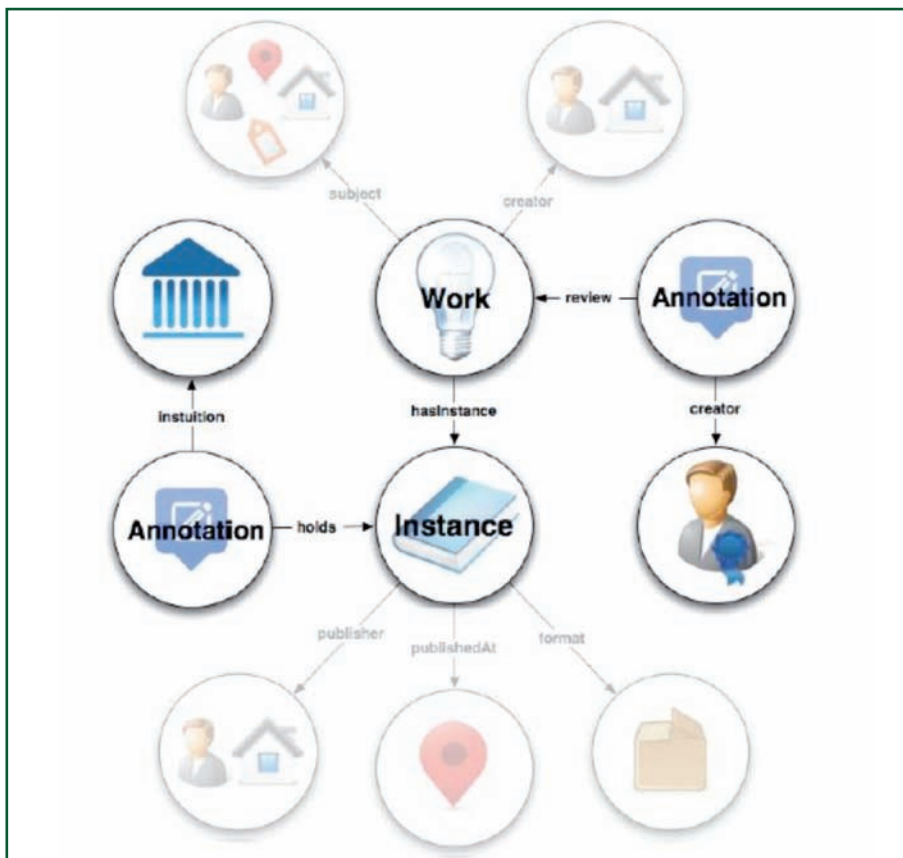


Figura 2 :Una rappresentazione grafica del BIBFRAME Linked Data Model all'interno di un contesto rispetto al quale l'integrazione delle informazioni deriva da svariate fonti

di tale modello legato alle annotazioni. Queste annotazioni possono essere di tipo "interno" (quando ad esempio una determinata biblioteca inserisce un commento/recensione personale all'Opera) o "esterno" (quando invece si aderisce a un servizio esterno di recensione per la diffusione dei dati).

Ad ogni modo, i principi fondanti del *Linked Data* fanno sì che queste annotazioni possano essere estese, rappresentate e gestite facilmente.

Ulteriori esempi pratici di tale modello di annotazioni ci vengono forniti dalla possibilità di aderire, da parte delle istituzioni, a servizi di *cover art* per sostituire le proprie versioni o

permettere ad utenti autorizzati di contribuire con immagini proprie. In tal senso, il servizio commerciale offerto da Amazon rappresenta un modello che potrebbe essere seguito dai fruitori abituali del patrimonio bibliotecario e culturale.

Nella figura 3 è mostrato l'uso delle annotazioni nell'Istanza BIBFRAME al fine di descrivere un oggetto all'interno di una collezione bibliotecaria ed il relativo contributo apportato da un utente autorizzato.

Un altro esempio pratico di tale modello basato sull'annotazione è fornito dal crescente bisogno che si avverte, nell'era digitale, di rintracciare informazioni autorevoli. Progetti co-

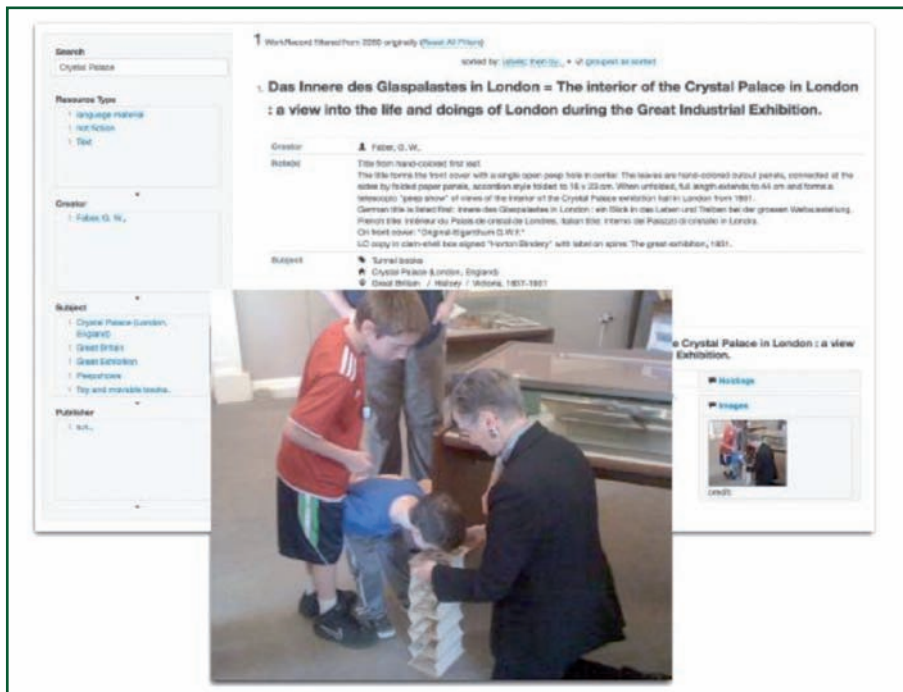


Figura 3 ::Esempio di fotografia fornita dall'utente quale Annotazione di un'Istanza BIBFRAME

me il *Reference Extract*⁴ si propongono di raccogliere i riferimenti di bibliotecari o esperti del settore nel rispondere alle domande attraverso i servizi offerti dalla Library of Congress o il servizio OCLC *Question Point*.⁵ Tali informazioni, tra cui quelle utilizzate per decretare le risorse più autorevoli, vengono dapprima raccolte e processate, per poi essere diffuse sul web attraverso vari sistemi.

L'istantanea della figura 4 mostra un esempio del prototipo sviluppato dal *Reference Extract Project* che, ai tradizionali risultati proposti dal motore di ricerca, affianca altre annotazioni autorevoli.

La possibilità di annotare facilmente le risorse bibliografiche, così da poter dare risposte attendibili alle domande dei fruitori, agevola lo

sviluppo di nuovi servizi che possono affiancarsi ai tradizionali motori di ricerca al fine di aiutare gli utenti finali nel determinare la credibilità delle risorse.

Ulteriori iniziative nell'ambito delle tecniche catalografiche di nuova generazione, riprendono i fondamenti del *Linked Data*; esse tuttavia non hanno pienamente sviluppato un'architettura tale da poter associare ed esprimere tali annotazioni strutturate su più livelli. Un problema pratico è rintracciabile, infatti, nel flusso di lavoro, e nello specifico, nella copia e condivisione dei record.

A tale proposito, BIBFRAME è un *framework* unico nel suo genere in quanto fornisce un ambiente integrato per la gestione dell'annotazione dei record, consentendo altresì di ge-

⁴ La Fondazione Mac Arthur finanzia il *Reference Extract* al fine di promuovere sia la professionalità dei bibliotecari che il valore della ricerca Web.
<http://www.oclc.org/news/releases/2011/20111.htm>

⁵ Question Point <http://www.questionpoint.org/>

stire il ciclo di vita degli eventi che permette, a sua volta, di individuare la tracciabilità degli errori o di riuscire a determinare chi sia l'attuale possessore di informazioni associate ad una determinata Istanza.

Ciò è simile al principio di provenienza adottato in ambito archivistico e museale, nonché nell'ambito del Web semantico, sebbene l'approccio BIBFRAME presupponga un processo di annotazione più semplice grazie all'utilizzo dei record.

La possibilità di arricchire le risorse con annotazioni ed eventuali servizi, è segno distintivo del web e di come la rete abbia rivoluzionato molti settori di ricerca.

Il processo di annotazione BIBFRAME costituisce il perno principale di un sistema retto sul potenziale conoscitivo delle biblioteche piuttosto che sugli interessi commerciali che ad oggi dominano la rete.

L'obiettivo finale del *Bibliographic Framework* è quello di sviluppare un modello di riferimento al quale altri potranno a loro volta guardare. Esso tiene in considerazione del fatto che ogni comunità ha una visione diversa delle proprie risorse e pertanto diverse necessità ri-

spetto alla descrizione delle medesime. Ciò è particolarmente evidente se, invece di considerare il documento nella sua fisicità, ci si concentra su immagini (fisse e in movimento), risorse cartografiche e archivistiche, collezioni museali e qualsiasi manufatto di matrice culturale. Molti modelli contenutistici stabiliscono relazioni di natura gerarchica che dapprima vengono rappresentate in un grafo RDF, per poi essere semplificate dal modello BIBFRAME. Ad esempio il concetto di Opera/Istanza espresso in BIBFRAME può riflettere le relazioni di FRBR - dopo aver applicato un'operazione di semplificazione per ridurre le cose il più possibile - in termini di un grafo piuttosto che in una relazione di tipo gerarchico. Di conseguenza, il prossimo passo sarà quello di conciliare le innovazioni del modello BIBFRAME con gli standard catalografici di RDA-*lite*.

Il vocabolario BIBFRAME

Il modello BIBFRAME è completamente espresso in RDF.

Il vocabolario RDF per la bozza BIBFRAME, a

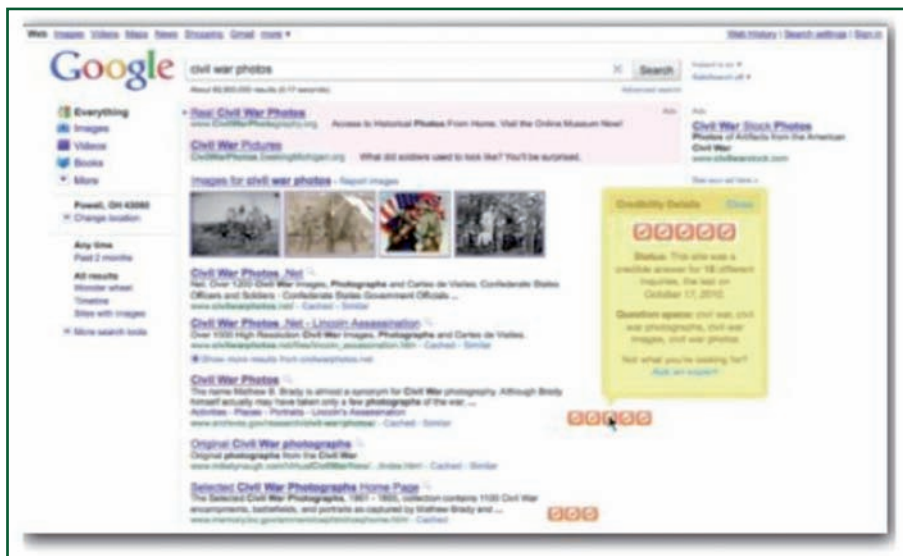


Figura 4 :Un esempio di notazione autorevole rispetto al termine di ricerca 'foto della guerra civile'.

breve disponibile, costituirà un pratico strumento di navigazione del modello RDF.

Il *namespace* proposto per il modello BIBFRAME è <http://loc.gov/bibframe/vocab>. Qui sarà possibile trovare a breve, una serializzazione in formato RDF/XML di tale vocabolario.

Se la segnalazione di ogni *namespace* è antichissima delle numerose iniziative bibliografiche *Linked Data*, risulta cruciale dichiarare la responsabilità e l'autorità che si cela dietro il *framework* BIBFRAME, così da evitare confusioni e ridurre al minimo la complessità dei formati.

Ai detentori degli standard bibliotecari sarà affidato il compito di salvaguardare i collegamenti tra gli elementi del modello BIBFRAME e vocabolari quali Dublin Core, FOAK, SKOS e ulteriori che potranno essere sviluppati d'ora innanzi per rispondere alle diverse esigenze del lavoro bibliotecario. La persistenza degli identificatori non è di natura tecnica bensì risponde a esigenze di tipo organizzativo e comunitario. In tal senso, si ritiene necessaria una regolamentazione continuativa che possa stabilire criteri di gestione dei cambiamenti e non.

<!-- Opera -->

```
<Report id = "http://bibframe/work/frbr-report">
  <title> Functional requirements for bibliographic records :</title>
  <titleRemainder> final report / IFLA Study Group on the Functional Requirements for
  Bibliographic Records ; approved by the Standing Committee of the IFLA Section on
  Cataloguing.</titleRemainder>
  <creator resource = "http://bibframe/auth/org/ifla" />
  <subject resource = "http://bibframe/auth/topic/cataloging" />
  <subject resource = "http://bibframe/auth/topic/bibliography" />
  <subject resource = "http://bibframe/auth/topic/frbr" />
  <abstract> The purpose of this study is to delineate in clearly defined terms the
  functions performed by the bibliographic record with respect to various media, vari-
  ous applications, and various user needs. The study is to cover the full range of func-
  tions for the bibliographic record in its widest sense- i.e., a record that encompasses
  not only descriptive elements, but access points (name, title, subject, etc.),
  other 'organizing' elements (classification, etc.), and annotations. </abstract>
  <language>English</language>
  <hasInstance resource="http://bibframe/inst/frbr-1997-09-01:0" />
  <hasInstance resource="http://bibframe/inst/frbr-1997-09-01:1" />
  <hasInstance resource="http://bibframe/inst/frbr-1997-09-01:2" />
</Report>
```

<!-- Instanza -->

```
<HardcoverBook id="http://bibframe/inst/frbr-1997-09-01:0">
  <date>1998</date>
  <place resource="http://bibframe/auth/geo/münchen" />
  <publisher resource="http://bibframe/auth/org/k.g.saur" />
  <isbn>359811382X</isbn></HardcoverBook>
```

<!-- Instanza -->

```
<DigitalResource id="http://bibframe/inst/frbr-1997-09-01:1">
  <link>http://www.ifla.org/files/cataloguing/frbr/frbr\_2008.pdf</link>
  <format>application/pdf</format> <date>1997-09-01</date> <publisher
  resource="http://bibframe/auth/org/ifla" />
```

</DigitalResource>

<!-- Instanza -->

```
<DigitalResource id="http://bibframe/inst/frbr-1997-09-01:2">
  <link>http://archive.ifla.org/VII/s13/frbr/frbr_current_toc.htm</link>
  <format>text/html</format> <date>2007-12-26</date> <publisher
  resource="http://bibframe/auth/org/ifla" />
```

</DigitalResource>

<!-- BIBFRAME Topic -->

```
<Topic id="http://bibframe/auth/topic/frbr"> <label>FRBR (Conceptual model)</label>
<hasIDLink resource="http://id.loc.gov/authorities/subjects/sh2007002541" /></Topic>
```

<!-- BIBFRAME Topic -->

```
<Topic id="http://bibframe/auth/topic/bibliography"> <label>Bibliography</label>
<generalSubdivision>Methodology</generalSubdivision>
<hasIDLink resource="http://id.loc.gov/authorities/subjects/sh85013838" /></Topic>
```

<!-- BIBFRAME Topic -->

```
<Topic id="http://bibframe/auth/topic/cataloging"> <label>Cataloging</label>
<hasIDLink resource="http://id.loc.gov/authorities/subjects/sh85020816" /></Topic>
```

<!-- BIBFRAME Organization -->

```
<Organization id="http://bibframe/auth/org/ifla"> <label>IFLA Study Group on the
Functional Requirements for Bibliographic Records</label>
<link>http://www.ifla.org/</link> <hasIDLink resource="http://id.loc.gov/authori-
ties/names/nr98013265" /></Organization>
```

<!-- BIBFRAME Organization -->

```
<Organization id="http://bibframe/auth/org/k.g.saur"> <label>K.G. Saur</label>
<link>http://www.degruyter.com/</link> <hasIDLink
resource="http://id.loc.gov/authorities/names/nr91037301" /></Organization>
```

<!-- BIBFRAME Organization -->

```
<Organization id="http://bibframe/auth/org/k.g.saur"> <label>K.G. Saur</label>
<link>http://www.degruyter.com/</link> <hasIDLink
resource="http://id.loc.gov/authorities/names/nr91037301" /></Organization>
```

<!-- BIBFRAME Place -->

```
<Place id="http://bibframe/auth/geo/münchen"> <label>Munich (Germany)</label>
<hasIDLink resource="http://id.loc.gov/authorities/names/n79059670" /></Place>
```

Un modello RDF come definito nella serializzazione XML che rappresenta le relazioni tra Opere e Istanze, è rappresentato nella figura 5.

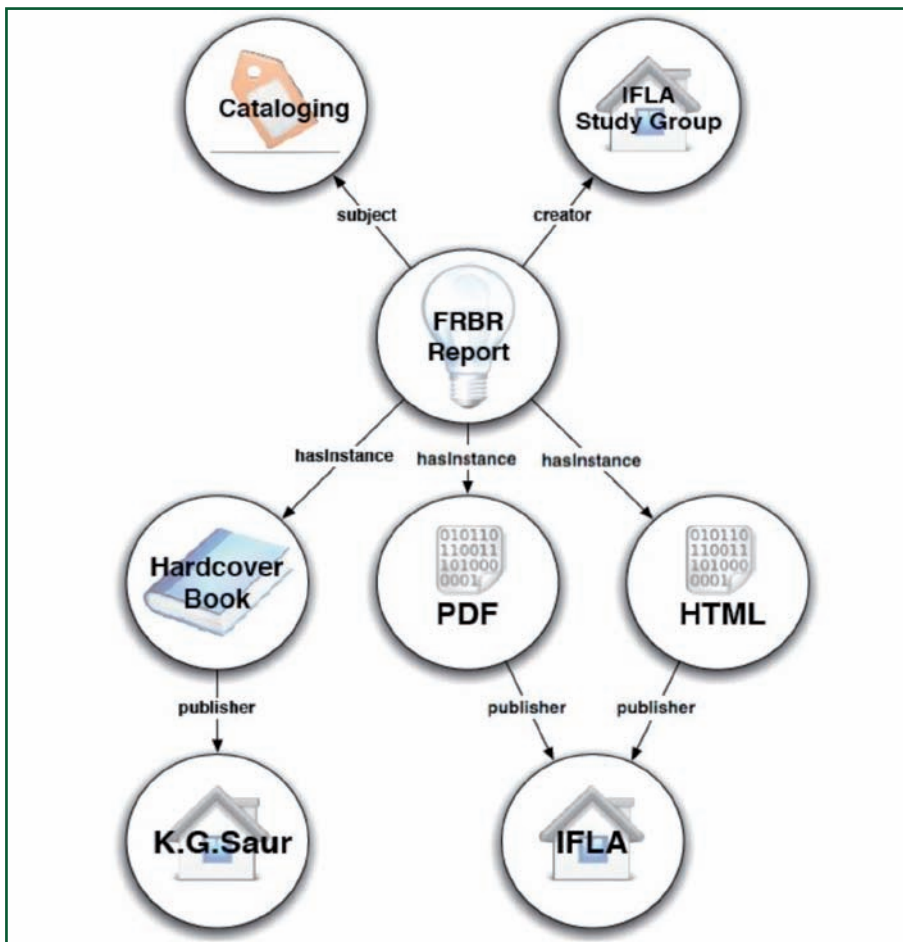


Figura 5 :Un modello di riferimento RDF collegato alla serializzazione XML. Il Report FRBR ha tre Istanze associate a due diversi editori.

Dopo aver fornito un esempio di codifica dell'Opera e delle relative Istanze e Authority ad essa associate, la seguente serializzazione rappresenta un'annotazione relativa alle *Library holdings* per "Functional Requirements for Bibliographic Records: Final Report", pubblicato da K. G. Saur.

```
<!-- Holdings Annotation -->
<Holdings id="http://bibframe/annot/holdings/frbr-1997-09-01:0">
  <annotates resource="http://bibframe/inst/frbr-1997-09-01:0" />
  <institution resource=" http://bibframe/auth/org/ohio.university.alden" />
  <callNumber>025.3 F979 1998</callNumber>
  <access>circulating</access> <status>available</status>
</Holdings>
```



```

<!-- BIBFRAME Organization -->
<Organization id="http://bibframe/auth/org/ohio.university.alden">
  <label>Ohio University, Alden Library</label>
  <city>Athens</city> <state>OH</state>
  <zip>45701</zip>
  <link>http://www.library.ohiou.edu/</link>
  <hasIDLink resource="http://id.loc.gov/authorities/names/n2003039990" />
</Organization>

```

Serializzazione del modello BIBFRAME

Possono esserci diverse serializzazioni del BIBFRAME *Linked Data Model*.

La serializzazione XML qui esposta (secondo il modello dati RDF), sebbene destinata a cam-

biare, ne fornisce un esempio concreto. Esso si propone di fornire una codifica serializzata di una determinata Opera, delle Istanze corrispondenti e delle informazioni di *authority* ad essa associate.

L'Opera in questione è *"Functional Requirements for Bibliographic Records: Final Report"*. Il record

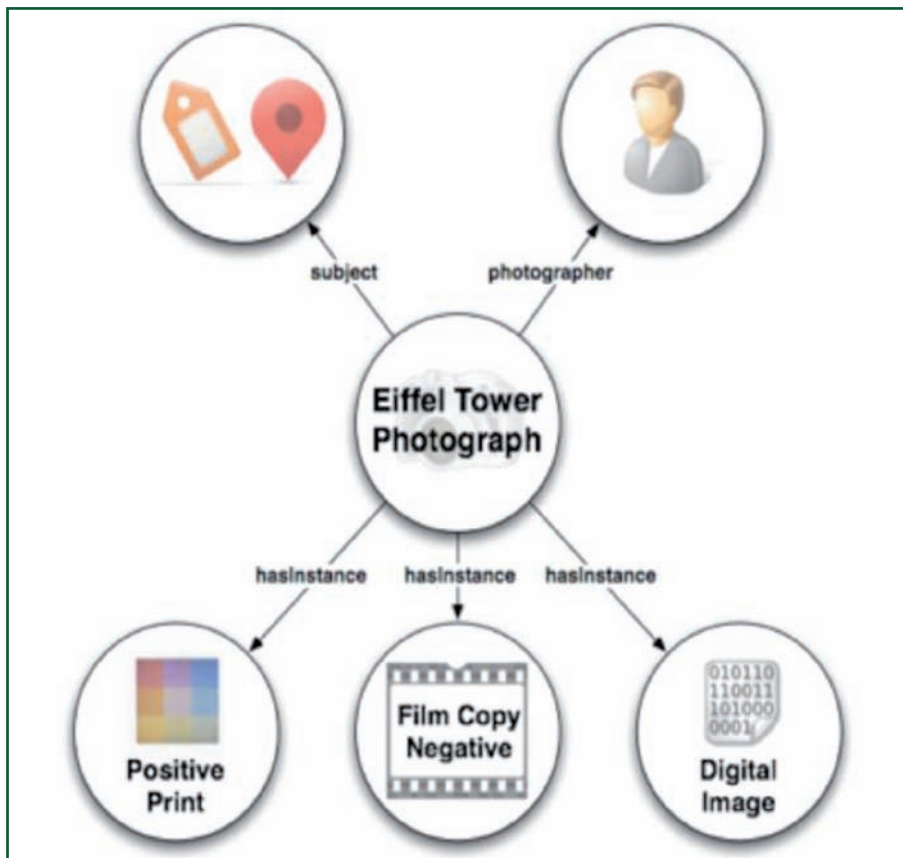


Figura 6: Un esempio di risorsa non testuale. La fotografia ha tre istanze ad essa associate.

originale in formato BIBFRAME è disponibile al link: <http://lccn.loc.gov/2001433363>.

L'esempio include tre Istanze (fisica, in PDF, in HTML) e le informazioni di *authority* (soggetti, autori, editori ecc.), sebbene non rappresenti una completa mappatura da MARC 21 verso BIBFRAME.

I link riportati nell'esempio che segue, sebbene non attivi, sono riportati solo a titolo illustrativo allo scopo di mostrare come impiegare URLs per definire risorse BIBFRAME.

Strumenti di supporto

Fondamentali, al fine di rafforzare il modello BIBFRAME, risultano essere gli strumenti e i servizi di supporto che favoriscono il passaggio da MARC a una struttura *Linked Data*. Il loro scopo è quello di tradurre il flusso dei dati dal formato MARC 21 al modello BIBFRAME.

La figura 7 fornisce un esempio di interfaccia *Linked Data* a faccette, modellata su dati BIBFRAME usando Exhibit⁶, un *framework open source* sviluppato da MIT Simile Project per la pubblicazione di pagine web interattive arricchite di annotazioni semantiche.

La figura 8 riporta una visualizzazione espansa delle intestazioni di autorità per autore e soggetto di una determinata opera all'interno del modello BIBFRAME.

Sono presenti degli indicatori visivi (ad esempio stelline o segni di spunta) che evidenziano i collegamenti che intercorrono tra le risorse originarie in formato

Marc 21 con i link che vengono attivati dalle intestazioni di autorità.

I collegamenti che forniscono intestazioni di autorità utilizzati nello *streaming* di questo flusso di dati, includono la *Library of Congress Linked Data Service (ID)*⁷ e la *VIAF Virtual International Authority File*⁸.

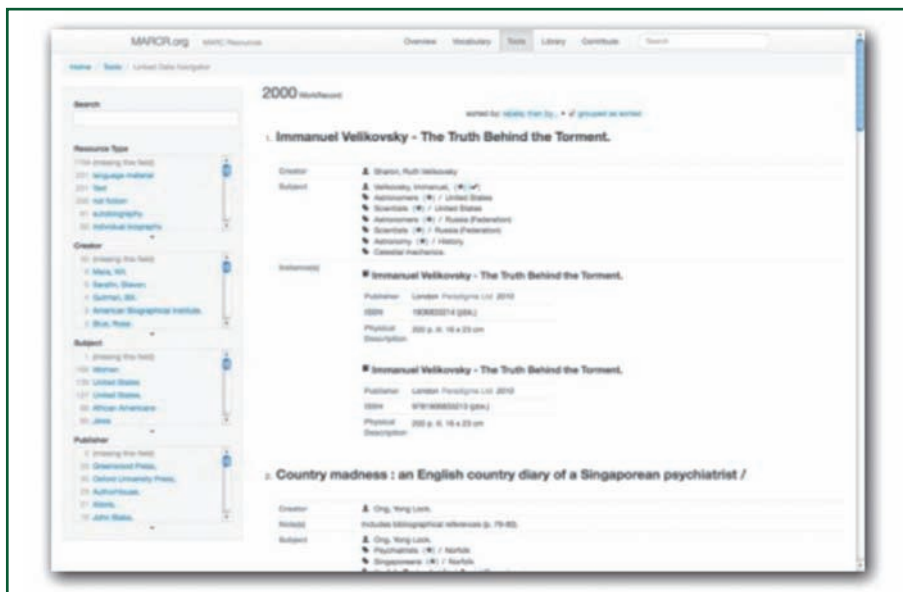


Figura 7: Un esempio di interfaccia *Linked Data* a faccette che trasferisce all'interno del modello BIBFRAME la struttura dei dati in MARC 21.

⁶ Sviluppato da MIT Simile Project - <http://www.simile-widgets.org/exhibit/>

⁷ Library of Congress Linked Data Service (ID) - <http://id.loc.gov/>

⁸ VIAF Virtual International Authority File - <http://viaf.org/>

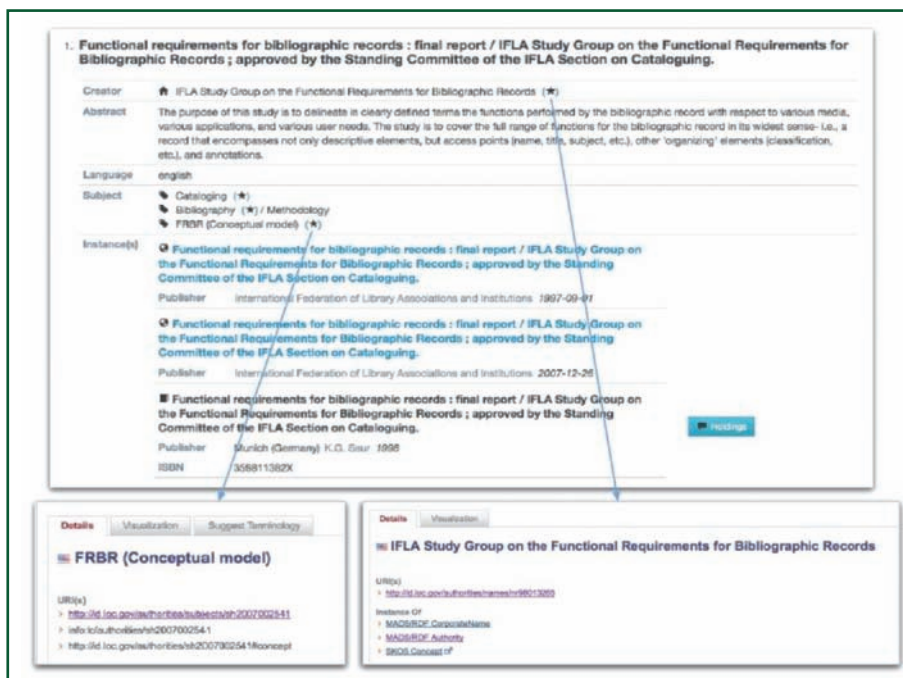


Figura 8 :Una visualizzazione espansa delle intestazioni di autorità per autore e soggetto.

Le stelline evidenziano i collegamenti che intercorrono tra le risorse originarie in formato Marc 21 con i link che vengono attivati dalle intestazioni di autorità. Vedi: <http://id.loc.gov/>.

L'obiettivo del progetto *Linking Open Data del W3C SWEO (Semantic Web Education and Outreach)* è di espandere il Web pubblicando diversi open dataset come RDF e impostando link RDF tra dati provenienti da differenti risorse.

Linked Data e LOD: quadro generale

Linking Open Data (LOD) è una comunità non ufficiale che si è occupata di diffondere i criteri per la creazione effettiva di un Web dei dati, secondo un'idea abbozzata da Tim Berners-Lee in un documento sui *Linked Data*. Il W3C ha da sempre sostenuto questa comunità, combinando la visione rispetto alle particolarità semantiche che arricchiscono il Web con il pragmatismo che invece caratterizza il Web 2.0.

Tratto dalle pagine *wiki* di LOD ospitate presso W3C:

L'enfasi che W3C manifesta rispetto all'utilizzo di RDF è comprensibile, considerato che ha supportato questa tecnologia per un decennio, anche se tuttavia stanno emergendo nuovi e influenti pareri che indirizzano LOD nel valutare che limitarsi strettamente al formato RDF non sia attualmente la strategia più valida. LOD supporta il modello concettuale di RDF, sebbene si concentri maggiormente su configurazioni interconnesse e condivise, piuttosto che sintattiche.

Il Gruppo di lavoro LOD, che occupa una posizione periferica rispetto al sistema centrale W3C, sostiene i vantaggi dell'*opening data* e considera RDF, Atom, JSON e simili esclusiva-

mente come degli strumenti a disposizione degli sviluppatori del Web per arricchire i propri dati.

Linked Open Data (LOD) 101

Volendo entrare più nello specifico sulle implicazioni di LOD, bisogna partire dai seguenti quattro criteri abbozzati da Tim Berners-Lee nel suo articolo *“Linked Data”*:

1. Usare URI per identificare oggetti presentati sul Web come risorse.
2. Usare HTTP URI in modo che gli utenti possano dereferenziare questi oggetti (effettuare operazioni di *look up*).
3. Fornire informazioni utili sull’oggetto quando la sua URI è dereferenziata.
4. Includere *link* ad altre URI collegate per migliorare la ricerca di altre informazioni in rete.

Il **primo principio** concerne il tentativo di esporre il più possibile le informazioni attraverso URI, non solo per pagine Web, ma anche per moduli, righe di database, dati personali e metadati, *log* delle transazioni, regole di mercato e politiche commerciali, servizi vari. Se la condivisione di tutto questo ha un senso, allora è importante che ogni elemento abbia un URI.

Per quanto riguarda la sicurezza, spesso ci si affida ai tradizionali programmi di protezione dati; le persone hanno una tale fiducia nel Web che lo utilizzano per fare affari, comprare oggetti e prenotare viaggi. La sicurezza dei dati garantita dalla rete è comprovata dall’utilizzo delle migliori procedure.

Il **secondo principio** presuppone l’abbandono dei complessi schemi ID (e lo stesso vale anche per gli schemi URI) per attenersi strettamente al protocollo http, che è riuscito a servire Internet in maniera eccellente. Ciò garantisce l’accesso ad una svariata quantità di strumenti e risorse.

Il **terzo principio** presuppone che i dati, ai

quali le persone possono accedere tramite URI, debbano essere espressi in un formato comune per la condivisione in rete.

A tal proposito, XML si candida come linguaggio preferenziale, sebbene non in tutte le sue varianti. Bisogna infatti usare XML affinché sia semanticamente trasparente, così da rendere i contenuti descritti in XML facilmente processabili dalla macchina.

RDF è il principale formato utilizzato dal Gruppo LOD, tuttavia, nonostante l’elevata trasparenza semantica offerta, il supporto per RDF non è così diffuso come quello per XML. Per trarre il meglio da entrambi i formati, è possibile utilizzare GRDDL, sistema che permette di avere visualizzazioni XML utilizzando lenti RDF.

Il **quarto principio**, infine, può essere altresì definito “criterio di condivisione”. Nei primi tre si è posto l’accento sulla necessità di creare dei “puntatori” Web per accedere ai dati, massimizzandone l’utilità. Il passo successivo risiede nel fornire, senza reticenze, il più alto numero di *link* possibili a partire da questi “puntatori”. Non è possibile prevedere fino a quale livello di profondità possa essere utilizzata la rete dei dati, ed è in quest’ottica che si inserisce l’obiettivo principale di LOD: favorire qualsiasi uso dei dati diverso da quello per il quale era stato concepito.

Bolle espandibili

Il Gruppo LOD mantiene un diagramma di *dataset* pubblici che possono essere utilizzati in accordo con i principi suddetti.

La figura 9 presenta una recente versione del diagramma: la grandezza di ogni bolla indica approssimativamente la quantità di dati presenti nel *dataset*.

Elementi d’interesse sono:

- Freshmeat, sito internet che fornisce una lista di programmi *open source*.
- MusicBrainz, un database *online* di informazioni riguardanti musica registrata e album.

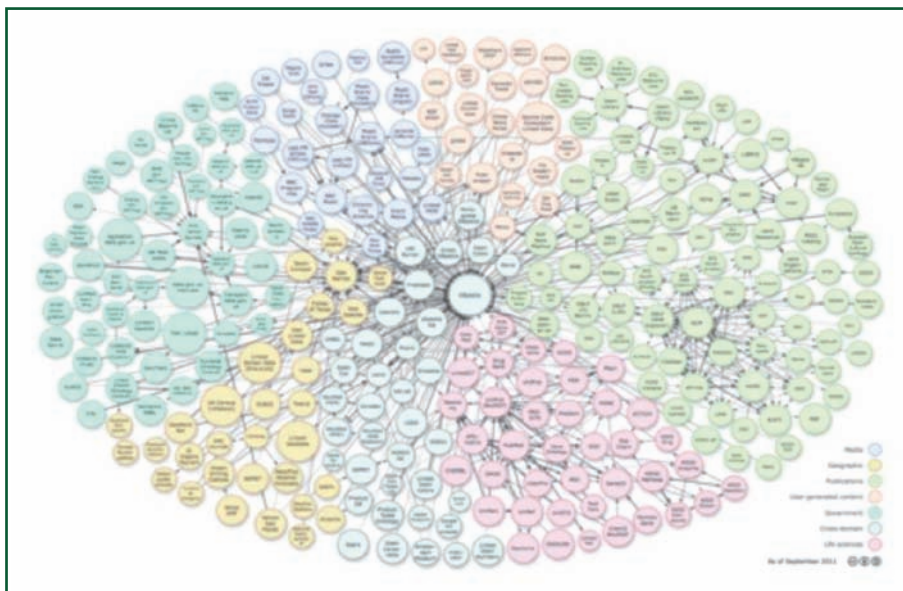


Figura 9 :Diagramma tematico della Linking Open Data cloud, di Richard Cyganiak e Anja Jentzsch. <http://lod-cloud.net/>

- Project Gutenberg, una lodevole iniziativa che si propone di fornire versioni elettroniche liberamente riproducibili di libri stampati.
- FOAF, un *network* di relazioni sociali che utilizza il formato RDF.
- DBPedia, un progetto LOD per l'estrazione di informazioni strutturate da Wikipedia.

Nonostante il suo indiscusso successo, molto può essere ancora fatto per migliorare la rete. A ben vedere, uno degli obiettivi principali di W3C è quello di un Web Semantico volto a creare una rete di dati semanticamente trasparenti. A tal proposito, il modello LOD ben si presta al raggiungimento di questo obiettivo, adeguandosi perfettamente ai più importanti criteri del Web 2.0. Ad esempio, prendiamo in considerazione *mashup*, un'applicazione che include contenuti da più sorgenti per creare un servizio completamente nuovo.

Grazie a LOD, non è necessario che essa implichi un processo consapevole e specifico per ogni sito; è possibile acquisire i dati in maniera semplice e trasparente da una varietà di fonti e servizi sparsi sul Web. Alcuni saranno liberi mentre altri avranno delle restrizioni d'accesso per ragioni commerciali o di sicurezza, problematiche, queste, che gli sviluppatori del Web hanno già in parte risolto.

LOD intende rendere più agibile la scoperta da parte degli utenti degli elementi presenti sul Web e la loro collocazione, in modo che questi ultimi possano riutilizzarli a proprio vantaggio nei modi più disparati.

I moderni progettisti del Web dovrebbero quindi iniziare a pensare in primo luogo a che tipo di risorse informative e non sono presenti nell'applicazione, cercando di fare il massimo per fornire un HTTP URI ben progettato e un formato dati ricco semanticamente, con un numero crescente di *link* collegati.

Applicazione del modello *Linked Managed Data*

Alla base del *Linked Managed Data* ci sono identificatori, relazioni, linee programmatiche e servizi.

Identificatori: La comunità bibliotecaria è da sempre stata pioniera nella creazione di identificatori forti e universali delle risorse bibliografiche e delle intestazioni di autorità associate. Fondamentale risulta essere l'applicazione di questo *framework* alle numerose entità che fanno parte della gestione dei dati bibliografici e dell'universo catalografico, utilizzando sistemi di identificazione per libri, serie, giornali, autori, risorse digitali attendibili e servizi vari. Ciò vale anche per i tentativi meno rigorosi di identificazione, come per l'utilizzo di parole chiave (*tag*) nelle folksonomie.

Il cammino verso la razionalizzazione, sebbene possa apparire arduo, è già stato intrapreso, così come dimostra la creazione di *authority files*.

Relazioni: Uno dei fattori che ha determinato il successo della rete risiede in una semplicità nello stabilire i *link*, tale da promuovere un sistema di interrelazioni testuali (*cross referencing*).

E se per gli identificatori è importante stabilire un sistema di autorità, con le relazioni, invece, il successo deriva dalle libere connessioni.

Il tristemente famoso codice di risposta 404 "not found" inizialmente preoccupò i puristi dell'ipertesto, tuttavia la storia ha dimostrato che la forza del *Web linking* risiede precisamente nella sua semplicità di definizione. L'algoritmo *PageRank* di Google esemplifica perfettamente la possibilità dei *link* di arricchirsi progressivamente. La comunità bibliotecaria può beneficiare della stessa facoltà che permette a Google di stabilire relazioni web tra applicazioni chiave.

Politiche: Alcune istituzioni che si occupano dei beni culturali gestiscono la loro comunità

con un sistema di licenze, iscrizioni e altri strumenti che stabiliscono termini e condizioni di condivisione delle risorse, tra cui cataloghi e record di autorità. Essi saranno poi inseriti nel *framework* dei metadati, sottoforma di *query* dinamica, insieme alle procedure stabilite per indicare i processi relativi a determinate risorse.

Servizi: I programmi bibliotecari promuoveranno le capacità della comunità di farsi carico dei nuovi sistemi promossi dalle istituzioni culturali in una prospettiva innovativa, favorendo le molteplici interazioni tra risorse autorevoli sul Web.

Come illustrato in precedenza, l'uso di sistemi orientati al web in materia di identificatori e relazioni, costituisce il fattore più importante. Il *framework* dei metadati verrà altresì usato per implementare la descrizione e favorire il reperimento dell'applicazione/servizio.

La sfida che la comunità deve affrontare risiede nel riconsiderare le biblioteche ed il corpus di conoscenze, dati e informazioni ad esse relazionati alla luce di questi nuovi sviluppi.

Iniziative bibliotecarie collegate

Molti esperti nel settore dei beni culturali sono arrivati contemporaneamente alla medesima conclusione di voler adottare il modello *Linked Data*. Tra le varie iniziative, la *Library of Congress* promuove il raggiungimento di una piena interoperabilità tra i dati. Al fine di divulgare gli sforzi messi in atto nel proporre un modello *Linked Data* per MARC, si è deciso di investigare esiti e risultati ottenuti da molteplici iniziative che caratterizzano questa ricerca. A tal riguardo, nei paragrafi successivi verranno prese in esame varie considerazioni basate sia su riflessioni personali, frutto delle esperienze lavorative nel settore, sia sulla documentazione fornita a supporto e da ulteriori comunicazioni riguardanti i suddetti progetti.

British Library

Una panoramica generale

La *British Library* sta sviluppando una versione della *British National Bibliography* sotto forma di *Linked Open Data*. Al momento, il progetto comprende solo opere pubblicate ma in futuro mira ad includere anche pubblicazioni periodiche, opere in più parti, risorse integrate (cioè a fogli mobili), *kits* e piani editoriali.

Realizzazione

L'intero lavoro, così come anche il presente documento, sono da considerarsi in continua evoluzione. La *British Library* ha reso disponibile un *dataset* (vedi il paragrafo *End-Points per accesso ai dati*), sebbene non ne garantisce l'invariabilità; non bisognerà meravigliarsi, quindi, se i suoi contenuti, parte di uno sviluppo continuo e in itinere, possano cambiare. La *British Library* aggiorna il *dataset* regolarmente e continua a lavorare per fronteggiare le varie problematiche.

Data Model e Serializzazione

Alla base del suddetto lavoro, vi è il *corpus* bibliografico della *British National Bibliography (BNB)* che conta più di tre milioni di titoli. Esso differisce dall'iniziale tentativo bibliografico di usare RDF: non solo infatti si pensa di codificare l'insieme di *records* MARC in formato RDF/XML, come già fatto precedentemente da altre biblioteche, ma si ricorre a "elementi di interesse" quali persone, luoghi ed eventi collegati ad un determinato documento. Sebbene si sia tenuto conto dell'indiscusso valore del catalogo BNB, si è cercato di svincolarsi dal formato e dalle procedure che negli anni hanno portato alla sua creazione.

Su <http://www.bl.uk/bibliographic/pdfs/bldatamodelbook.pdf> è possibile consultare il grafico del modello dati appena citato.

Le procedure *Linked Data* adottate in rete da istituzioni governative, commerciali e accademiche, presentano questi "elementi di interesse" riutilizzando schemi preesistenti, possibil-

mente di tipo descrittivo; rimangono tuttavia aree per le quali non è possibile rifarsi a strutture preesistenti perché inadeguate a rendere un determinato significato. Per ovviare a tali problematiche, è stato stilato il vocabolario *British Library Terms (BLT)* il cui aggiornamento andrà di pari passo con l'inserimento dei dati.

Questo schema, il cui *namespace* è <http://www.bl.uk/schemas/bibliographic/blterms#>, insieme ai seguenti vocabolari, viene utilizzato per descrivere le risorse presenti nella *British National Bibliography (BNB)*:

- Bibliographic Ontology (Bibo)
- Bio: a Vocabulary for Bibliographical Information
- Dublin Core International Standard Bibliographic Description (ISBD)
- Org: An Organization Ontology
- SKOS
- RDF Schema
- OWL
- FOAF
- The Event Ontology (<http://purl.org/NET/c4dm/event.owl#>)
- WGS84 Geo Positioning and the preliminary version of RDA

I dati sono collegati a svariate fonti '*Linked Open Data*', nello specifico: VIAF, LCSH, Lexvo, *GeoNames* (per il paese di pubblicazione), MARC per la codifica del paese e della lingua, Dewey.info e RDF Book Mashup.

End-Points per accesso ai dati

Il sottoinsieme di dati della *British National Bibliography (BNB)*, incluso il materiale pubblicato e/o distribuito in Inghilterra dal 1950, è consultabile ai seguenti indirizzi:

- SPARQL endpoint:
<http://bnb.data.bl.uk/sparql>
- Describe endpoint:
<http://bnb.data.bl.uk/describe>
- Search service:
<http://bnb.data.bl.uk/search> and
<http://bnb.data.bl.uk/items>

- The Data Hub:
<http://thedatahub.org/dataset/bluk-bnb>
(file contenente tutte le triple)

Osservazioni

Gli sforzi messi in atto dalla *British National Bibliography (BNB)* sono stati particolarmente rilevanti nella definizione del modello BIBFRAME. Nello specifico, la modellazione dei soggetti BNB ha contribuito efficacemente alla teorizzazione del modello di intestazione di autorità in BIBFRAME.

Deutsche Nationalbibliothek

Una panoramica generale

La Deutsche Nationalbibliothek (DNB) offre un servizio *Linked Data* che permetterà alla comunità del web semantico di usare per intero il patrimonio bibliografico nazionale a sua disposizione, compresi i dati relativi alle intestazioni di autorità. L'obiettivo a lungo termine è quello di attrarre nuovi gruppi di destinatari, puntando a soddisfare le richieste e i bisogni specifici di ognuno di essi.

A tal proposito, la DNB ha prodotto una documentazione volta ad esplorare i bisogni di nuove categorie, come ad esempio, i fornitori di servizi commerciali (dai gestori dei motori di ricerca e dei sistemi di gestione della conoscenza agli enti di ricerca e le organizzazioni *no-profit*).

L'obiettivo dichiarato è quello di apportare, con il suo servizio dati, un contributo decisivo all'infrastruttura globale dell'informazione, spianando la strada ai moderni servizi web commerciali e non.

Realizzazione

La DNB ha iniziato ad avvalersi dei *Linked Data* nel 2010 per pubblicare i suoi dati relativi alle intestazioni di autorità. Il servizio ha continuato ad espandersi e da Gennaio 2012 include anche dati bibliografici. Per iniziare, c'è stata la conversione dei dati bibliografici

delle collezioni più importanti e delle pubblicazioni periodiche (riviste, giornali e collezioni) del *German Union Catalogue of Serials (ZDB)* in formato *Linked Data*.

Data Model e Serializzazione

DNB rende disponibili i servizi *Linked Data* secondo le serializzazioni RDF/XML e RDF/*turtle*. I *records* dei dati bibliografici modellati in RDF non presentano lo stesso livello di strutturazione come i loro corrispettivi MARC. Vengono inoltre utilizzate proprietà specifiche a seconda del tipo di risorse rappresentate (pubblicazioni periodiche, libri, collezioni, articoli, serie, ecc.). Ovviamente l'implementazione, nella sua fase iniziale, ha riguardato un graduale lavoro di affinamento e diffusione del modello. Questo graduale adattamento coinvolge anche la scelta dei vocabolari utilizzati e il motivo è legato al fatto che le risorse attualmente esistenti per strutturare le ontologie non sempre possono descrivere a pieno la complessità dei dati, come nel caso di quelli di tipo bibliografico. Si è deciso pertanto di utilizzare comunque questo tipo di risorse per fornire una descrizione del dato bibliografico essenziale, combinando vocabolari già utilizzati per la descrizione di ontologie, quali:

- Bibliographic Ontology (Bibo)
- International Standard Bibliographic Description (ISBD)
- Dublin Core
- RDF Schema
- OWL
- FOAF

End-Points per accesso ai dati

La Deutsche Nationalbibliothek (DNB) offre un servizio di accesso all'intero set dei dati in formato *Linked Data* al seguente indirizzo:

http://www.dnb.de/EN/Service/DigitaleDienst/LinkedData/linkedata_node.html.

È possibile effettuare un accesso ad istanze individuali degli stessi, all'indirizzo:

<http://thedatahub.org/dataset/deutsche-na>

tionalbibliografie-dnb.

Sono disponibili inoltre le intestazioni di autorità della Gemeinsame Normdatei (*Integrated Authority File*) all'indirizzo:

<http://thedatahub.org/dataset/dnb-gemeinsame-normdatei>.

I file di autorità dei paesi germanofoni, la Gemeinsame Normdatei (GND) e i dati bibliografici del servizio *Linked Data* adotteranno la licenza *Lizenz Creative commons Zero CCO BY-NC-ND*. Ciò riguarda le risorse il cui URI inizia con <http://d-nb.info/ddc>

Osservazioni

L'approccio alla descrizione delle risorse basato sui "tipi" è stato incluso nel modello BIBFRAME, di cui si è ampiamente trattato nella prima parte del presente elaborato. Alla luce del progetto BIBFRAME va altresì letto l'apprezzabile tentativo di affinare e diffondere il modello nel tempo. Gli *authority files* della DNB dei paesi germanofoni sono stati rilevanti rispetto alla definizione di un modello base per gli *authority files* di BIBFRAME.

OCLC/WorldCat

Una panoramica generale

Il 20 giugno 2012 l'OCLC ha aggiornato il catalogo WorldCat.org per includere il linguaggio di marcatura basato sulla struttura di Schema.org. Questo lavoro fornisce collegamenti aggiuntivi tra le descrizioni bibliografiche e ulteriori servizi *Linked Data* come VIAF (per i *files* di autorità) e di progetti per la gestione tassonomica dei soggetti (tra cui FAST e DDC). Grazie all'adozione del sistema di marcatura proposto da Schema.org per tutti i testi, riviste e pubblicazioni bibliografiche, l'intera versione pubblicamente disponibile di WorldCat è ora accessibile ai *crawler* Web intelligenti, come Google e Bing, che possono fare uso di questi metadati all'interno di indici di ricerca e altre applicazioni.

Realizzazione

L'OCLC, pur continuando ad incrementare le funzionalità offerte dal servizio, garantisce costante continuità di accesso ai dati bibliografici distribuiti secondo il modello *Linked Data*. Fin dall'inizio gli sforzi si sono concentrati nel cercare di utilizzare il più possibile il vocabolario fornito da Schema.org, estendendo questo vocabolario solo se necessario. Questa prima versione viene descritta come sperimentale ed è pertanto destinata a cambiare.

Data Model e Serializzazione

Il *Linked Data Model* di OCLC include, oltre a Schema.org, un'estensione proprietaria della biblioteca. Ogni elemento WordCat include estensioni RDFa e *microformats*.

End-Points per accesso ai dati

La versione del catalogo *WordCat.org* in *Linked Data* è stato diffuso da OCLC sotto licenza *Open Data Common Attribution*.

Osservazioni

Per la progettazione e lo sviluppo del lavoro di OCLC, è stato coinvolto il gruppo Zepheira. L'obiettivo di questo sforzo è stato quello di proiettare i dati bibliografici disponibili in WorldCat in un vocabolario incentrato sulla struttura di Schema.org per potenziare i risultati di quei motori di ricerca che utilizzano questo sistema di marcatura. Inoltre, questo sforzo è stato un primo passo per identificare ciò che è necessario al di là del vocabolario Schema.org per favorire una più efficace ricerca all'interno dei materiali bibliografici. Questo progetto, combinato con le documentate caratteristiche di MARC (in particolar modo per la descrizione delle risorse), ha contribuito alla fondazione del modello BIBFRAME.

Schema.org

Sebbene non sia esclusivamente pensato per le biblioteche, Schema.org è un importante progetto, nato un anno fa, per il Web e i *Linked Data*.

Nel 2011 i tre maggiori motori di ricerca, Google, Bing e Yahoo, hanno lanciato l'iniziativa Schema.org al fine di fornire ai *webmaster* gli strumenti per integrare le informazioni con un *set* di metadati. L'iniziativa consente di migliorare la qualità dei risultati di ricerca ed è pensata per favorire l'evoluzione verso un Web dei dati.

La compatibilità con il suddetto modello è diventata un parametro imprescindibile per molte organizzazioni, che mirano a potenziare la loro presenza sul web e il valore aggiunto della rete associato ai propri dati.

Anche le biblioteche si muoveranno verso Schema.org; a tal proposito sono già stati abbozzati dei moduli ed è stato affrontato un dibattito su come poter rappresentare il materiale bibliotecario. OCLC ha prodotto, in via sperimentale, una rappresentazione di WorldCat compatibile con Schema.org ma questo lavoro, al momento, non gode di un riconoscimento ufficiale.

Osservazioni

Sebbene gli obiettivi commerciali di Schema.org non coincidano con quelli della comunità bibliotecaria, le possibili ricadute di questo progetto sulla suddetta comunità sono considerevoli. Se da un lato il modello BIBFRAME è focalizzato sulla definizione di un modello comune che possa promuovere l'interoperabilità tra i *Linked Data*, dall'altro il *design* (IO METTEREI PROGETTO) è influenzato anche da Schema.org.

Un'ulteriore ricaduta del lavoro BIBFRAME sarà quella di coinvolgere la comunità Schema.org stessa, al fine di supportare maggiormente le esigenze delle biblioteche.

RDA (Resource Description and Access)

Una panoramica generale

RDA è il modello catalografico che sostituisce AACR2; gli editori sono l'*American Library*

Association, la *Canadian Library Association* e il *Chartered Institute of Library and Information Professionals* (CILIP) nel Regno Unito.

La *Joint Steering Committee for Development of RDA* (JSC) è responsabile per il mantenimento del modello RDA. Questo comitato è composto da rappresentanti dell'*American Library Association*, dell'*Australian Committee on Cataloguing*, della *British Library*, del *Canadian Committee on Cataloguing*, del *Chartered Institute of Library and Information Professionals*, della *Deutsche Nationalbibliothek* e della *Library of Congress*. RDA è un'applicazione dei modelli concettuali FRBR (e FRAD) e comprende un insieme di istruzioni modellate secondo FRBR.

Nel 2009 le tre biblioteche nazionali degli USA (*Library of Congress*, *National Library of Medicine* e *National Agricultural Library*) hanno deciso di adottare una linea comune rispetto all'implementazione o meno di RDA, sulla base dei risultati di un test che ne valutava la possibile realizzazione rispetto a vincoli di natura economica, tecnica e operativa.

Realizzazione

RDA è stato rilasciato inizialmente a giugno 2010.

All'inizio del 2012 il *Test Coordinating Committee* di RDA ha raccomandato alle tre biblioteche nazionali degli USA (*Library of Congress*, *National Library of Medicine* e *National Agricultural Library*) di adottare RDA solo quando saranno rispettati i parametri emersi dall'indagine sopra citata.

L'implementazione avrà inizio non prima del 1. gennaio 2013.

Nel contempo la *Library of Congress* ha annunciato che, a partire dal 31 marzo 2013, la catalogazione rispetterà in toto il modello RDA.

Anche altre biblioteche, come la *British Library*, la *Library and Archives Canada*, la *National Library of Australia*, la *National Library of New Zealand* e la *Deutsche Nationalbibliothek* si propongono di implementare RDA nel 2013.

Data Model e Serializzazione

Il modello dati RDA, sebbene basato su *Functional Requirements for Bibliographic Resources* (FRBR), non stabilisce un metodo specifico per la codifica o la visualizzazione dei dati.

Esistono mappature da RDA a FRBR, MARC 21 e MODS.

Osservazioni

Vedi la sezione FRBR, immediatamente successiva.

Functional Requirements for Bibliographic Resources (FRBR)

Una panoramica generale

FRBR è un modello concettuale entità-relazione sviluppato dall'IFLA, incentrato sulle esigenze degli utenti rispetto alle operazioni di accesso e ricerca in cataloghi bibliografici *online* e altri *database* bibliografici. FRBR rappresenta un approccio incentrato sulle relazioni tra entità, che consente la navigazione gerarchica tra le relazioni stesse, attraverso la realizzazione di collegamenti.

A supporto di questo modello, FRBR si concentra sulle seguenti funzioni:

- **trova** (*find*): per trovare le entità corrispondenti ai criteri di ricerca dell'utente;
- **identifica** (*identify*): per identificare un'entità;
- **seleziona** (*select*): per selezionare un'entità adeguata alle esigenze dell'utente;
- **ottieni** (*obtain*): per acquisire o accedere all'entità descritta.

Realizzazione

Molte ricerche si sono concentrate sulla possibile ricaduta che l'implementazione di FRBR potrebbe avere sugli OPAC. La OCLC, ad esempio, ha promosso vari progetti di ricerca FRBR con *record WorldCat*. FRBR è stato altresì implementato dall'*Australian Literature Gateway* (AustLit) con alcune modifiche e ag-

giunte al modello iniziale.

La *Library of Congress*, invece, metteva a disposizione uno strumento per convertire *set* di risultati derivanti da una ricerca in una struttura gerarchica secondo il modello FRBR.

Molte aziende commerciali, inoltre, hanno messo a punto implementazioni di FRBR.

Si annoverano esperienze come quelle di Virtua, il catalogo sviluppato da VTLs Inc che è stato diffuso dando la sua piena adesione al modello FRBR (Virtua 2005), e di Portia creato da VisualCat, un sistema integrato di catalogazione capace di fondere tipi diversi di metadati in un unico *framework* semantico basato su RDF e FRBR (VisualCat 2005).

Ulteriori recenti implementazioni sono visibili su FRBRblog: *Library Thing*, *Drupal*, *Australian Music Centre* e la *University of Toronto*.

Data Model e Serializzazione

Il modello entità-relazione è composto da tre gruppi: entità, attributi e relazioni tra entità. Il primo gruppo di entità si riferisce agli aspetti sotto i quali può essere vista una produzione intellettuale (opera, espressione, manifestazione, item); il secondo gruppo si riferisce alle entità responsabili dei contenuti artistici e intellettuali (persone o enti collettivi); il terzo gruppo si riferisce ai soggetti delle opere (concetti, oggetti, eventi, luoghi, più tutte le entità dei due gruppi precedenti).

In quanto modello concettuale, FRBR non definisce una sintassi specifica.

Varie iniziative, tra cui il *Variation/frbr project* dell'Università dell'Indiana (<http://www.dlib.indiana.edu/projects/vfrbr/index.shtml>), sono state promosse al fine di descrivere uno schema XML per FRBR.

Osservazioni

Sia RDA che FRBR hanno giocato un ruolo chiave nel riqualificare gli sforzi catalografici, alla luce di principi funzionali (*"from strings to things"*) e nel fornire uno schema di requisiti a supporto della catalogazione futura.

L'approccio olistico rispetto alle operazioni di recupero e accesso, come definito dal lavoro FRBR, è stato un principio guida per il modello proposto in questo documento.

ONIX

ONIX è una famiglia di *standard* per la trasmissione di metadati dettagliati relativi a libri, periodici e altri media, che utilizza elementi di dati comuni. Gli standard comprendono *ONIX for Books*, *ONIX for Serials*, e ONIX per termini della licenza. *ONIX for Books* è lo *standard* internazionale per la descrizione e la trasmissione delle informazioni sulla produzione editoriale in formato elettronico.

ONIX for Serials è un gruppo di formati XML per trasmettere informazioni su pubblicazioni periodiche e abbonamenti.

ONIX è sviluppato e sostenuto da EDItEUR, da *Book Industry Communication* in Inghilterra e da *Book Industry Study Group* negli USA. Esistono inoltre gruppi di utenti in Australia, Belgio, Canada, Finlandia, Francia, Germania, Italia, Olanda, Norvegia, Russia, Spagna, Svezia e Repubblica Coreana.

Realizzazione

Ad aprile 2009 EDItEUR ha annunciato (<http://www.bisg.org/news/news.php?n=d&id=360&cid=1>) una nuova versione di *ONIX for Books*: ONIX 3.0.

Questa versione di ONIX è la prima dal 2001 a non essere retrocompatibile con le versioni precedenti e, cosa ancor più rilevante, fornisce un mezzo per una migliore gestione dei prodotti digitali. Una versione riveduta (3.0.1) è stata successivamente rilasciata a gennaio 2012.

Una lista delle organizzazioni che hanno implementato *ONIX for Books* nelle proprie aziende è consultabile all'indirizzo: <http://www.editeur.org/111/users-and-services-directory.html>. Essa include editori, distributori commerciali, compagnie logistiche internazionali, svi-

luppatori di *software*, fornitori di servizi digitali e rappresenta solo una parte delle organizzazioni che utilizzano ONIX.

Attualmente gli autori del presente documento non sono a conoscenza di alcuna implementazione rispetto ad *ONIX for Serials*.

Data Model e Serializzazione

In quanto standard basato su XML, ogni versione di *ONIX for Books* consiste in un XML DTD (Document Type Definition) e/o Schema XML. Questo viene distribuito assieme alla documentazione relativa alle specifiche per la trasmissione di ciascun ONIX *message* o *data file*. Insieme a queste specifiche, EDItEUR fornisce vari strumenti XML, oltre alle norme per implementare ONIX, assicurando la gratuità di tutti questi materiali.

ONIX non è un *database*, bensì un modo per trasferire dati tra *database* diversi; molti membri di EDItEUR e di altre organizzazioni offrono *software* commerciali o applicazioni web per potenziare l'implementazione della messaggistica ONIX. Altri invece hanno sviluppato soluzioni proprie che implementano gli standard di comunicazione ONIX.

I DTD in formato XML di ONIX sono basati sul modello Indecs, creato a partire da un comune ed essenziale modello commerciale. L'importante idea di fondo è riassunta nelle parole di Godfrey Rust (uno dei creatori originari di Indecs):

"le persone costruiscono oggetti, li usano e (dal punto di vista commerciale) fanno affari con essi".

Il modello Indecs presenta quattro diversi livelli concettuali:

- **Astrazione:** l'oggetto è descritto nella sua forma intangibile, come concetto;
- **Espressione:** un evento, ovvero, una creazione (può essere un'esecuzione);
- **Manifestazione:** è la realizzazione concreta di un'espressione (può essere un libro, un *file* digitale, come un MP3);
- **Item:** una singola istanza di ciascuna manifestazione.

Ci sono molte similitudini tra FRBR e Indecs, sebbene abbiano diverse esigenze funzionali, e ciascuno ha dato maggiore rilevanza nello sviluppo del relativo modello, rispetto agli elementi che riteneva più importanti. Entrambi, tuttavia, forniscono la propria personale visione in materia di risorse e relazioni: ognuno ha le sue esigenze. In linea di massima, i due modelli sono compatibili; di conseguenza una reale integrazione di metadati provenienti da schemi FRBR e Indecs risulta fattibile, sebbene con i dovuti accorgimenti.

Osservazioni

Il progetto ONIX/Index è indirettamente collegato a BIBFRAME attraverso i molteplici tentativi di adeguamento di ONIX a RDA. *ONIX for Serials* apre la strada a nuovi campi d'indagine e prepara il terreno per un eventuale modello BIBFRAME per le pubblicazioni periodiche.

Conclusioni e prospettive future

L'iniziativa *Bibliographic Framework Transition* condotta dalla *Library of Congress* ha lo scopo di tradurre MARC 21 in un *Linked Data Model* (LD).

Il presente documento abbozza un modello iniziale per l'interscambio di dati in un contesto *Linked Data* basato sull'analisi e la sintesi di attività collegate.

La strada verso il perfezionamento di questo progetto si snoda lungo il seguente percorso:

- Analisi costante delle iniziative ad esso collegate;
- *Vocabulary Navigator*, ovvero uno strumento per la ricerca all'interno delle definizioni utilizzate per identificare Opere, Istanze e intestazioni di Autorità, relazionando questi termini a codici MARC 21;
- Sviluppo del modello RDF di BIBFRAME sulla base della serializzazione XML, comprensivo di esempi di codice;

- Strumenti di supporto per la trasformazione da MARC 21 verso BIBFRAME;
- *Browser* operanti su dati strutturati secondo il modello *Linked Data* di BIBFRAME utili a dimostrare il funzionamento o la mera rappresentazione del formato MARC 21 quale *Linked Data*, al fine di provarne i benefici (sia dal punto di vista della catalogazione che degli utenti).

Glossario dei termini utilizzati

HTML (HyperText Markup Language): principale linguaggio di marcatura per pagine Web.

HTTP (Hypertext Transfer Protocol): un protocollo a livello applicativo per sistemi informativi distribuiti, collaborativi, ipertestuali, come ad esempio il web.

JSON (JavaScript Object Notation): un formato di interscambio di dati tra computer basato su testo, comprensibile anche all'uomo, utilizzato per la rappresentazione di semplici strutture di dati e *array* associativi.

Linked Data: un metodo di esporre, condividere e connettere dati sul web.

Linked Open Data (LOD): un metodo di esporre, condividere e connettere dati liberamente disponibili sul web.

RDFa (Resource Description Framework-in-attributes): si tratta di una raccomandazione W3C che aggiunge un *set* di estensioni a livello di attributo XHTML per l'incapsulamento di metadati per l'arricchimento di documenti web. La mappatura del modello dati RDF permette sia l'inserimento di triple RDF in documenti XHTML sia la loro estrazione.

REST (Representational state transfer): un tipo di architettura *software* per i sistemi di ipertesto distribuiti, come il World Wide Web.

RDF (Resource Description Framework): una famiglia di specifiche proposte da W3C (World Wide Web Consortium), originariamente pensate come un modello per esprimere metadati. Attualmente rappresenta un metodo generale per la descrizione concettuale e la modellazione di informazioni implementate in risorse web, attraverso una varietà di formati sintattici.

RDFS (Resource Description Framework Schema): una famiglia di specifiche del W3C che, aggiungendosi agli standard RDF, definisce nuovi tipi di classi con proprietà certe e fornisce gli strumenti necessari per creare un vocabolario RDF (a volte denominato ontologia).

Serializzazione: il processo di conversione di una struttura di dati o di un oggetto in un formato che può essere memorizzato (in un file, in un'area di memoria, o trasmesso attraverso una connessione di rete) e successivamente ricreato nello stesso stato o in uno differente.

SPARQL: un linguaggio e protocollo per l'interrogazione di *dataset* RDF.

Web Semantico: un *framework* comune per la condivisione e il riutilizzo di informazioni sul Web, superando i confini delle applicazioni, delle imprese e delle comunità. Basato su standard Resource Description Framework (RDF), è uno sforzo cooperativo guidato dal W3C, al quale partecipano molti ricercatori e partner industriali.

SKOS (Simple Knowledge Organisation System): una famiglia di linguaggi formali creata per rappresentare tesauri, classificazioni, tassonomie e qualsiasi tipo di vocabolario

strutturato. SKOS si basa su RDF e RDFS e il suo obiettivo principale è di consentire una facile pubblicazione di vocabolari strutturati per il Web Semantico.

URI (Uniform Resource Identifier): una stringa di caratteri utilizzata per identificare un nome o una risorsa su Internet. Lo scopo principale di questa identificazione è consentire l'interazione con le rappresentazioni della risorsa in rete (tipicamente il *World Wide Web*) utilizzando protocolli specifici. Utilizzare gli URI per la gestione ontologica delle conoscenze favorisce la loro riusabilità da parte dei gruppi interessati.

Per ulteriori dettagli:

http://en.wikipedia.org/wiki/Uniform_Resource_Identifier

URL (Uniform Resource Locator): un sottoinsieme di URI (Uniform Resource Identifier) che specifica dove una risorsa identificata è disponibile e il relativo meccanismo per il suo recupero sul web.

W3C (World Wide Web Consortium): il principale organismo internazionale di normalizzazione per il *World Wide Web*.

XML (Extensible Markup Language): un insieme di regole per la codifica dei documenti in formato elettronico.

XSLT (XSL Transformations): un linguaggio dichiarativo, basato su XML utilizzato per la trasformazione di documenti XML in altri documenti XML. Il documento originale non viene modificato, ma piuttosto, ne viene creato uno nuovo partendo dal contenuto di quello esistente.



ICCU

Istituto centrale per il catalogo unico
delle biblioteche italiane e per le informazioni bibliografiche
<http://www.iccu.sbn.it>

Copyright © ICCU - Roma

Finito di stampare nel mese di giugno 2014
a cura dell'Istituto Poligrafico e Zecca dello Stato S.p.A. - Roma