

Associazione italiana biblioteche  
Gruppo di studio AIB Catalogazione,  
indicizzazione, linked open data e web semantico  
(CILW)

**Progressi dell'informazione e  
progresso delle conoscenze  
Granularità, interoperabilità e  
integrazione dei dati**

a cura di Roberto Raieli

Prefazione di Vittorio Ponzani

Testi di

Maria Teresa Biagetti, Enzo Borsellino, Loredana Cerrone,  
Elena Corradini, Laura Crociani, Patrizia De Martino,  
Camilla Fusetti, Mauro Guerrini, Antonella Iacono,  
Klaus Kempf, Luca Martinelli, Giovanni Michetti,  
Roberto Raieli, Elisabetta Viti, Carlotta Vivacqua,  
Maurizio Vivarelli, Paul Gabriele Weston

Appendice a cura del  
Gruppo di lavoro LOD - Sezione AIB Piemonte

Roma  
Associazione italiana biblioteche  
2017

Per la valutazione *ex ante* delle pubblicazioni monografiche l'Associazione italiana biblioteche ricorre a due esperti del settore, di cui almeno uno individuato all'esterno del Comitato scientifico.

Il testo viene riesaminato da almeno uno dei due esperti dopo la revisione richiesta agli autori.

Il Comitato scientifico è composto da Giovanni Di Domenico, Anna Galluzzi, Alberto Petrucciani.

Alcuni capitoli del volume sono basati sulle relazioni tenute nel corso del "Convegno AIB CILW 2016. La rinascita delle risorse dell'informazione: granularità, interoperabilità e integrazione dei dati", svoltosi il 21 ottobre 2016 presso la Biblioteca nazionale centrale di Roma: <<http://www.aib.it/attivita/congressi/c2016/giornata-studi-aib-cilw-2016/>>.

Il convegno è stato organizzato dal Gruppo di studio AIB Catalogazione, indicizzazione, linked open data e web semantico (CILW), dalla Sezione AIB Lazio e dalla Biblioteca nazionale centrale di Roma.

Il Gruppo di studio AIB CILW è composto da Marinella Cisternino, Elena Corradini, Antonella Iacono, Roberto Morelato, Roberto Raieli e Antonella Trombone.

Editing Palmira M. Barbini

Immagine di copertina: *Linking Open Data cloud diagram 2017*, di Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentzsch e Richard Cyganiak: <<http://lod-cloud.net/>>. Disponibile con licenza CC-BY-SA.



Quest'opera è stata rilasciata con licenza **Creative Commons Attribuzione – Non commerciale – Condividi allo stesso modo**.

Per leggere una copia della licenza visita il sito web <<http://creativecommons.org/licenses/by-nc-sa/4.0/>> o spedisce una lettera a Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

© 2017 Associazione italiana biblioteche  
Produzione e diffusione: Associazione italiana biblioteche  
Viale Castro Pretorio 105 - 00185 Roma  
Tel. 064463532, fax 064441139  
e-mail [aib@aib.it](mailto:aib@aib.it), <http://www.aib.it>  
ISBN 978-88-7812-254-3

## Indice

### **Prefazione**

Vittorio Ponzani 7

### **Premessa**

Roberto Raieli,  
*La rinascita delle risorse della conoscenza?* 11

### **Prima parte**

#### **Dai cataloghi alla navigazione semantica: storia, teoria e pratica delle strade condivise**

Roberto Raieli, *La comunità della conoscenza: senso e possibilità del progetto LOD-semantic web* 37

Maurizio Vivarelli, *Vedere le informazioni: dati, persone, mediazione documentaria* 77

Paul Gabriele Weston - Camilla Fusetti,  
*Le risorse bibliografiche verso il web semantico* 93

Maria Teresa Biagetti, *Interoperabilità semantica tra fonti eterogenee del patrimonio culturale: ontologie e linked open data* 123

Mauro Guerrini - Carlotta Vivacqua, *Linked data nei progetti delle biblioteche europee* 139

Klaus Kempf, *Oltre il catalogo: creazione di metadati standard e linked open data nel mondo di lingua tedesca* 161

### **Seconda parte**

#### **Linked open data e semantic web: progetti ed esperienze di convergenza**

Elena Corradini, *I linked open data e i servizi di reference nel mondo digitale* 175

Antonella Iacono, <i>Utenti, cataloghi e linked data: interfacce e nuove funzioni dell'utente nel web semantico</i>	201
Laura Crociani - Elisabetta Viti, <i>Nuovo soggettario e WebDewey italiana: strategie di mapping e integrazione</i>	223
Loredana Cerrone - Patrizia De Martino, <i>Biblioteche accademiche: linked open data e integrazione nel web con SHARE Catalogue</i>	237
Giovanni Michetti, <i>Linked data nel dominio archivistico: rischi e opportunità</i>	255
Enzo Borsellino, <i>Musei e MAB dopo le 'riforme' Franceschini</i>	279
Luca Martinelli, <i>Wikidata: la soluzione wikimediana ai linked open data</i>	287

## **Appendice**

Gruppo di lavoro AIB Piemonte sull'integrazione e innovazione dei servizi bibliotecari - sottogruppo LOD, <i>Pillole di LOD: brevi suggestioni senza pretesa di esaustività su web semantico e linked open data</i>	303
---	-----

<b>Biografie degli autori</b>	321
-------------------------------	-----

# Linked data nei progetti delle biblioteche europee

Mauro Guerrini e Carlotta Vivacqua\*

## Abstract

Numerose biblioteche europee hanno avviato, fin dai primi anni del Duemila, sperimentazioni ed esperienze con le tecnologie dei linked data per la pubblicazione e il miglior utilizzo dei propri dati sul web; alcune di esse espongono già risultati molto avanzati. In molti casi i dati non sono strutturati nativamente in RDF, cioè secondo il modello definito per i linked data, bensì sono convertiti da formati usati in precedenza. In questo capitolo si analizzano i casi data.bnf.fr della Bibliothèque nationale de France (BnF), The British National Bibliography as linked open data della British Library, Linked data service della Deutsche Nationalbibliothek (DNB) e Datos.bne.es della Biblioteca Nacional de España.

## *Linked data in European libraries projects*

Many European libraries have started since the early years of the millennium experiments and experiences with the linked data technologies for publishing and the best use of their own data on the web; some of them already show very advanced results. In many cases the data are not natively structured in RDF, i.e. according to the model defined for linked data, but they are converted from previously used formats. In this chapter we analyze the cases data.bnf.fr belonging to the Bibliothèque nationale de France (BnF), The British National Bibliography as linked open data belonging to the British Library, Linked data service belonging to the Deutsche Nationalbibliothek (DNB), and Datos.bne.es belonging to the Biblioteca Nacional de España.

## Premessa

Numerose biblioteche europee hanno avviato, fin dai primi anni del Duemila, sperimentazioni ed esperienze con le tecnologie dei linked data per la pubblicazione e il miglior

\* Mauro Guerrini, Università di Firenze, mauro.guerrini@unifi.it. Carlotta Vivacqua, Biblioteca nazionale centrale di Firenze, carlotta-v@hotmail.it. Ultima consultazione siti web: 13 dicembre 2016.

utilizzo dei propri dati sul web; alcune di esse espongono già risultati molto avanzati. In molti casi i dati non sono strutturati nativamente in RDF, cioè secondo il modello definito per i linked data, bensì sono convertiti da formati usati in precedenza.

Le iniziative europee presentano caratteristiche comuni:

- il modello FRBR sottende tutti i progetti: la documentazione fa sempre riferimento alle sue entità e all'utilizzo di un'ontologia nata dal modello;
- la struttura sfrutta al massimo i dati esistenti, collegandoli ad altri dataset per l'arricchimento informativo, realizzato tramite l'*interlinking*;
- la filosofia è la medesima, il paradigma dei linked data e del web semantico: ciò che differenzia i progetti è la modalità di presentazione dei dati, che cambia in funzione di una loro maggiore chiarezza di lettura da parte degli utenti;
- i progetti richiedono importanti investimenti di risorse economiche e umane.

Analizziamo i progetti principali, partendo dal francese, il primo realizzato e a cui gli altri hanno fatto riferimento.

## **Data.bnf.fr**

Il progetto data.bnf.fr della Bibliothèque nationale de France (BnF)<sup>1</sup> utilizza i linked data e applica gli standard del web semantico per la pubblicazione dei cataloghi della biblioteca e i dati della biblioteca digitale Gallica. Data.bnf.fr aggrega informazioni provenienti da cataloghi diversi. Il progetto è guidato dal desiderio di sperimentare nuove possibilità tecniche per favorire le ricerche da parte degli utenti, di cui sono stati analizzati i comportamenti nella formulazione

<sup>1</sup> Weinz Romain, *Linked open data for new library services: the example of data.bnf.fr*, «JLIS.it», 4 (2013), n. 1, p. 403-415, <<http://tinyurl.com/y73vvqa3>>.

delle *query*. Data.bnf.fr è stato, pertanto, concepito secondo la ‘metodologia agile’ (*agile software development*), tecnica che permette di strutturare le evoluzioni dell’applicazione sulla base dei variegati e sempre nuovi bisogni dei lettori.

*Des fiches de référence sur les auteurs, les oeuvres et les themes*, titolo per esteso del progetto, da cui è assente il termine ‘catalogo’, fornisce informazioni relative ad autori, opere e soggetti. A ottobre 2016 sono presenti oltre 2 milioni di autori e quasi 200 milioni di autori collegati a oltre 8 milioni di risorse provenienti dal catalogo generale della BnF. La Bibliothèque nationale de France provvede a una capillare diffusione dei dati tramite specifici servizi e tecnologie; rende, inoltre, i dati disponibili al riuso in modalità aperta e in formato RDF, segnalandoli su data.gouv.fr, portale dei dati pubblici dello Stato francese, col solo vincolo di citazione delle fonti. Nonostante la forte simbiosi con Gallica, data.bnf.fr ha una vita propria, il cui scopo è coprire l’intera consistenza dei cataloghi e delle risorse di BnF e Gallica. Il punto cruciale è esporre la maggior parte delle risorse digitali e delle risorse analogiche. La BnF ha implementato data.bnf.fr quale interfaccia web per riunire dati e collegamenti, ed esprimere le informazioni bibliografiche in modo ricco, altamente strutturato secondo standard condivisi a livello internazionale.

Per essere conforme al paradigma del linked open data, il progetto fa propri alcuni assunti per la descrizione delle risorse:

- mantenere URI permanenti, strutturati in modo che siano comprensibili agli utenti;
- realizzare un efficiente meccanismo di negoziazione dei contenuti;
- utilizzare un modello di dati RDF compatibile, con vocabolari standard (SKOS, RDA e FOAF);
- utilizzare vocabolari esistenti, se possibile;

- utilizzare vocabolari locali, creati *ad hoc*, solo per le classi e gli oggetti che sono specifici della BnF e che non troverebbero corrispondenza con i vocabolari esistenti;
- allineare i dati con dati provenienti da database esterni, dalla Library of Congress alla Deutsche Nationalbibliothek, a Geonames ecc.

Il dataset RDF, periodicamente aggiornato, è accessibile e scaricabile dalla home page del sistema. Il *data model* del progetto esprime le scelte compiute, le ontologie utilizzate, tra cui RDA (soprattutto per le relazioni), FOAF (per le entità del tipo autore), DC (per gli elementi descrittivi) e i collegamenti a progetti e dataset esterni.

### *L'impianto teorico e tecnico di data.bnf.fr*

L'applicazione è stata progettata per essere utilizzabile da browser basati sull'uomo che naviga tramite le pagine di un sito web. Essa genera pagine fornendo informazioni standardizzate, riferimenti e collegamenti su autori, opere e soggetti. Le pagine generate sono di facile reperibilità e comprensione: di un autore, per esempio, l'utente trova i link alle pagine web delle opere rilevanti, da e sull'autore, in due sezioni separate; per un'opera trova un link alla pagina del suo autore e quelli alle altre manifestazioni dell'opera. Data.bnf.fr include le fasi di trasformazione dei dati originali in RDF, il loro arricchimento, il collegamento a risorse interne ed esterne tramite *interlinking*, la pubblicazione in pagine HTML per la navigazione da parte degli utenti e dei motori di ricerca. Il progetto raggruppa ed espone online in formato RDF dati provenienti da fonti eterogenee; i dati possono essere facilmente indicizzati dai motori di ricerca e collegati ad altre risorse, sia interne alla BnF (BnF catalogue général, BnF archives et manuscrits, Gallica ecc.) sia esterne (Catalogue collectif de France, Catalogue du Système Universitaire de Documentation Sudoc, WorldCat, VIAF, Wikipedia).



Il processo di lavorazione dei dati si articola in diversi passi: dalla loro raccolta (*harvesting*) da dataset diversi, all'arricchimento e al collegamento con altre risorse interne ed esterne a BnF, fino alla successiva pubblicazione di pagine in HTML. Il processo richiede la trasformazione di metadati formattati in MARC o EAD con tecniche di modellazione in RDF. La modellazione ha un collegamento diretto con l'allineamento e arricchimento dei dati che devono essere estratti e processati. I contenuti e i collegamenti sono riuniti in conformità ai concetti basati sulle entità del Gruppo 1, 2 e 3 di FRBR. Quindi i dati sono integrati all'interno di un'architettura che prevede sia la pubblicazione di pagine HTML, sia la visualizzazione dei dati grezzi in RDF. Questo sforzo elaborativo e di applicazione ha significato confrontarsi con standard web comuni e non biblioteca-centrici, tesi alla costruzione di un'interoperabilità con fonti esterne, in modo da consentire l'allineamento delle risorse provenienti da cataloghi diversi in altrettanti formati. Il servizio presenta un modello *semantic web friendly* che comprende utenti, ricercatori, sviluppatori e comunità non tradizionalmente incluse nell'utenza della BnF.

### *Identificatori permanenti*

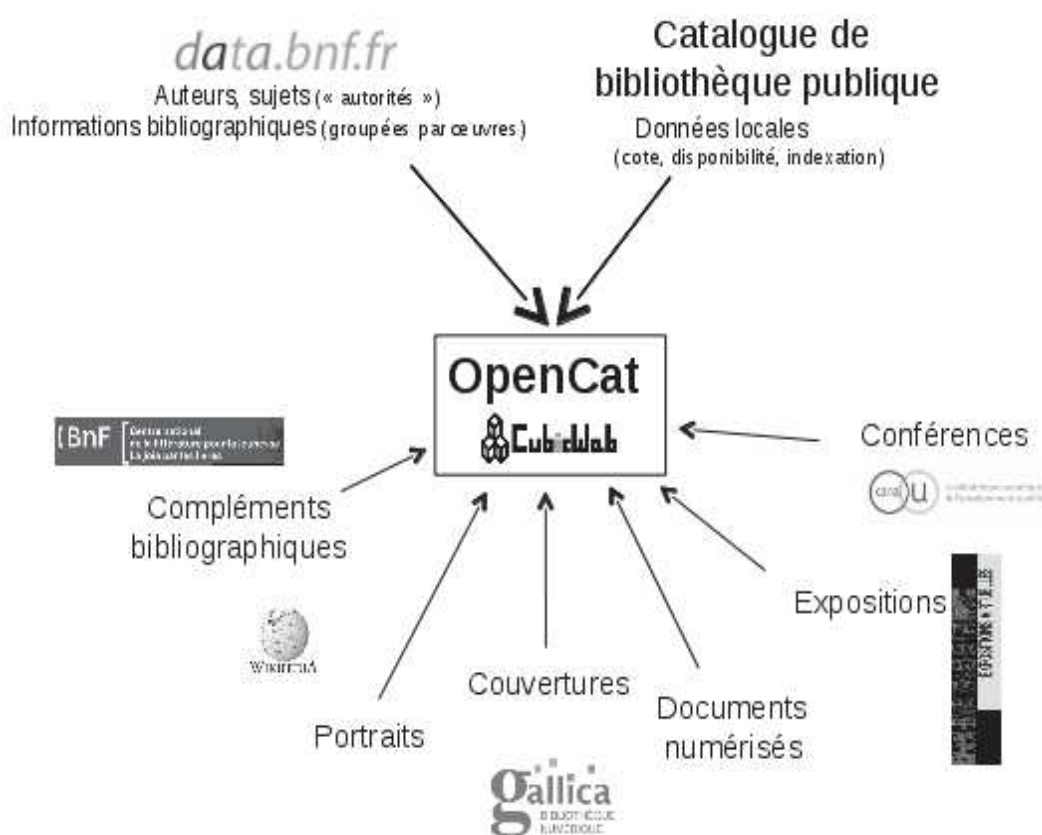
Nel 2013 la BnF ha deciso di avviare il progetto OpenCat<sup>2</sup>, insieme alla Biblioteca comunale di Fresnes<sup>3</sup> e alla società di software Logilab, specializzata in tecnologie LOD, appaltatrice di BnF per lo sviluppo di data.bnf.fr. Sul sito web di Logilab è disponibile una sintesi del lavoro compiuto:

Il progetto OpenCat ha permesso di prefigurare ciò che saranno, tra pochi anni, i cataloghi delle biblioteche che riutilizzano i dati disponibili sul web e in particolare su data.bnf.fr. Supportato dal Ministero della Cultura e sviluppato in partnership con la Bibliothèque nationale de France e la Biblioteca comunale di Fresnes, l'esperimento ha dato prova

<sup>2</sup> <<https://tinyurl.com/ybdwtwfb>>.

<sup>3</sup> <<http://bm.fresnes94.fr/>>.

del modo di creare il proprio catalogo depositando un elenco di libri. Gli elementi dell'elenco vengono impostati automaticamente in relazione alle risorse di data.bnf.fr. Si ottiene così un catalogo navigabile online che utilizza identificatori univoci e perenni (identificatori ARK di BnF), con record di qualità che descrivono opere, autori e soggetti (modellazione FRBR) e che include collegamenti a risorse esterne (Wikipedia, Gallica, etc.). In futuro, numerosi fornitori di sistemi per la gestione delle biblioteche incorporeranno funzioni di questo tipo, che consentiranno di ridurre i costi di gestione dei cataloghi, approfittando delle risorse disponibili sul web<sup>4</sup>.



**Figura 1** – Confluenze in OpenCat

Dalle pagine BnF relative alla sperimentazione si apprende che il prototipo – costruito con le stesse tecniche di data.bnf.fr – include:

- raggruppamenti di opere (FRBR) realizzati per data.bnf.fr;

<sup>4</sup> <<https://www.logilab.fr/blogentry/1978>>.

- dati locali della biblioteca di Fresnes, corrispondenti al *corpus* di dati comuni con data.bnf.fr (circa 4.000 opere);
- aggiunte bibliografiche assenti dai due database (raccomandazioni e avvisi del Centre national de littérature pour la jeunesse);
- informazioni contestuali non bibliografiche (illustrazioni di copertina, ritratti di autori, informazioni biografiche);
- link alle risorse online (testi, suoni, video di Gallica, mostre virtuali della BnF, conferenze scientifiche di Canal-U, la webTV dell'insegnamento superiore).

L'utente ottiene risultati di ricerca provenienti da BnF, arricchiti da informazioni provenienti da altre fonti, rimanendo nella medesima interfaccia di ricerca; ciascun risultato restituito ha gli identificatori permanenti BnF, che fungono da perno per la navigazione verso altre risorse.

Dalla sperimentazione OpenCat è emersa l'urgenza di una soluzione definitiva per gli identificatori permanenti di data.bnf.fr, problematica tipica della tecnologia LOD e dell'ambiente del web semantico. La soluzione al problema è il documento *Les identifiants ARK*<sup>5</sup>, disponibile nella sezione About data.bnf.fr, alla voce FRBR data model, del sito.

### *Identificatori attuali: proposte e necessità della BnF*

Scopo degli identificatori è garantire la 'citabilità', ovvero la capacità di denominare e di individuare la risorsa in modo certo. Negli identificatori univoci si uniscono altri due elementi fondamentali: l'unicità e la 'persistenza'. La citabilità dipende dalla scelta degli identificatori semantici o significativi, cioè portatori di significato. Si tratta di identificatori ampiamente adottati sul web, ma che possono comportare problemi: potrebbero mancare di stabilità col mutare della natura della risorsa, potrebbe essere 'rotta' la relazione semantica tra l'identificatore e la risorsa. Inoltre, nel passaggio da una lingua a un'altra, o nel corso del tempo, può

<sup>5</sup> <[http://www.bnf.fr/documents/ark\\_presentation\\_bermes\\_2006.pdf](http://www.bnf.fr/documents/ark_presentation_bermes_2006.pdf)>.

cambiare il significato di una parola o di un acronimo, può essere esplicito per un periodo e divenire incomprensibile in un altro. BnF reputa importante riflettere su contingenze simili e considerare la scelta di un sistema di identificazione opaco. Gli identificatori opachi vengono generati da macchine tramite l'uso di software, ma anch'essi possono presentare difficoltà nel mantenere un collegamento tra l'identificatore e la risorsa. Essi sono generati artificialmente, quindi possono essere molto lunghi, pertanto impossibili da usare per gli utenti umani.

L'analisi dei vantaggi e degli svantaggi degli identificatori semantici e degli identificatori opachi richiama la premessa relativa alla difficoltà di scelta: se gli identificatori significativi facilitano la citabilità, e quindi la disponibilità delle risorse, quelli opachi garantiscono la persistenza e sembrerebbero preferibili in un contesto in cui la conservazione delle risorse è ipotizzata a lungo termine. Si pone il dilemma tra conservazione e comunicazione dei dati, tra archiviazione e indicizzazione. La soluzione potrebbe prevedere più sistemi identificativi.

Un'autorità di denominazione riconosciuta a livello mondiale offre l'indipendenza necessaria per assegnare identificatori di risorse persistenti, unici e adattabili, che i browser interpretano direttamente e mediante un *resolver*. Confrontare e valutare i diversi sistemi è un'attività complessa, poiché nessuno garantisce valori assoluti in termini di persistenza.

Il grado di standardizzazione è un elemento chiave dell'unicità degli identificatori e della loro persistenza. La scelta di un sistema di identificatori persistenti nella BnF è stata compiuta, dando priorità sia alle esigenze di visibilità delle proprie risorse web e di costruzione di servizi associati, sia all'attuazione del sistema di gestione delle risorse digitali coerente in un repository OAIS compatibile chiamato SI digitale, pensato per la loro conservazione a lungo termine. Il sistema dovrebbe garantire indipendenza dal

punto di vista economico e tecnico, ovvero in termini di integrabilità nell'architettura esistente, in caso di abbandono.

Con questi vincoli, i responsabili della BnF hanno esplorato varie soluzioni:

- un *naming* OAI (Open Archives Initiative)<sup>6</sup>, associato a un sistema di *resolver* come POI (PURL-Based Object Identifier)<sup>7</sup>; OAI non ha, però, soddisfatto integralmente le varie esigenze di un'ampia accessibilità, citabilità e indicizzazione delle risorse;
- la creazione di un sistema simile a quello della NLA (National Library of Australia), sulla base di un dominio con prospettiva di compatibilità coi sistemi internazionali. La soluzione avrebbe comportato un lavoro molto complesso.

### *La soluzione*

La scelta BnF è caduta sul sistema di identificatori persistenti ARK (Archival Resource Key)<sup>8</sup>, creato e mantenuto dalla California Digital Library<sup>9</sup>. Gli elementi costitutivi di questo sistema sono due: le autorità di indirizzamento NMAH (Name Mapping Authority Hostports)<sup>10</sup> e i qualificatori.

Il NMAH è l'indirizzo del server che rende trasparente l'identificatore ARK. È possibile disporre di più autorità di indirizzamento per lo stesso identificatore, permettendo, nel caso di BnF, di rispettare il contesto di visualizzazione delle risorse: il catalogo generale e Gallica. Il processo permette di mantenere lo stesso nome di dominio tra l'applicazione d'accesso (il catalogo o Gallica) e la risorsa stessa. I motori di ricerca indicizzano correttamente le risorse nel campo

<sup>6</sup> <<http://www.openarchives.org/>>.

<sup>7</sup> <<https://tinyurl.com/yc269uaa>>.

<sup>8</sup> <<http://www.cdlib.org/inside/diglib/ark/arkspec.pdf>>.

<sup>9</sup> <<http://www.cdlib.org/>>.

<sup>10</sup> <<http://www.cdlib.org/services/uc3/arkspec.pdf>>.

Gallica. L'autorità di denominazione è libera di sviluppare una gerarchia e rendere visibile le varianti, sviluppando così la granularità di una risorsa, facendo riferimento, per esempio, alle sue diverse versioni. Il qualificatore non è, tuttavia, coperto da garanzia di persistenza. Ciò richiama la questione degli identificatori astratti e concreti: i primi sono persistenti e puntano verso un oggetto logico, mentre i secondi designano, o possono designare, una parte sempre identica di un oggetto logico.

Un esempio: l'identificatore

ARK:/12148/bpt6k85329c

indica l'oggetto logico che corrisponde alla digitalizzazione di 'Wheler, George. Voyage de Dalmatie, de Grèce et du Levant', mentre l'identificatore

<http://catalogue.bnf.fr/ark:/12148/bpt6k85329c/f4.pagination>

fa riferimento all'oggetto fisico che corrisponde alla quarta immagine di questo oggetto logico, nel contesto del catalogo e col servizio di invio della paginazione. Di questi due ID, solo il primo è persistente; il secondo contiene informazioni difficili da rendere persistenti, tra cui informazioni semantiche ('catalogo', 'paginazione') che hanno il solo scopo di rendere un servizio.

Gli oggetti digitali che integrano SI digitale della BnF sono diversificati e sono presentati con i relativi ID (ISSN, ISBN, gli URI, i DOI ecc.).

Per gli accessi – la problematica più vicina agli utenti – utilizzare identificatori ARK significa garantire granularità a livello di oggetto e migliorare la gestione dei dati.

Le Portrait de Dorian Gray / par Oscar Wilde ; traduction d'Eugène Tardieu ; bois gravés de Fernand Siméon - 1920  
 Description [ The picture of Dorian Gray (français)]  
 Auteur Wilde, Oscar ( 1854-1900 ). Auteur du texte

**i** Informations détaillées

Notice complète

Titre : Le Portrait de Dorian Gray / par Oscar Wilde ; traduction d'Eugène Tardieu ; bois gravés de Fernand Siméon  
 Auteur : Wilde, Oscar ( 1854-1900 ). Auteur du texte [Ne voir que les résultats de cet auteur](#)  
 Éditeur : Mornay (Paris)  
 Date d'édition : 1920  
 Contributeur : Tardieu, Eugène (18...-19... ; angliciste). Traducteur  
 Contributeur : Siméon, Fernand (1884-1928). Graveur  
 Type : text  
 Type : monographie imprimée  
 Type : printed monograph  
 Langue : Français  
 Langue : français  
 Format : 373 p. : fig., planche ; in-8°  
 Format : application/pdf  
 Format : Nombre total de vues : 397  
 Description : [ The picture of Dorian Gray (français)]  
 Description : Collection : Les Beaux livres ; 2  
 Droits : domaine public  
 Droits : public domain  
 Identifiant : **ark:/12148/bpt6k314017x**  
 Source : Bibliothèque nationale de France, département Réserve des livres rares, RES P-Y2-1535  
 Relation : <http://catalogue.bnf.fr/ark:/12148/bpt6k314017x>  
 Provenance : Bibliothèque nationale de France  
 Date de mise en ligne : 16/05/2016

**Figura 2** – Esempio di visualizzazione dell’identificatore ARK all’interno di una notizia bibliografica in Gallica

### *Modellazione in RDF*

La modellazione dei dati in RDF nel progetto francese si dirige verso un allineamento alle finalità del web semantico: produrre dati utili, affidabili, controllati e integrati nella rete. La soluzione tecnica che anche la Francia realizza per il proprio ingresso nel web semantico è la creazione di collegamenti tramite URI dereferenziabili e permanenti, e la creazione di un meccanismo di negoziazione dei contenuti e dell’accesso ai dati grezzi. Il risultato è disponibile sul sito web in diverse sintassi RDF, quali: RDF/XML, RDF-N3, RDF-NT, JSON, JSON-LD. La BnF fornisce gli URI per tutte le risorse, insieme a identificatori permanenti (ARK), così da permettere di scoprire le risorse. Un *endpoint* SPARQL permette di visualizzare e reperire i dati in RDF o in altri formati.

La modellizzazione ha previsto uno step di allineamento *clustering* (a grappolo) per opera, in base al quale nelle pagine di ‘opera’ e ‘autore’, tutte le manifestazioni di una singola opera sono riunite sotto lo stesso autore, grazie al col-





### *Sfide future*

Data.bnf.fr lancia sfide a lungo termine e ad ampio spettro: in seguito a studi statistici, si prevede che nei prossimi dieci anni le biblioteche pubbliche francesi utilizzeranno le tecnologie del web semantico per collegare i propri cataloghi locali con i metadati prodotti dalla BnF in quanto agenzia bibliografica nazionale di riferimento.

Il presupposto tecnico è che i metadati non dovranno più essere estratti da server a server della BnF verso biblioteche locali; piuttosto i riferimenti delle collezioni delle biblioteche locali potrebbero essere messi a disposizione da BnF in un centro di metadati costruito sulla base delle tecnologie open source sviluppate per data.bnf.fr. Le biblioteche locali potrebbero, quindi, linkare le proprie informazioni bibliografiche ai metadati della BnF dai loro sistemi locali. La catalogazione locale tradizionalmente intesa potrebbe, pertanto, ridursi al minimo, col vantaggio di consentire ai bibliotecari di concentrarsi sul collegamento delle risorse e sull'arricchimento dei metadati con informazioni locali, piuttosto che provvedere alla descrizione diretta delle risorse.

### *Policy di licenza*

Nel 2011, per rendere i dati aperti tecnicamente e legalmente, la BnF ha adottato la licenza pubblica raccomandata dalla missione *open data* del governo francese (Etalab)<sup>11</sup>, paragonabile a una licenza di tipo CC-BY. Essa è gratuita, aperta, incoraggia il riutilizzo e la riproduzione, compresi gli scopi commerciali. La licenza richiede l'attribuzione ed è compatibile con altre licenze internazionali, quali l'*Open government licence* del Regno Unito.

<sup>11</sup> <<https://www.etalab.gouv.fr/>>.

## **Bnb.data.bl.uk**

Il progetto [bnb.data.bl.uk](http://bnb.data.bl.uk) della British Library pubblica una parte della BNB (*British National Bibliography*)<sup>12</sup> in linked open data; include risorse monografiche e seriali per un totale di circa 3.1 milioni di record. BNB come linked open data è nata nel 2010, è stata lanciata nel 2011 su una piattaforma Talis<sup>13</sup> che ha offerto una gamma di opzioni per l'interrogazione dei dati, tra cui un *endpoint* SPARQL. Nel primo anno di attività il numero di visite all'*endpoint* SPARQL ha raggiunto la quota di 9.000.000 dalle 38.000 del primo mese. La discussione aperta in conseguenza dei numeri registrati si è conclusa poi nel 2012 quando Talis ha annunciato il ritiro dal business del web semantico per il lento sviluppo del mercato rispetto alle previsioni.

Nel 2013 il progetto è stato trasferito su una nuova piattaforma, ospitata da TSO<sup>14</sup>. La BNB viene aggiornata mensilmente sulla piattaforma e sulla pagina Download. Il dataset in RDF, disponibile sotto licenza Creative Commons CC0 1.0 Universal Public Domain Dedication, è consultabile con un *editor* SPARQL e uno SPARQL *endpoint*. È possibile, inoltre, lanciare ricerche utilizzando la stessa interfaccia del progetto.

### *Struttura e sezioni del progetto*

La sezione più importante per comprendere l'impianto del progetto è la terza: British Library open data pages; essa

<sup>12</sup> <<https://data.blog.gov.uk/2014/04/22/the-linked-open-british-national-bibliography/>>.

<sup>13</sup> Talis è una società di software inglese che ha aperto la strada del web semantico e della tecnologia dei linked data. Nata da un'unione di biblioteche a Birmingham, nel 2011 ha realizzato la prima piattaforma del progetto BNB linked open data; la società ha venduto la sua divisione software di gestione delle biblioteche al colosso dei servizi Capita, concentrandosi sulla tecnologia del web semantico e sulle possibilità di marcare dati sul web con metadati che descrivono il loro significato.

<sup>14</sup> <<https://tinyurl.com/yccy7hs2>>.

contiene informazioni per la comprensione delle varie sezioni e presenta i link ai due documenti fondativi del progetto e della sua mission; ovvero, presenta all'utente gli strumenti per navigare correttamente e consapevolmente. Il primo dei due mostra il ragionamento sotteso al progetto di BNB come linked open data che si è reso necessario col modificarsi dei contesti internazionali in cui la British Library è inserita. Evocativo appare il titolo dato alla presentazione: *Making the connection, linked open data at the British Library*. Il secondo documento, *Publishing the British National Bibliography as linked open data*, spiega dettagliatamente il progetto, insieme agli spigoli da smussare e alle differenze di opinione riscontrate. Il documento evidenzia in più punti la volontà di aprire le porte della British Library a comunità diverse, tradizionalmente non contemplate tra gli utenti.

### *Procedure di realizzazione*

La realizzazione del progetto è avvenuta per passaggi gradualmente; il personale coinvolto ha sviluppato competenze in standard e strumenti bibliografici per gestire e manipolare grandi volumi di dati, prevalentemente in MARC, e acquisito esperienze di HTML, XML e XSLT, senza che nel team venissero inclusi programmatori o architetti di dati. La formazione specifica è stata assicurata da Talis<sup>15</sup>. Grazie a questa sinergia sono stati sviluppati nuovi strumenti, per esempio, per generare URI o collegare dati a set di dati esterni. Sono state identificate le entità, quindi assegnati gli URI, scegliendo di coniarne di propri piuttosto che rivolgersi a fonti esterne. Come per data.bnf.fr, si è posto il dilemma sull'opportunità di usare URI opachi o trasparenti, cioè leggibili per l'essere umano.

Il passo successivo nel processo ha coinvolto la descrizione delle entità e di come si relazionano tra di loro. Il team

<sup>15</sup> <<http://support.talis.com/hc/en-us>>.

inglese ha optato per un utilizzo di classi e oggetti di vocabolari RDF esistenti, verificando le ontologie utilizzate in altri progetti LOD.

La policy della BNB intende raggiungere un pubblico ampio che va oltre quello abituale delle biblioteche; per questo motivo il progetto non ha sfruttato l'ontologia FRBR, considerata complessa da applicare dagli sviluppatori Talis (scelta opposta alla BnF), mentre si è allineata ai linguaggi degli altri progetti: DCMI Metadata Terms, FOAF, SKOS; ha, tuttavia, utilizzato schemi biblioteconomici come ISBD elements, Bibo, RDA Group 2 Elements, Bio. La British Library ha creato classi e proprietà specifiche per facilitare la ricerca e la navigazione da una risorsa all'altra; per esempio, le classi BLT: TopicLCSH e BLT: TopicDDC, che consentono agli utenti di compiere una ricerca raffinata basata su una voce presente in LCSH o su una notazione DDC. La biblioteca ha analizzato i dati bibliografici e li ha convertiti, ponendo al centro del modello la risorsa, identificata da un URI, da cui partono collegamenti con entità (istanze e classi). Il dataset è arricchito da link esterni a diversi dataset, tra cui GeoNames, MARC Country code, VIAF, Lexvo, LCSH, ID.loc.gov, Dewey. Gli URI coniatati dalla British Library sono stati generati prima nei record MARC e successivamente collegati ai vari dataset esterni come VIAF, con l'aggiunta dei collegamenti LCSH. Il processo è stato realizzato con la *suite* Catalogue utility bridge.

### *Prossimi sviluppi*

Nell'immediato futuro gli obiettivi riguardano lo sviluppo dell'offerta legata ai dati *open*; è necessario estendere e raffinare il modello iniziale, arricchire BNB linked open data con collegamenti ad altre risorse, per esempio, ISNI, LC/NACO e DBpedia. È in programma la valutazione di collegamenti di maggiore granularità relativi ai luoghi, rendendo possibile una mappatura esatta tra Geonames e i luoghi di pubblicazione. La prospettiva è valutabile solo in con-

sequenza di una disambiguazione dei dati MARC. Sviluppo di più corto raggio è invece costituito dal collegamento di BNB con altre bibliografie nazionali, in particolare con la Deutsche Nationalbibliothek.

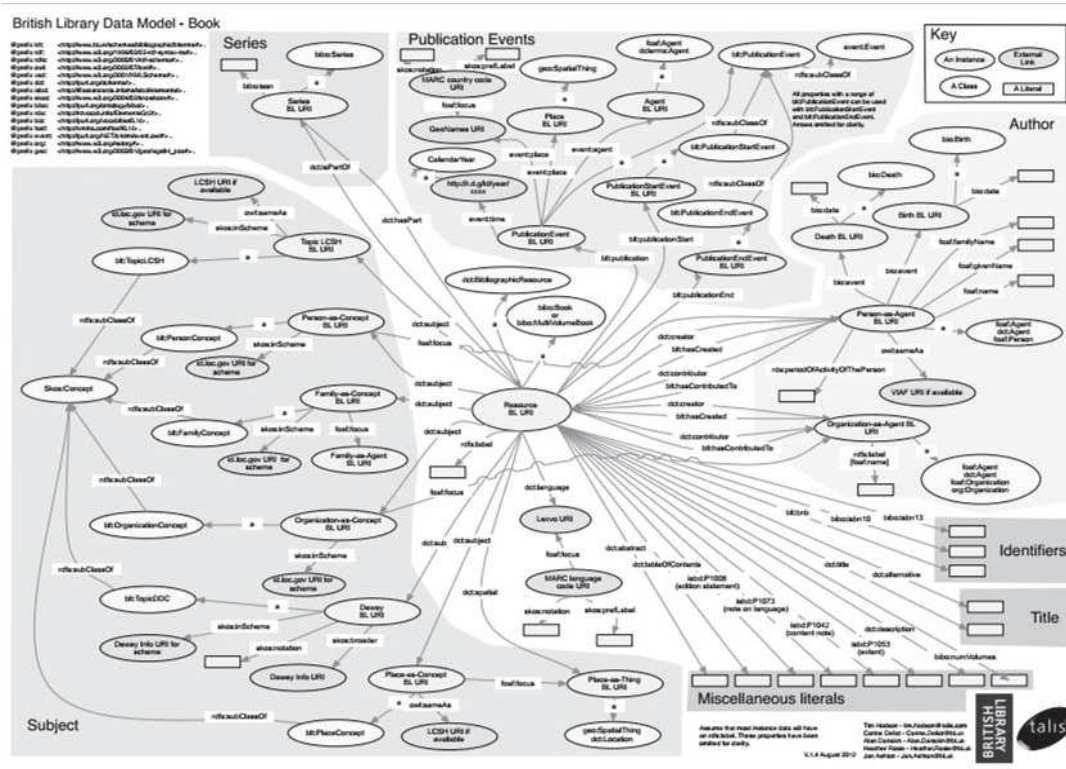


Figura 4 – Data model della sezione libri del progetto bnb.data.bl.uk

## Il linked data service della Deutsche Nationalbibliothek

La Deutsche Nationalbibliothek (DNB) ha avviato nel 2010 il progetto di conversione e pubblicazione dei propri record d'autorità in linked open data<sup>16</sup>; nel 2012 ha esteso il progetto ai record bibliografici. A gennaio 2015 sono pubblicati in RDF, sotto licenza CC0 (Creative Commons Zero), i dati delle principali collezioni della biblioteca (eccettuata la musica a stampa e la ExilSammlungen) e dei seriali della Zeitschriftendatenbank (ZDB). I record bibliografici e

<sup>16</sup> <<http://www.dnb.de/EN/lids>>.

d'autorità sono stati analizzati nei loro formati originari per individuare un sottoinsieme di elementi necessari alla identificazione delle risorse. L'analisi ha seguito le raccomandazioni<sup>17</sup> del DINI WG KIM Bibliographic Data Group<sup>18</sup>, costituito dai rappresentanti delle reti e delle grandi biblioteche di area linguistica tedesca. Il gruppo di lavoro ha definito criteri di armonizzazione delle rappresentazioni bibliografiche in RDF valide per i paesi di lingua tedesca. Nella modellazione dei dati sono stati utilizzati vocabolari e ontologie esistenti e ampiamente utilizzati in progetti analoghi, tra cui Dublin Core terms, Dublin Core elements 1.1, Bibliographic Ontology (BIBO), Resource Description and Access Vocabularies, ISBD elements, RDF Schema, Upper Mapping and Binding Exchange Layer (UMBEL)<sup>19</sup>, FOAF, Schema.org: Library extension terms<sup>20</sup>.

I metadati e le interfacce online sono forniti senza garanzia di esenzione da errori relativi alla violazione dei diritti (personali e d'autore) e da esaustività descrittiva. La conversione dei dati in RDF avviene tramite Metafacture<sup>21</sup>, sviluppato da Culturegraph.org. La rappresentazione RDF dei dati test (test in rappresentazione RDF/XML) è disponibile, gratuitamente, alla voce Aktionen nella visualizzazione completa del catalogo della DNB o nel Data shop. Interessanti sono la sezione relativa ai criteri con cui vengono assegnati gli URN e la sezione dedicata ad AREDO, un servizio di cooperazione della DNB per sostenere le biblioteche, gli istituti di istruzione, i musei, gli archivi e le istituzioni della memoria nella conservazione a lungo termine delle risorse. Alla pagina Data service si trovano elencati i dati disponibili in formato MARC21 e in formati leggibili tramite WWW o FTP. I record sono forniti in MARC21 e

<sup>17</sup> <<https://wiki.dnb.de/pages/viewpage.action?pageId=68060017>>.

<sup>18</sup> <<https://wiki.dnb.de/display/DINIAGKIM/Titeldaten+Gruppe>>.

<sup>19</sup> <<http://umbel.org/umbel#>>.

<sup>20</sup> <<http://purl.org/library/>>.

<sup>21</sup> <<https://github.com/culturegraph/metafacture-core/wiki>>.

MARC21XML, molti anche in RDF/XML, e in DNB casual (solo per i dati bibliografici).

Dal primo ottobre 2015 la biblioteca ha iniziato la catalogazione con RDA. L'influenza delle linee guida è evidente, per esempio, nell'attenzione rivolta agli elementi quali *content*, *media* e *carrier*, nella definizione degli attributi qualificanti e identificanti la risorsa. È possibile avere accesso ai linked open data in diverse modalità: tramite lo specifico URI della risorsa, come dataset RDF, tramite interfacce SRU (Search/Retrieve via URL) e OAI-PMH.

### *Panoramiche SRU e OAI*

Nella sezione del progetto dedicata alla panoramica SRU sono date spiegazioni esaustive su questo protocollo, utile per compiere ricerche mirate nei database. Le *query* vengono formulate in CQL (Common Query Language), mentre i risultati sono in formato XML. SRU è stato pubblicato dalla Library of Congress come sviluppo ulteriore di Z39.50. Seguono micro sezioni di spiegazione e schemi dei cataloghi tedeschi accessibili in questa modalità, e dei formati di disponibilità dei dati. Il protocollo per la comunicazione è l'OAI-PMH (OAI-Protocol for Metadata Harvesting)<sup>22</sup>, che consente la sincronizzazione di ingenti quantità di dati.

### *FAQ di dnb.de.dataservice*

Le FAQ sono particolarmente utili e forniscono un quadro esaustivo tramite cui muoversi con agilità all'interno del progetto. Il progetto tende a trasformare l'intero catalogo in linked data. In caso di dubbio o di richieste ulteriori dnb.de.dataservice rimanda a una mailing list tramite cui è possibile ricevere assistenza, leggere commenti e interagire. Viene, infine, data la possibilità di consultare la cronistoria del progetto con un link che rimanda alle precedenti versioni. La sezione termina con la documentazione sui linked

<sup>22</sup> <<http://oaiharvestmangr.sourceforge.net/>>.

data prodotti e sui dati disponibili in versione scaricabile (previo accesso alla pagina personale).

## **Datos.bne.es**

Il progetto di pubblicazione in modalità linked open dati della Biblioteca Nacional de España, ancora in versione *beta*, riguarda manoscritti, libri moderni e antichi, mappe, disegni, stampe e fotografie, spartiti, registrazioni audio e videocassette. Il modello rispecchia il francese data.bnf.fr, a partire dalle tre tipologie di entità trattate: Autori, Opere, Soggetti. Si tratta di uno sviluppo operato dalla Biblioteca e dall'Ontology engineering group (OEG)<sup>23</sup> dell'Universidad Politécnica de Madrid (UPM). I dati sono accessibili e scaricabili dal portale e sono editi sotto licenza Creative Commons Public Domain Dedication, aperta, gratuita e finalizzata alla promozione del loro riuso.

Riutilizzo e apertura sono parole d'ordine per l'intero settore pubblico spagnolo, che ha aderito al progetto di riuso delle informazioni mediante un decreto del 2011, che fa propria la proposta della Conference of European National Librarians (CENL).

### *Fasi e versioni*

Il progetto inizia nel 2010. La prima serie di dati collegati è stata ottenuta nell'aprile 2011 e un insieme rappresentativo nel dicembre successivo. Sulla base degli stessi dati, aggiornati al 2014, è stata costruita un'interfaccia di visualizzazione con un motore di ricerca che sfrutta le relazioni tra i dati e sperimenta nuove modalità di presentazione dei dati, compresi quelli di arricchimento da fonti esterne. Sono stati ampliati la gamma e il numero di record bibliografici e d'autorità. Il progetto presenta un link alle copie digitaliz-

<sup>23</sup> <<http://www.oeg-upm.net/>>.



zate della Biblioteca Digital Hispánica e dispone di un'ontologia interna che riutilizza le proprietà di diversi vocabolari. I dati provengono sia da cataloghi di biblioteche e liste d'autorità della BNE, sia da scansioni di opere della Biblioteca Digital Hispánica.

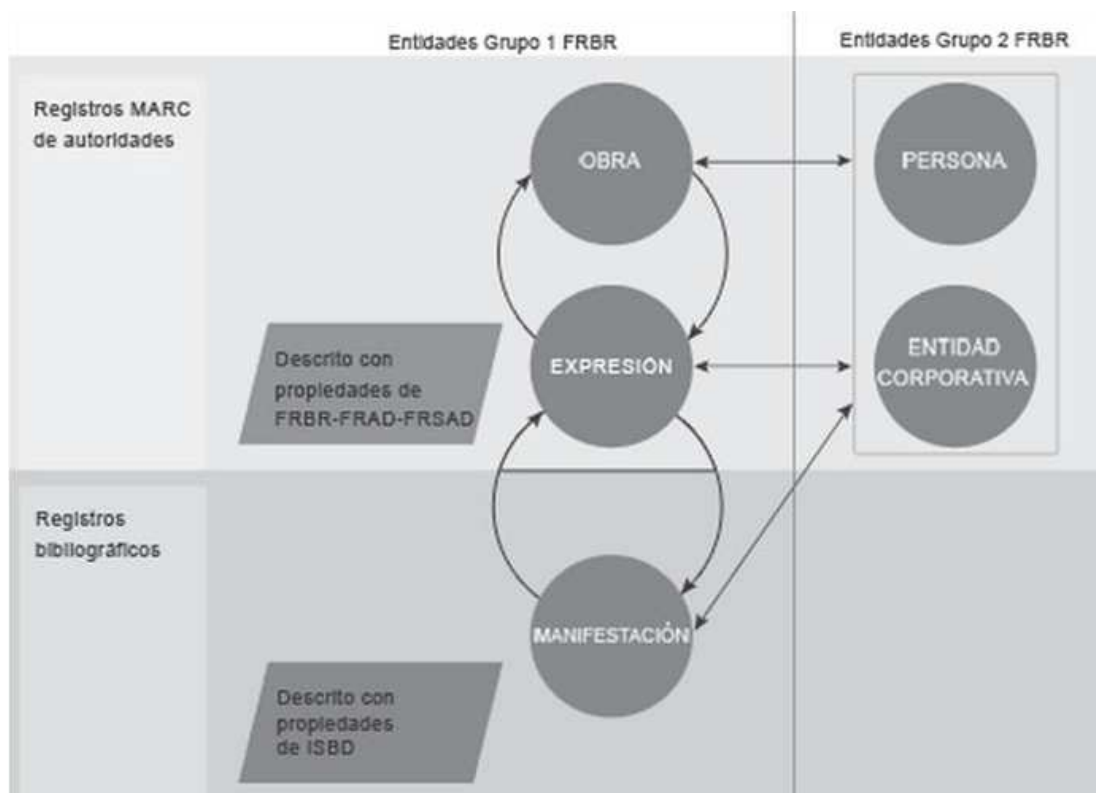
Sono stati estratti oltre 4 milioni di record relativi a risorse di varia tipologia (manoscritti, monografie antiche e moderne, risorse elettroniche, periodici, mappe, stampe, fotografie, musica stampata, suoni e registrazioni audiovisive) e oltre 4 milioni di record d'autorità di persone, enti, conferenze, opere, espressioni, materiali, nomi geografici e termini di genere utilizzati nelle registrazioni bibliografiche. Le serie di record vengono originariamente codificate secondo MARC21.

#### *Arricchire un'esperienza di navigazione in LOD*

Libros interactivos è una sezione dedicata a un altro progetto che la BNE segue insieme a Telefónica, una compagnia di telecomunicazioni spagnola. L'obiettivo è mettere in risalto le risorse particolarmente preziose delle collezioni della BNE, sia arricchendo ciascuna opera di informazioni sul contesto culturale e temporale in cui sono state concepite, sia arricchendo l'esperienza dell'utente con contenuti multimediali che permettono una nuova fruizione delle opere.

#### *Trasformare e rappresentare dati in RDF*

La rappresentazione dei dati in RDF è stata compiuta tramite l'uso delle ontologie e dei modelli proposti dall'IFLA. la BNE ha assunto FRBR come modello concettuale di riferimento e di presentazione dei dati. Il modello di dati usa l'ontologia FRBRer.



**Figura 5** – Schema concettuale dell’uso del modello FRBR nel progetto datos.bne.es

È molto utilizzato lo schema dei dati ISBD, insieme ad altre ontologie, tra cui RDA, Dublin Core, e BIBO. Ciascuna risorsa ha un certo numero di collegamenti esterni, tra cui VIAF, Sudoc, Dbpedia, Wikipedia, Libris, ISNI, Library of Congress Name Authority File, Deutsche Nationalbibliothek, Bibliothèque nationale de France (con data.bnf.fr).