



# Collegare i metadati delle biblioteche al web: l'esperienza della Germania

Gabriele Meßmer

Una domanda che ci siamo posti non appena ha preso il via il progetto dei Linked Open Data (LOD) nella Biblioteca di Stato Bavarese è stata «Qual è il valore di un catalogo di più di 23 milioni di record?». Molte generazioni di bibliotecari hanno svolto un ottimo lavoro nella creazione di record catalografici leggibili dalla macchina, oggi chiamati metadati, con l'intento di descrivere libri, musica a stampa, manoscritti, carte geografiche, di sviluppare authority file, documentando così il materiale posseduto. Oggi, per aumentare il valore di questi dati, assai costosi da produrre, è essenziale non solo offrire interfacce di recupero dell'informazione, ma anche rendere disponibili i database catalografici e dare libero accesso ai record. In Germania sono state alcune biblioteche della rete North-Rhine-Westphalian le prime a pubblicare, nel marzo 2010, i loro record come LOD. Parallelamente, l'HBZ ha creato un sito internet LOD. Da allora, un numero sempre maggiore di biblioteche ha avviato il dibattito sugli open e linked open data e su come rendere accessibili i loro dati in modo gratuito. In Germania esistono sei reti di biblioteche che mantengono cinque diversi cataloghi unici. Sebbene le reti di catalogazione tedesche lavorino in stretta collaborazione fra loro da diversi anni, condividendo gli authority files, proponendo e scambiando record per il loro riutilizzo, nei cataloghi unici persistono

ancora molti casi di diversi record per lo stesso documento.

A causa della speciale situazione che caratterizza i cataloghi delle cinque diverse biblioteche della rete, non esiste in Germania un identificatore unico per i record catalografici. Molti record hanno già un OCLC number, che viene in parte utilizzato come identificatore, ma non tutti i record ne hanno uno, e a volte record che descrivono lo stesso libro possono avere due o più identificatori diversi. Dopo l'avvio del progetto LOD in Germania, è chiaro come sia diventata basilare la necessità di un identificatore comune e persistente per ogni titolo e, idealmente, solo un unico identificatore per record diversi che descrivono la stessa risorsa. È a questo scopo che nasce Culturegraph,<sup>1</sup> un servizio di LOD che genera un identificatore specifico per tutti i tipi di documenti posseduti dalle biblioteche in Germania. Questo identificatore dovrebbe essere unico e persistente, usato come riferimento alla descrizione dei vari documenti e possedere una sintassi ben definita.

In un prima fase le reti delle biblioteche tedesche hanno fornito record di monografie e di opere in più volumi pubblicate dopo il 1945, per essere inserite nel database Culturegraph. Successivamente i record sono stati messi a confronto e sono stati creati raggruppamenti (clusters) con record che – sebbene leggermente differenti – si è ritenuto definissero lo stesso oggetto. Il criterio di schiacciamento prevede l'utilizzo di una serie di identificatori come ISBN, ISMN, OCLC number e altri. Al momento attuale in Culturegraph è disponibile un servizio di risoluzione e di recupero di singoli titoli o gruppi di essi. Il prossimo passo sarà quello di costruire un database più ampio, formato da tutti i record posseduti dalle reti delle biblioteche tedesche, un obiettivo che rappresenta una sfida quando è applicato ai primi libri a stampa, caratterizzati da titoli lunghi ed elaborati.

---

<sup>1</sup><http://www.culturegraph.org>.

Il centro di servizio bibliotecario North-Rhine-Westphalian (hbz) ha commissionato al giurista berlinese Till Kreutzer l'analisi degli aspetti giuridici degli open library data. Nel 2011 egli ha pubblicato delle linee guida (*Open Data – Freigabe von Daten aus Bibliothekskatalogen*) che costituiscono una buona base per rispondere a tutte le domande di natura giuridica sugli open data. Le linee guida contengono alcuni capitoli sulla protezione dei dati e, in modo particolare, di singoli campi. Sono presi in considerazione database e collezioni di dati come i cataloghi, e viene esaminata la possibilità di aprire i cataloghi e pubblicare i record come open data.

Il primo servizio di LOD della Biblioteca nazionale tedesca è stato la pubblicazione, nel 2010, dell'authority file dei nomi personali e dei soggetti. Oggi, la Biblioteca nazionale tedesca rende disponibili tre serie di dati: record relativi alla collezione principale (escludendo i record per la musica a stampa), il catalogo unico dei periodici tedeschi (Zeitschriftendatenbank, ZDB) e l'authority file tedesco nella sua interezza (Gemeinsame Normdatei, GND). Il modello per i dati bibliografici è documentato in un saggio (*The Linked Data Service of the German National Library: Modelling of bibliographic data*) disponibile sul web anche in inglese. Tutti i data pools sono pubblicati sotto licenza Creative Commons Zero (CC0).

## **Il Progetto B3Kat – open data**

Nel 2010 anche la Commissione per la catalogazione e i metadati delle reti delle biblioteche bavaresi e di Berlino-Brandeburgo ha iniziato a dibattere sulla questione degli open e linked open data. Queste biblioteche utilizzano un database catalografico comune chiamato B3Kat. Il catalogo contiene più di 23 milioni di record, appartenenti a 180 biblioteche. I record sono memorizzati in MAB, lo speciale formato di scambio tedesco, e sono collegati ai record

dell'authority file nazionale. È stato istituito un piccolo gruppo di lavoro al fine di giungere velocemente a un risultato. Inizialmente, il gruppo ha identificato le ragioni che possono spingere a possedere un catalogo di metadati aperti:

- rendere i dati locali e regionali visibili a livello mondiale, e non averli più nascosti nel deep web;
- completare e aumentare il valore di risorse/servizi già esistenti sul web, come ad esempio Wikipedia, fornendo link a risorse autorevoli;
- fornire dati per nuovi servizi web-based,
- integrare dati nel web semantico con la possibilità di riutilizzo dei dati completi e arricchiti nel proprio ambiente di catalogo;
- contribuire al movimento open access e promuoverlo.

Nella parte A del progetto è stato sviluppato un OAI-PMH repository nell'ambiente Aleph500 di B3Kat. I record, mantenuti nel sistema Aleph e strutturati secondo il formato di scambio tedesco MAB, sono stati convertiti in MARCXML e successivamente forniti come open data. La sfida, in particolare, consisteva, da una parte, nel mappare il maggior numero possibile di campi MAB verso il formato MARCXML, al fine di includere un massimo di informazioni; dall'altra, essa prevedeva l'inclusione di informazioni di base sulle biblioteche che possiedono tali record. Dal momento che il catalogo contiene più di 23 milioni di record – un enorme mole di dati – vi erano due vie per prelevare i record: il set di dati completo, diviso in tre parti, bloccato ad una certa data, e un OAI repository che comprende l'aggiornamento costante dei record esistenti o di quelli creati ex novo.<sup>2</sup> Sia il set di dati completo che l'OAI repository sono

---

<sup>2</sup>Informazioni sull'open data pool sono disponibili all'indirizzo web [https://opacplus.bib-bvb.de/TouchPoint\\_touchpoint/help.do?helpContext=opendata\\_en](https://opacplus.bib-bvb.de/TouchPoint_touchpoint/help.do?helpContext=opendata_en).

accessibili gratuitamente. È possibile scaricare l'intero set di dati o selezionare solo i record di una specifica biblioteca, selezionare un singolo record (se il B3Kat ID number è conosciuto) o ottenere specifici set di record che possono essere identificati in quanto includono determinati codici o campi. Il catalogo sarà pubblicato due volte l'anno. Al marzo 2012, il set completo o sue parti sono stati scaricati più di 400 volte.

## **Il Progetto B3Kat – linked open data**

L'open data pool in formato MARCXML serve di base al LOD pool (parte B del progetto). Questa è stata la via più semplice per il passo successivo, ovvero la trasformazione in RDF, poiché vi sono già degli strumenti per questa procedura. Molti campi forniti in formato MARCXML hanno dovuto essere mappati verso il modello di dati RDF e pubblicati come RDF data. Sono stati utilizzati gli URI quando possibile; quindi per ogni record è stato creato un URI basato sul numero identificativo del B3Kat (che iniziava con BV), e il namespace è stato riservato specificamente per B3Kat lod.b3kat.de. Questo namespace è stato registrato nel registro dei domini della Germania (DENIC), che sta al livello più alto del dominio .de.

Per collegare dati ad altri dati nel miglior modo possibile, è stato implementato un certo numero di collegamenti, arricchiti del contenuto speciale dei nostri record: per esempio, collegamenti all'authority file integrato tedesco, a WorldCat, al codice di lingua ISO 639-2, ai soggetti della Library of Congress e alla Classificazione Decimale Dewey. Attualmente, il database in RDF contiene circa 600 milioni di triple RDF.

Nel marzo 2012, il database è stato scaricato più di 680 volte e 7600 sono state le visite registrate. Per il momento, il set di LOD viene

## La stampa a Firenze : 1471 - 1550 ; omaggio a Roberto Ridolfi at lod.b3kat.de

<http://lod.b3kat.de/title/BV000434000>



Min per i Beni Culturali e Ambientali, Bibl Riccardiana, Firenze Catalogo a cura di D E Rhodes

Property	Value
dc:description	Min per i Beni Culturali e Ambientali, Bibl Riccardiana, Firenze Catalogo a cura di D E Rhodes
bibo:editor	<ul style="list-style-type: none"><li>&lt;http://id-nb.info/gnd/101713622&gt;</li><li>&lt;http://id-nb.info/gnd/11913263X&gt;</li></ul>
ifbr:exemplar	<ul style="list-style-type: none"><li>&lt;http://lod.b3kat.de/bib/DE-12/Item/BV000434000&gt;</li><li>&lt;http://lod.b3kat.de/bib/DE-12/Item/BV000434000&gt;</li><li>&lt;http://lod.b3kat.de/bib/DE-29/Item/BV000434000&gt;</li><li>&lt;http://lod.b3kat.de/bib/DE-35/Item/BV000434000&gt;</li><li>&lt;http://lod.b3kat.de/bib/DE-384/Item/BV000434000&gt;</li><li>&lt;http://lod.b3kat.de/bib/DE-739/Item/BV000434000&gt;</li></ul>
dcterms:extent	55, [20] S. III.
marcrel:hnr	<http://id-nb.info/gnd/101713622>
foaf:homepage	<http://worldcat.org/oclc/230828221>
bibo:isbn	8822232534
dcterms:issued	<ul style="list-style-type: none"><li>1984</li><li>1984 (xsd:int)</li></ul>
dc:language	<http://id.loc.gov/vocabularies/iso639-2/und>
rdagr1:placeOfPublication	Firenze
dc:publisher	Otschki
is:owl:sameAs of	<http://lod.b3kat.de/ISBN/8822232534>
owl:sameAs	<http://www.culturegraph.org/about/BVBBV000434000>
dc:subject	<ul style="list-style-type: none"><li>015.45/51 (dcterms:DDC)</li><li>Ausstellung (de)</li><li>Geschichte 1471-1550</li><li>&lt;http://id-nb.info/gnd/4008598-3&gt;</li><li>&lt;http://id-nb.info/gnd/4017581-9&gt;</li><li>&lt;http://id-nb.info/gnd/4027041-5&gt;</li><li>&lt;http://id-nb.info/gnd/4126475-7&gt;</li><li>&lt;http://dewey.info/class/015/about&gt;</li><li>&lt;http://lod.b3kat.de/nk/AN22200&gt;</li><li>&lt;http://lod.b3kat.de/nk/1U1050&gt;</li><li>&lt;http://lod.b3kat.de/ss/g/6.12&gt;</li></ul>
dc:title	La stampa a Firenze : 1471 - 1550 ; omaggio a Roberto Ridolfi
rdf:type	bibo:Book

Diese Seite zeigt Informationen aus dem SPARQL endpoint <http://lod.b3kat.de/sparql>.

As Turtle | As RDF/XML | Browse in Disco

Figura 1: Esempio di un record in B3kat.

pubblicato solamente ogni due anni, mentre gli aggiornamenti per gli open data avvengono in tempo reale.<sup>3</sup>

Sia il database degli open data che dei LOD di B3Kat sono stati messi in linea nei primi giorni del dicembre 2011. Al momento è il più

<sup>3</sup>Informazioni sul set di dati RDF, il modello dei dati, le ontologie usate e SPARQL sono disponibili alla pagina web del progetto <http://lod.b3kat.de>. Disponibile solo in tedesco.

grande set di record bibliografici disponibile in Germania.

## Aspetti giuridici

Un argomento importante nella discussione sugli open data è rappresentato dagli aspetti giuridici. Il gruppo di lavoro ha dovuto affrontare diverse questioni:

- I fornitori di servizi acconsentiranno alla pubblicazione dei loro record?
- Verrà accettata da tutte le biblioteche la pubblicazione dei loro record?
- Ci sono campi/tags che non dovrebbero essere pubblicati, come ad esempio i soggetti o le URL?
- Abstracts o indici integrati nei record potrebbero arricchire il catalogo?
- Quale tipo di licenza dovrebbe disciplinare gli open data?

Alcune biblioteche tedesche hanno deciso di non pubblicare i record nella loro interezza come open o LOD, ma di omettere dagli open data alcuni campi come le URL. Alcuni bibliotecari ritengono che le parti di un record particolarmente costose da produrre, come ad esempio i soggetti, non dovrebbero essere pubblicate gratuitamente. Il gruppo di lavoro Bavaria-Berlino-Brandeburgo ha invitato comunque a pubblicare record quanto più completi e esaustivi possibili – al fine di renderli davvero significativi per tutte le parti interessate e assicurare che questo servizio sia anche utile allo scambio generale di dati tra biblioteche e reti. Al momento, solo le URL collegate agli indici resi disponibili da fornitori di servizi commerciali non possono essere pubblicate per ragioni di copyright.

Prima di pubblicare i metadati è inoltre necessario considerare e definire le condizioni giuridiche per il loro riutilizzo. Vi sono due modelli:

- pubblicare dati sotto una licenza speciale oppure
- rinunciare ai diritti.

Per conformarci completamente al concetto e ai prerequisiti dei LOD, è necessario che questi siano posti sotto una licenza completamente gratuita. I record B3Kat sono quindi pubblicati sotto CC0 Universal Public Domain Dedication, che viene usato anche per i metadati nel contesto di Europeana. Questo permette il massimo uso dei record e il fornitore non ha nessun carico amministrativo per controllare le licenze degli utenti.

## Conclusioni

Il progetto è giunto a un punto d'arrivo, ma non è terminato. Ci sono ancora molte cose da fare: entrambe le pagine web dei progetti devono essere pubblicate anche in inglese. Le informazioni sulla licenza devono essere integrate nei record, sia in quelli MARCXML sia in quelli RDF. Inoltre, è necessario un processo di aggiornamento per mantenere i LOD sempre aggiornati, il che implica il trattamento di record corretti ed eliminati. Una forte richiesta da parte degli utenti è quella di pubblicare il set completo di 23 milioni di record nel repository OAI-PMH basato sul formato MARCXML e non dividerlo in tre parti come avviene attualmente, ma al momento questo non può essere realizzato per motivi funzionali e di hardware. Entrambe le parti del progetto hanno dato buoni risultati. L'implementazione è stata portata avanti velocemente, i set sono richiesti frequentemente e abbiamo imparato molto sulla pubblicazione di metadati



in MARCXML (la nostra esperienza finora è stata prevalentemente con il formato tedesco MAB) e su come farlo in un repository OAI. Tecnici e catalogatori devono lavorare congiuntamente per ottenere il meglio dai nostri record già esistenti. Pubblicare metadati non è un lavoro da fare semplicemente oltre al lavoro quotidiano. Necessita di tempo e denaro; bisogna calcolare bene i costi del personale, dell'hardware e del software, e soprattutto il tempo necessario per mantenere i dati aggiornati.

Fornire metadati può inoltre significare uno spostamento in termini di gestione dei dati all'interno della comunità bibliotecaria. Finora, l'invio di dati per ragioni diverse è dipeso dal fornitore del database e un consistente lavoro deve essere dedicato alle diverse definizioni di file e trasferimenti, sempre dipendenti da chi invia e non da chi riceve. Con Z39.50, e ancora di più con OAI, dipendeva dal fornitore di dati definire la metodologia per il loro invio, la struttura e i campi da trattare. Adesso con gli open data e, maggiormente con i LOD, sta all'utente compiere le sue scelte ed eventualmente un rimodellamento dei record. Tuttavia, la curva di apprendimento nell'analizzare i dati forniti al fine di un loro migliore utilizzo è ancora da affinare, e deve essere realizzata una più completa standardizzazione e armonizzazione dei formati e dei contenuti, al fine di rendere l'utilizzo degli open data un metodo agevole di cooperazione e riutilizzo dei record fra biblioteche.

Nell'estate 2012 sarà creato un altro repository OAI nel contesto del progetto Europeana Libraries, a cui partecipa fra le altre istituzioni, la Biblioteca di Stato bavarese. Uno dei risultati del progetto Europeana Libraries è il *Report on the alignment of library metadata with the Europeana Data Model (EDM)*. Questo repository userà open MARCXML data, ma conterrà unicamente metadati di oggetti digitali. Questi verranno arricchiti con collegamenti a thumbnail. Non ultimo, questo repository – anche in MARCXML – servirà come

data pool per portali tedeschi ed europei che presentano metadati per tutti i tipi di materiali digitali. Il vantaggio di EDM è quello di arricchire i record in Europeana e quindi renderli interoperabili e adatti al web semantico. Con EDM per Europeana e CIDOC CRM, il formato comune dei dati per la futura biblioteca digitale tedesca, possiamo vedere chiaramente quali ulteriori requisiti devono essere rispettati quando si trattano risorse digitali. Non si tratta solo di thumbnail, ma di ogni tipo di immagine (sicuramente non in alta risoluzione come nell'originale, ma in una risoluzione più bassa), di metadati strutturali presentati in files METS/MODS ed infine dell'informazione a testo pieno.

I Linked Open Data (LOD) porteranno a nuovi servizi che saranno sviluppati nel prossimo futuro. Questi si basano sul bagaglio di conoscenze proprio dei bibliotecari: metadati, strutture, normalizzazione, controllo di qualità, numeri standard etc. Il bibliotecario di oggi non è più un catalogatore, ma uno specialista e gestore di metadati: in questo modo la ricchezza di informazione contenuta nei record delle biblioteche deve poter essere sfruttata al meglio ed integrata nel web.

## Riferimenti bibliografici

Kreutzer, Till. *Open Data – Freigabe von Daten aus Bibliothekskatalogen*. Köln: Hochschulbibliothekszenrum des Landes Nordrhein-Westfalen, 2011. <http://www.hbz-nrw.de/dokumentcenter/veroeffentlichungen/open-data-leitfaden.pdf>. (Cit. a p. 3).

*Report on the alignment of library metadata with the Europeana Data Model*. 2011. <http://www.europeana-libraries.eu/documents/868553/1eade085-34ac-487f-82af-d5cd2545e619>. (Cit. a p. 9).

*The Linked Data Service of the German National Library: Modelling of bibliographic data*. Leipzig: Deutsche National Bibliothek, 2012. [http://www.dnb.de/SharedDocs/Downloads/EN/DNB/service/linkedDataModellierungTitelraten.pdf?\\_\\_blob=publicationFile](http://www.dnb.de/SharedDocs/Downloads/EN/DNB/service/linkedDataModellierungTitelraten.pdf?__blob=publicationFile). (Cit. a p. 3).

---

**Ai fini di una corretta indicizzazione, si invitano i lettori a citare esclusivamente il testo in lingua inglese; l'unico, infatti, che presenta l'indicazione del numero di pagina, l'abstract, le keywords e le date del processo redazionale.**

Meßmer, G. "Linking library metadata to the web: the German experience". *JLIS.it*. Vol. 4, n. 1 (Gennaio/January 2013): Art: #5507. DOI: [10.4403/jlis.it-5507](https://doi.org/10.4403/jlis.it-5507). Web.

