



Catalogare in modo aperto: disintegrazione e distribuzione del record

Martin Malmsten

Il contesto

LIBRIS, il catalogo collettivo svedese, è nato nel 1972 con l'obiettivo di razionalizzare il lavoro in biblioteca con il supporto della tecnologia dell'informazione. Per questo è stato creato un ambiente di catalogazione cooperativo in cui ogni bibliotecario avrebbe descritto la propria collezione, con il presupposto di base, nello spirito di quei tempi, che ogni biblioteca avrebbe contribuito con ciò che di unico essa possedeva, ma al contempo avendo accesso a tutte le descrizioni delle altre biblioteche partecipanti. L'accesso al database era ristretto alle biblioteche aderenti via terminal 3270 e con una connessione via modem al mainframe LIBRIS. Nel 1997 è stato messo a punto un servizio per ricercare e mostrare dati nel web, rendendo di fatto l'informazione accessibile a tutti e ovunque, a condizione che avessero accesso alla rete. Più o meno nello stesso periodo è stato reso disponibile Z39.50¹, rendendo possibile agli utenti della catalogazione derivata ricercare da remoto e scaricare dati da LIBRIS. Questo è stato un passaggio fondamentale anche

¹<http://www.loc.gov/z3950/agency>.

perché consentiva alle macchine di accedere all'informazione sulle collezioni descritte in LIBRIS. Sebbene fosse considerato sufficiente per la catalogazione derivata, sempre che venisse usato lo stesso formato, Z39.50 era carente in altre aree, poco *user friendly*, e problematico nel collegarsi ad altri sistemi al di fuori della comunità bibliotecaria. Nel 2006 fu approntato e inaugurato un nuovo catalogo, un "catalogo di nuova generazione", e con esso una serie di numeri Application Programming Interface (API) con l'intento di rendere più semplice per tutti creare servizi costruiti sui dati di LIBRIS. Ne derivò che per ogni record bibliografico fosse creata una URI, o meglio una URL, rendendo possibile per il sistema stesso e le applicazioni esterne collegarsi a record singoli in modo semplice e permanente. Per utile e importante che fosse conoscere (molto) su API e cAPIre i formati disponibili (MARC21, DC, MODS, ecc.), tutto però si basava ancora sulle descrizioni contenute all'interno dei record. Inoltre, una volta trovati, nei record erano presenti pochi link alle risorse al di fuori di LIBRIS anche se i collegamenti ai siti quali Google Books erano comunque forniti all'utente. Nel 2008 fu diffusa la versione di LIBRIS nel suo insieme come linked data (Malmsten), compresi i dati di autorità che descrivono le persone e le organizzazioni, e le intestazioni di soggetto. Furono aggiunti i collegamenti alle risorse esterne, ad esempio quelli descritti da Library of Congress Subject Headings (Library of Congress Subject Headings (LCSH)), Wikipedia e Virtual International Authority File (Virtual International Authority File (VIAF)) e immediatamente LIBRIS divenne parte di quel grafico in rAPI da espansione di metadati generati da una serie di entità prevalentemente esterne al settore GLAM. Questa iniziativa ha destato grande interesse in particolar modo da parte di altri enti governativi e di altre organizzazioni che miravano a collegarsi a, oppure a scaricare completamente, parti di dati di autorità. E questo non deve sorprendere in quanto un identi-

ficatore, per esempio di un autore famoso, è utile per le biblioteche, per gli archivi e per altre istituzioni di beni culturali.

Catalogare in modo aperto

A partire da settembre 2011 la Bibliografia Nazionale e l'Authority File svedese, due sottoinsiemi del database LIBRIS, sono messi a disposizione nello stesso formato in cui sono stati creati (MARC21). Questa decisione non solo di esporre i record sia nella loro forma originale e in Resource Description Framework (RDF)/linked data è strategica. In questo modo chiunque può vedere, valutare, fare riferimento e in ultima analisi, contribuire al lavoro svolto dalla Biblioteca Nazionale, Ciò implica che la visibilità e l'apertura conduce ad una maggiore qualità dei dati.

Per evitare eventuali restrizioni quando si tratta di ri-utilizzare i dati, la Biblioteca Nazionale ha scelto² CC0³ per la Bibliografia nazionale e l'Authority File, mettendo i set di dati disponibili al pubblico. L'unica eccezione è il campo MARC 667 (nota non pubblica), che viene filtrato a causa di motivi di integrità personale. E' stata scelta tale licenza perché sono stati notati problemi con le licenze di attribuzione quali ODB-BY e CC-BY quando si tratta di ri-uso di dati nel corso del tempo, ad esempio, i cosiddetti "attribution stacking". L'obiettivo è quello di rilasciare il set di dati nel formato originale con CC0, anche se alcuni record o parte di record in LIBRIS provengono da altre organizzazioni (LC, BNB, DB, OCLC, ecc.): ciò richiederà un certo tempo e lavoro. Chi volesse accedere ai dati può farlo in due modi: attraverso Atom feeds e/o utilizzando il protocollo OAI-PMH. I feed sono rappresentati, essenzialmente, ottenendo i dati dal sistema di

²<http://librisbloggen.kb.se/2011/09/21/swedish-national-bibliography-and-authority-data-released-with-open-license>.

³<http://creativecommons.org/choose/zero>.

catalogazione in tempo reale. Ciò significa che chiunque può ottenere le modifiche apportate all'Authority File entro pochi secondi dal cambiamento. Il lavoro dei catalogatori è stato reso disponibile in sostanza nel momento stesso in cui digitano. Per inciso, la scelta di non fornire i file completi per il download ("lo scarico grezzo") dei dati sta ad indicare che il set dei dati è vivo, mentre il grezzo è fondamentalmente stantio e/o obsoleto nel momento stesso in cui viene scaricato. Comunque, sia Atom che OAI-PMH possono essere usati per scaricare il set completo dei dati, quindi forse la distinzione è solo accademica. Quanto appena descritto va visto come un investimento strategico nella direzione dell'apertura e la conclusione logica dell'idea cooperativa che sottende a LIBRIS, solo su scala più ampia: contribuire con, e condividere, ciò che è unico per se stessi ed avere accesso al resto. Questa in sintesi la promessa dei linked data, e cioè che si può distribuire la creazione dell'informazione senza doverla costantemente aggregare per renderla utile.

Utilizzare i linked data

Tuttavia, per raccogliere di fatto i benefici dei linked data, dobbiamo utilizzarli anche come parte integrante dei nostri sistemi, e non solo esporli. Questo comporta una serie di interessanti conseguenze. In primo luogo, poiché i linked data ci consentono di collegarci con qualsiasi dato, ovunque questo venga creato, viene a cadere la distinzione tra set di dati interni ed esterni. Questo ha un impatto profondo sull'impostazione del sistema in quanto bisogna fare affidamento su protocolli abitualmente usati per set di dati esterni internamente. La questione del controllo diventa poi una questione di fiducia piuttosto che un fatto tecnologico. Se non si riesce ad avere il controllo sull'informazione bisogna decidere di chi fidarsi, sebbene questo sia un prezzo decisamente basso da pagare per un

mondo di dati. In genere coloro di cui ci si fida sono anche gli utenti dei dati prototti. In terzo luogo, poiché un numero sempre maggiore di informazioni nei nostri record è collegato ad una qualche fonte al di fuori del nostro controllo, sia essa una persona in Wikipedia o un'intestazione di soggetto in id.loc.gov, il concetto di record diventa in qualche modo meno interessante.

Fluidità

Spesso non c'è necessità di aggregare i dati prodotti da altri, ma piuttosto di reagire al fatto che qualcosa a cui ci si collega è cambiato. Di nuovo, questo rende problematici i download completi di set di dati a intervalli discreti poiché

1. la ragione del cambiamento può non essere evidente e
2. avere importazioni multiple di batch che si riferiscono agli stessi dati causa decisamente dei problemi.

Un obiettivo per un sistema davvero collegato deve essere la capacità di diffondere alle parti interessate i cambiamenti, o le informazioni sui cambiamenti in modo fluido in tempo reale o quasi. Se creare dei feed rende possibile per gli utenti richiedere gli aggiornamenti, più spesso di quanto si possa immaginare la risposta è no, che non ci sono aggiornamenti. Questo contribuisce a rendere il sistema molto inefficiente, in quanto si devono inviare molte richieste per assicurarsi che i set di dati siano in sincronia. Ci sono almeno due iniziative che riguardano questo argomento: pubsubhubbub⁴ e ResourceSync⁵. Usando degli hub a cui l'editore può segnalare una modifica e a cui l'utente può abbonarsi si crea una rete efficiente tramite la quale diffondere l'informazione.

⁴<https://code.google.com/p/pubsubhubbub>.

⁵<http://www.niso.org/workrooms/resourcesync>.

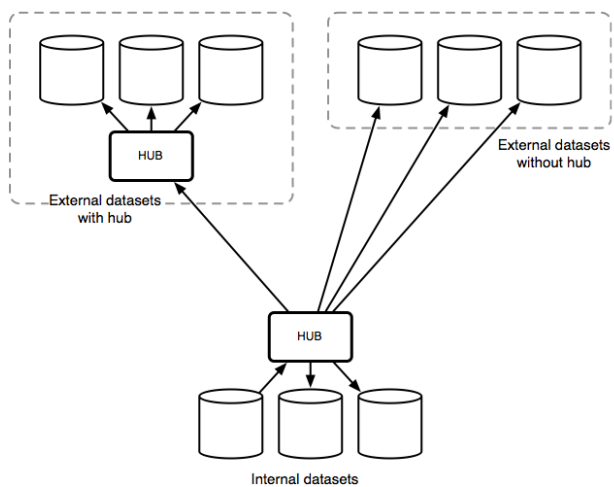


Figura 1: Un cambiamento in un set di dati si propaga attraverso hubs, sia per set di dati interni ed esterni

Conclusioni

In effetti, usare i linked data, invece che esporli soltanto, elimina in qualche modo la distinzione tra set di dati interni ed esterni. Il controllo diventa una questione di fiducia, non un fatto tecnologico. Il record si disintegra nel momento in cui i dati vengono distribuiti.

Riferimenti bibliografici

Malmsten, Martin. «Making a Library Catalogue Part of the Semantic Web». *International Conference on Dublin Core and Metadata Applications, DC-2008–Berlin Proceedings*. 2008. (Cit. a p. 2).

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.

Malmsten, M. "Cataloguing in the open: the disintegration and distribution of the record". *JLIS.it*. Vol.4, n.1 (Gennaio/January 2013): Art: #5512. DOI: [10.4403/jlis.it-5512](https://doi.org/10.4403/jlis.it-5512). Web.

