



I linked open data verso library management e discovery solution di futura generazione

Axel Kaschte

Ex Libris è nota per essere una compagnia innovativa che sposa davvero le nuove tecnologie. E i linked data sono una tecnologia molto avanzata e rivoluzionaria. Nell'agenda di questo evento vediamo che vengono trattati molti aspetti dei linked data, io mi soffermerò su cosa cerca una compagnia commerciale come Ex Libris in questo nuovo scenario. Tutte le volte che guardo qualcosa di nuovo, mi torna in mente *Il piccolo principe*. Egli ha questa eccezionale capacità di guardare a certe cose con nuovi occhi. Probabilmente ricorderete la domanda: «Che cosa vedi qui?» mentre il Piccolo Principe mostra un disegno. Molte persone rispondono immediatamente che si tratta di un cappello. Ma tutti sappiamo (dal libro) che la risposta è: un serpente che ha ingoiato un elefante. È una curva molto simile a quel cappello e, come potete immaginare, ora vi pongo io la domanda: «Cosa vedete qui?». Non è il serpente, questo ve lo posso dire. È una curva a campana. Per essere più precisi, è la curva a campana relativa all'adozione delle innovazioni di Rogers Everett (*Diffusion of Innovations*), che ha inventato questo modo di rappresentare in-

formazioni del genere, e mostra l'uso delle nuove tecnologie fatto nel tempo. Quindi, agli esordi di una nuova tecnologia, si vede che c'è un 2.5% di innovatori. Queste sono le biblioteche che cavalcano le nuove tecnologie, e quindi che cercano davvero di investire nelle novità. Segue un 13% di coloro che si adattano in tempi brevi. È già una percentuale notevole, ma è sempre chiamata innovativa. Seguono poi la maggioranza in anticipo, la maggioranza in ritardo e via dicendo. Per fare degli esempi di prodotti che ci sono familiari nell'era di internet, ci sono un televisore e un quotidiano. Se uno di questi si trova a destra della curva a campana sull'adozione delle nuove tecnologie non vuol dire che non è usato – ma il contrario. Vuol dire che tutti lo hanno già e che lo stanno usando. La curva a campana fa riferimento al tasso di crescita del numero di utenti aggiunti al mese. Attualmente Facebook è in cima, il che vuol dire che la percentuale di crescita è ancora eccezionale ma che ci saranno sempre meno nuove persone ad aggiungersi. Ci sono già seicento milioni di utenti. Potete vedere anche nuovi siti web come ad esempio Vimeo, e tecnologie come l'iPhone. Il Blackberry si colloca leggermente più in alto. L'iPhone è più recente. Il Kindle di Amazon, l'e-book reader, sono ancora nella prima fase di adattamento. È sul mercato da tre anni ma la curva di adattamento mostra quanta strada c'è ancora da fare per questo prodotto. Vedete, c'è anche il discorso del profitto nel mercato di massa. Ora il punto è che, noi come compagnia commerciale siamo ovviamente interessati all'utile, ma ci interessa anche servire i clienti come voi con servizi commerciali. Ogni qualvolta che c'è una nuova tecnologia l'obiettivo è renderla disponibile a molti, renderla veramente disponibile come soluzione conveniente, in quel caso è la compagnia commerciale che ha il modello migliore. Come si applica questo alle biblioteche? Integrated Library System (ILS), sono stati inventati molto tempo fa, circa trent'anni fa o anche di più, e come potete vedere ci sono ancora delle biblioteche, soprattutto in Asia,

che ancora non lo hanno. Questo spiega come mai il tasso di crescita delle nuove biblioteche sia lento, ma queste sono ancora tecnologie allo stato dell'arte e sono ancora richieste. Poi ci sono cose come la meta-ricerca, come i link resolution. È molto interessante che, se guardiamo le nostre statistiche, un numero elevato di biblioteche acquista ancora i *link solution*. Ce ne sono molte che ancora non entrano nel mondo delle risorse elettroniche ma stanno per farlo o lo faranno a breve. Per questo obiettivo avranno bisogno di un *link resolver*, il che spiega perché questo si trova nella zona alta della curva a campana. Diamo uno sguardo a Discovery, la generazione successiva dell'OPAC, dove vediamo entrare in gioco le tecnologie dei motori di ricerca: Discovery si trova leggermente prima dell'apice della curva a campana. Guardiamo le statistiche: Primo, il nostro *discovery solution* è in rapida crescita ma non è ancora in cima¹. E possiamo vedere anche indici centrali di risorse elettroniche come Primo Central, una mega-aggregazione di articoli disponibili elettronicamente per la ricerca, per i ricercatori e così via. È partito appena due anni fa, è stato adattato molto rapidamente e siamo in fase di rapida espansione. Queste sono solo categorie di prodotti, ora affiancherò ad esse le tecnologie. ILS è una tecnologia non più in espansione. Cercate Application Programming Interface (API), Open URL, i motori di ricerca, e poi cloud come *e-technology* che consente a questo tipo di servizio di avere un unico indice centrale per ogni biblioteca ed è caratterizzato anche da un elevato rapporto costi-benefici. Questo chiarisce bene perché le compagnie commerciali sono così brave ad elaborare tali modelli. Se sono in grado di fornire a molte biblioteche soluzioni molto favorevoli dal punto di vista costi-benefici, questo è il modello che le biblioteche dovrebbero utilizzare. In questo modello ci sono altri particolari da considerare. Geoffrey Moore(Moore) ha introdotto il concetto di abisso. Ci sono

¹<http://www.exlibrisgroup.com/category/PrimoOverview>.

molti prodotti che si trovano ancora in fase di innovazione e che non diventeranno mai prodotti di massa. Cadranno appunto nell'abisso. Ex Libris ha creato due cose che sono precipitate nell'abisso: una ERM – probabilmente molti di voi hanno sentito parlare di Verde – e un sistema di gestione dei beni digitali. Se ci riflettiamo, non è solo Ex Libris ad aver fallito ad introdurre prodotti come questi nel mercato di massa. Ci sono solo poche centinaia di biblioteche che li usano, contro quattromila che usano Aleph in tutto il mondo. Non c'è più espansione. L'abisso è semplicemente un modo di presentare quella tecnologia che non risponde nel migliore dei modi alle esigenze della biblioteca: non ha mai raggiunto il livello di una best practice. Sarò molto franco. Bisogna fare un'analisi così come farebbe una compagnia commerciale ed essere in grado di dire: «va bene, è stato un errore. È stato denaro speso male, sia per noi che per voi». Ciò che stiamo introducendo ora, e che il mercato sta seguendo come idea, è una soluzione al problema dell'automazione nei silos. Se procediamo con i silos, se procediamo uniti – nel senso che c'è un'unica soluzione per la gestione della stampa e per la gestione delle risorse elettroniche e probabilmente per la gestione dei beni digitali – e cioè se viene introdotto un unico ambiente, ci accorgiamo che questo è proprio quello che le biblioteche vogliono oggi. E questo è quello che stiamo iniziando a fare in questo momento; il nostro primo cliente probabilmente lo avrà il mese prossimo. Nel nostro caso è il software chiamato Alma ed è la tecnologia cloud che lo rende possibile. Vi ho mostrato tutto questo perché vi voglio condurre all'idea che le compagnie commerciali devono essere pronte a percepire quando è il momento giusto per sposare una nuova tecnologia in modo da farne un prodotto di massa. Guardiamo a certe altre tecnologie affermate in ILS, ne abbiamo sentito parlare oggi: AACR2, MARC21, Cataloghi collettivi, file d'autorità. Sono in circolazione già da un po' di tempo. Abbiamo sentito parlare

molto anche di tecnologie emergenti quali Resource Description & Access (RDA). Ne ha parlato Alan Dunskin della British Library e lo abbiamo sentito chiedere «Per favore, aiutateci a recuperare il tempo perduto e ad usarlo; fornite il *tool set* lì dove ha luogo la catalogazione». In altre parole, è il momento giusto e Ex Libris si impegna a fornire presto gli strumenti necessari alle sue applicazioni. Poi c'è Resource Description Framework (RDF), di cui si parla in tutto il seminario. E le fonti degli *open data*. Tutte queste nuove tecnologie si collocano molto lontano sulla sinistra della curva a campana. Sono ancora in realtà nella fase sperimentale, come lo sono molti aspetti dei *linked open data*. Abbiamo visto i diagrammi degli oggetti e le loro relazioni, abbiamo libri, dipinti, autori e pittori. Abbiamo gli oggetti e i loro creatori – e molte altre relazioni. In un certo senso i dati bibliografici erano già molto collegati in passato. Non erano aperti, non era usando Uniform Resource Identifier (URI), ma all'interno dei sistemi erano già collegati. Quindi, se pensiamo a soluzioni come i *discovery system* (per esempio il Catalogo Collettivo Austriaco a Vienna) vediamo che possiamo fare clic sull'autore ed avere tutte le manifestazioni delle opere dell'autore, e lo stesso per i soggetti. Come vedete, i collegamenti sono lì e potete navigare, ma rimanendo esclusivamente all'interno dell'ambiente della biblioteca. C'è già anche un link permanente alla manifestazione. Con questo si arriverà sempre allo stesso punto. Non è ancora URI perché non ci si collega ai dati, si arriva alla stessa pagina ma per lo meno è un modo permanente di giungervi. Costruire un API per fornire i dati in forma strutturata è la fase successiva. Ma ora c'è un ulteriore punto che si può vedere – questo è quanto è stato fatto nel Catalogo Collettivo Austriaco includendo Wikipedia come fonte di dati al di fuori dei metadati della biblioteca, viene usato il record di autorità con il suo identificatore per collegarsi con Wikipedia, se si clicca lì si ottengono informazioni appunto da Wikipedia. Un esempio banale,

si potrebbe pensare. Nelle varie biblioteche ci sono diversi esempi del genere nei vari discovery solution. Questi ancora non fanno uso del vero meccanismo URI. I link vengono costruiti al momento ed è qualcosa che funziona in quanto la piattaforma discovery e i dati sottesi consentono di presentare ciò agli utenti. Con questo esperimento in loco possiamo dare uno sguardo al livello di soddisfazione degli utenti finali. È qualcosa che vogliono davvero e su cui cliccano? Se no, non ce ne preoccupiamo. Come vedete questo è il genere di cose che cerchiamo di fare nella fase sperimentale. Quindi che cosa c'è di realmente nuovo nei *linked open data* dal nostro punto di vista? Non è una novità che si possano fare collegamenti tra manifestazioni e autori. Non è nuova nemmeno la possibilità di fare collegamenti tra i soggetti. Possiamo inserire collegamenti tra diversi tipi di dati bibliografici ma la struttura dei dati stessi è altamente specializzata, nessuno al di fuori della biblioteca di fatto la può leggere ed è molto difficile scambiare ed interagire al di fuori della biblioteca. Dal mio punto di vista, quindi, se devo riassumere ad una persona non addetta ai lavori che cos'è importante per le biblioteche nel contesto dei *linked open data*, direi che è rendere disponibili all'esterno i dati bibliografici e, forse ancora di più, prendere i dati bibliografici dall'esterno. In altri termini, far sì che il mondo della biblioteca sia parte dell'onnicomprensivo World Wide Web. Questo concetto viene ripreso nel lavoro del World Wide Web Consortium, W3C. Nel maggio 2010 è stato creato un gruppo incubatore per studiare i linked data nelle biblioteche e relativi sviluppi software. Questo gruppo ha studiato casi veri ed ha prodotto il *Library Linked Data Incubator Group Final Report*.² I casi riportati si riferiscono a come rendere disponibili i dati bibliografici nel mondo dei linked data. Abbiamo visto diversi casi del genere nelle presentazioni di questo seminario ed io cercherò qui di farne una sintesi inserendoli in uno

²<http://www.w3.org/2005/Incubator/llid/XGR-llid-20111025/>.

schema molto semplice fatto di tre aree di lavoro. Ma in primo luogo seguiamo quello che ha fatto il gruppo incubatore, e cioè categorizzare tutti i casi in 8 gruppi. Il primo gruppo è relativo alla gestione dei dati bibliografici da cui si passa poi ad uno scenario di linked data. Ad esempio, British National Bibliography, Bibliotheque National France, Bavaria State Library, Open Library. Il secondo gruppo è sui dati di autorità, stesse istituzioni ma dati differenti. Sto ripetendo cose che conoscete già solo per condurvi al punto in cui vedrete un modello. Il terzo gruppo riguarda il lavoro sull'allineamento del vocabolario. Ci sono state molte presentazioni nel seminario su questo argomento. Il quarto gruppo è sugli archivi che lavorano per mettere insieme i loro dati nella *linked data cloud*. Europeana – cosa alquanto interessante – viene citata nel gruppo degli archivi sebbene probabilmente lavori in tutti i gruppi. Il quinto gruppo è sulla citazione dei set di dati scientifici espressi nei linked data, cosa abbastanza nuova, mai fatta finora. Stanno cercando di potenziare le pubblicazioni, il che vuol dire che le pubblicazioni sono già dotate dei metadati sui dati di ricerca usati. Il sesto gruppo si occupa di oggetti digitali nel mondo delle biblioteche. L'obiettivo è creare un archivio di testi digitali come linked data così che i metadati, il testo e gli oggetti extra a cui il testo si riferisce siano forniti in un formato onnicomprensivo. Qui abbiamo dei casi al di fuori del mondo delle biblioteche, come la UK Open Government Data Initiative che ci offre molti esempi dai quali si possono estrarre dati e vedere come sono collegati tra loro. Il settimo gruppo è relativo alla costruzione delle collezioni. I bibliotecari hanno parlato delle strutture Functional Requirements for Bibliographic Records (FRBR), dove il livello di lavoro è altissimo. Ma che dire se si va ad un livello ancora più elevato e si comincia a descriverle le collezioni. Ci sono già dei casi in cui si cerca di definirle in set di dati e anche di più. L'ottavo (e ultimo) gruppo di casi è sui social network e sugli ambienti incro-

ciati. In questo gruppo non sembra che i casi siano collegati ad una visione classica della biblioteca ma gli utenti abituali sono molto attivi in questi ambiti scambiando informazioni soprattutto sulla letteratura che stanno usando, come ad esempio via Mendeley.

Tutto il lavoro distribuito in 8 gruppi si traduce in tre grandi aree di lavoro.

- Un'area è sulla preparazione dei dati, sulla creazione di dati che devono essere lì in strutture collegate per poter essere usati. Quest'area di lavoro si concentra sulla creazione di strumenti in grado di gestire ad alto livello la trasformazione e l'archiviazione.
- Allo stesso tempo c'è l'area in cui si definiscono le regole per la trasformazione dei dati. C'è una sorta di interazione tra le due. Ci sono dei progetti in cui si è tentato di inserire i dati in strutture collegate, ma, dopo aver fatto esperienza, si è dovuto riconoscere che no, le regole di trasformazione vanno cambiate, si devono rifare, quindi si è in una fase molto ripetitiva al momento.
- Infine, ciò che conta è quello che di fatto arriva all'interfaccia dell'utente finale e quale miglior uso possono farne rispetto ad oggi, che sarebbe poi la terza area di lavoro.

In qualche modo si direbbe che ci siano molti più progetti con l'obiettivo di far confluire i dati nella *linked data cloud* di quanti ce ne siano su come utilizzare di fatto questi dati, cosa che non si verificava prima. Questo unito alla elevata frequenza di cambiamenti nel definire le regole di trasformazione porta noi di Ex Libris alla conclusione che i linked data sono ancora in modalità di ricerca.

Un esempio di ricerca è Europeana, se ne è parlato proprio oggi nel seminario. Europeana ha un suo sito web, molto creativo – perché dico ricerca? Perché vedo i problemi che stanno affrontando.

Il portale centrale di Europeana non è in grado di riferire lo stato dell'arte della performance. Il problema sorge se si rimane fedeli ad uno dei principi fondamentali dei linked data – collegare tra loro dati di varie fonti. Queste fonti sono infatti silos di dati e per far sì che questi siano reperibili da un punto centrale bisogna fare una ricerca federata. Abbiamo in proposito l'esperienza dello scorso decennio nel fare meta-ricerche nei database delle biblioteche, per offrire soluzioni di alto profilo bisogna creare un indice centrale. Questo significa raccogliere dalle varie fonti, che di fatto significa una varietà di formati delle fonti, e ancor più importante adattare questi formati. Nel mondo RDF ciò è estremamente problematico e a mio avviso un'area di ricerca in sé. Questo è noto alle persone molto competenti che lavorano per Europeana. Per fare progressi in questo campo è stato avviato nel marzo di quest'anno un nuovo progetto finanziato dall'Unione Europea, chiamato DM2E – «Digitised Manuscripts to Europeana ». Consiste prevalentemente nel digitalizzare più materiale possibile, di farlo nel minor tempo possibile e di creare i metadati in modo semplice. Tuttavia, il Work Package 2 è sull'infrastruttura dell'interoperabilità. Poiché molte istituzioni che fanno la digitalizzazione sono biblioteche, queste hanno i loro sistemi e li usano per raccogliere i metadati in formati classici, come MARC. Siccome Europeana usa strutture di linked data, verrà creato un robusto tool di trasformazione RDF come parte del *Work Package 2* (WP2). Ex Libris è partner in WP2, partecipa a questa ricerca e di fatto crea prodotti che saranno open source e che sarà possibile aggiungere ad un sistema di biblioteca già esistente. Il tool prenderà per esempio MARC-XML e lo trasformerà in RDF. Abbiamo già parlato di esempi simili nella British National Bibliography, oggi pomeriggio abbiamo appreso come ha lavorato la Bavarian State Library. Tuttavia, questi esempi non usano tecnologia comune, si tratta di qualcosa ancora in fase sperimentale. Il tool creato nell'ambito di

WP2 consentirà diversi formati di input quali MARC21, UNIMARC, DC, MODS, trasformando questi in una rappresentazione RDF che fondamentalmente non è altro che un formato diverso di esportazione. Come fase successiva, verrà effettuata una trasformazione nel modello di dati di Europeana. Entrambi i passaggi si basano su regole di mappatura, ed in realtà in questo caso l'obiettivo è rendere molto facile cambiare le regole perché ci troviamo nella fase di definizione dei vocabolari e per questo ci piacerebbe giocare un po'. Attualmente non si conosce ancora il vocabolario definitivo che sarà usato, in ogni progetto a cui si fa riferimento nel report dei casi di cui abbiamo parlato prima si usa un'ontologia diversa. Creare un tool che consenta di giocare con le ontologie è il nostro contributo a questa fase di ricerca. In sintesi: perchè noi di Ex Libris dovremmo investire in prodotti che utilizzano la tecnologia dei *linked open data*? Per via dell'interoperabilità, soprattutto con altri domini, nel settore della ricerca. Probabilmente rimodulando la gestione dei metadati – la catalogazione molto probabilmente assumerà dei connotati completamente diversi, consisterà più nell'includere risorse esterne come i link che non nel digitare dati. Che cosa dovrebbe fare Ex Libris? Il primo passo è contribuire a colmare la lacuna nel mondo dei *linked open data* rendendo disponibile la pubblicazione dei dati bibliografici in strutture di linked data in modo che ogni biblioteca possa pubblicare con facilità i propri dati. Il secondo passaggio è utilizzare fonti di dati esterne. Questo vale sia per il sistema di ricerca che per il flusso di lavoro della catalogazione. Probabilmente c'è una terza fase in cui si modellano i servizi del futuro che noi non vediamo ora, ma che si affiancano sempre alle nuove tecnologie. I linked dati bibliografici sono nella fase di ricerca. Siamo nel ciclo: come fare, che cosa farne. Ogni volta che si fa un passo avanti ci si rende conto che: oh, potrei fare questo ed è per questo motivo che sentiamo di dover contribuire come partner di ricerca nel pro-

getto DM2E di Europea. Anche con il nostro gruppo di utenti, specialmente nell'area di Primo, ci sono molte iniziative.

Riferimenti bibliografici

Moore, Geoffrey A. *Crossing the Chasm*. New York: Harper Collins, 1992. (Cit. a p. 3).
Rogers, Everett M. *Diffusion of Innovations*. Glencoe: Free Press, 1962. (Cit. a p. 1).

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.

Kaschte, A. "1 linked open data verso library management e discovery solution di futura generazione". *JLIS.it*. Vol. 4, n. 1 (Gennaio/January 2013): Art. #5492. DOI: [10.4403/jlis.it-5492](https://doi.org/10.4403/jlis.it-5492). Web.

