



Library Linked Data: un'evoluzione

Karen Coyle

Ringrazio gli organizzatori di questo seminario, e tutti voi che parteciperete alle conversazioni di questi giorni e successive.

Per me è particolarmente gratificante essere qui a Firenze per partecipare a questo seminario su Library linked data. Dico 'particolarmente' perché sono trascorsi solo quattro anni da quando ho cercato di spiegare questo nuovo concetto proprio qui a Firenze, durante un altro convegno, e devo ammettere che il mio tentativo non ha avuto grande successo. In questi ultimi anni ho imparato molto di più su questo argomento, e i concetti del web semantico si sono diffusi sia nel mondo dell'informatica che in quello della scienza, e delle istituzioni dei beni culturali. Oggi siamo qui per fare un passo avanti nell'evoluzione del web semantico, non solo per le biblioteche, ma per tutti gli utenti del web, che sono, o che potrebbero diventare, utenti di biblioteche.

Il mio obiettivo, oggi, è di introdurre alcuni concetti di base che forniranno il contesto a quanto diremo in questi giorni. Un argomento come il web semantico tende a spostare la discussione sui dettagli – sempre importanti, naturalmente – ma è essenziale tenere d'occhio i punti di riferimento che ci servono per indirizzare questo argomento complesso nella direzione più utile alla nostra comunità, particolarmente in questo nostro momento storico, economico e tecnologico.

Per capire il nostro futuro dobbiamo sicuramente conoscere il nostro passato. Nel caso delle biblioteche, il nostro passato è lungo, e potrei solo fare un accenno ai secoli di esperienza e tradizione che oggi ci hanno condotto qui. Molte relazioni che ascolterete vi daranno una visione del nostro futuro. E allora nei pochi minuti che ho qui a disposizione, non voglio parlare del passato né del futuro, ma del presente. Spero così di offrire un contesto utile alle discussioni di questi due giorni, e possibilmente usare il presente per collegare quello che eravamo (come bibliotecari ed esperti nella cura della sapienza umana) a quello che possiamo diventare.

Il nostro mondo attuale

- viene da noi sempre di più vissuto attraverso computer e altre tecnologie (telefonini, iPad), tutti collegati alla rete

- è enormemente interattivo; tutti sono in grado di poter creare (magari solo una pagina su Facebook), di partecipare, e di essere visti e sentiti
- è un mondo pluralistico in termini di cultura, politica, ed economia. Nel mondo fisico, esterno alla rete, il potere è concentrato in una porzione pari all'1%, ma il blog di qualcuno che appartiene al restante 99% può attirare l'attenzione di milioni di lettori.

Le nostre risorse informatiche e culturali

- o sono nate digitali o stanno per essere digitalizzate
- sono relativamente accessibili sulla rete globale, senza confini geografici, ma sono anche costose da utilizzare, perché richiedono tecnologie avanzate (come dispositivi, reti, infrastrutture non solo di fili ed elettricità), e familiarità con queste tecnologie.

Gli utenti di oggi

- si aspettano di fare ricerche e interagire con l'informazione senza alcuna formazione, usando preferibilmente un singolo campo di ricerca
- interagiscono con la biblioteca attraverso software e hardware molto variabili, che non sono sotto il controllo della biblioteca.
- Per gli utenti di oggi "accedere" vuol dire "ottenere una copia"; e "ottenere una copia" vuol dire togliere le risorse dal contesto e dall'organizzazione intellettuale forniti dalla biblioteca o database o pagina web. Ogni utente si trova con un hard drive pieno di documenti, che non hanno alcuna connessione tra loro.

La comunicazione di oggi è ben diversa da solo due decenni fa:

- E' soprattutto a distanza, non faccia-a-faccia. Se osservi due ragazzi l'uno accanto all'altro che inviano SMS con i telefonini, c'è una buona probabilità che stiano chattando tra loro.
- E' rapida e breve; un libro richiede anni per essere scritto e giorni per essere letto; un blog richiede ore per essere scritto e minuti per essere letto; un Tweet richiede al massimo un minuto per essere scritto, e pochi secondi per essere letto.
- Tutti i nuovi stili comunicativi sono basati sull'interazione; il lettore può intervenire in un blog o rispondere a un Tweet, o anche commentare un articolo su un giornale online; il testo di un SMS è solo una frase nel flusso di una comunicazione continua. I giovani di oggi magari starebbero meglio con un discorso socratico che con un libro stampato, fisso, inerte.
- Mentre diminuisce l'importanza della parola scritta, specialmente su carta, cresce l'uso di foto e video, non solo come comunicazione di massa ma anche come comunicazione individuale, e non solo come intrattenimento, ma come mezzo fondamentale di formazione; invece del librone di istruzioni nella scatola del software di una volta, si trovano in rete dei video su "come si fa." Youtube riceve sessanta ore di video al minuto.



- Comunicazioni che in passato erano informali, come una chiacchierata in corridoio o una discussione in aula, sono adesso fissate nel formato digitale. Non solo, ma trattiamo queste comunicazioni come se fossero uguali a quelle più formali di una volta, utilizzandole, per esempio, in processi o anche per la ricerca scientifica.

Insomma, il mondo di oggi è in rete e interattivo; la comunicazione che è informale ma digitale fa parte della documentazione della nostra storia; il testo stampato è di un'altra epoca. Non sparisce, ma chiaramente non rappresenta una tecnologia moderna.

La rete ha cambiato ogni cosa.

Le biblioteche devono affrontare questo cambiamento; è questione di vita o di morte, di esistenza o di scomparsa. La centralità di un'istituzione fondata sulla civiltà pre-web è in dubbio e la sua sopravvivenza non è per niente scontata.

Ebbene, qual è oggi lo stato delle biblioteche?

Le biblioteche contengono un patrimonio culturale enorme. Organizzare, conservare, e mettere questi beni al servizio del pubblico è un lavoro complicato e costoso. Ma il problema di oggi nelle biblioteche non è solo la salvaguardia del passato; il presente ci pone di fronte a sfide enormi. Non solo per il fatto che ogni anno cresce il numero dei titoli di libri stampati, mentre le risorse finanziarie delle biblioteche diminuiscono, ma come l'esempio di Youtube ci insegna, ogni minuto un numero incalcolabile di nuove risorse si aggiunge alla nostra cultura digitale e nessuna di queste è sotto il controllo bibliografico delle biblioteche.

Se in altre epoche la biblioteca si poteva considerare la fonte principale di informazione, oggi tutto ciò non è più vero. Ma questo, in sé, non è un problema. Dovremmo essere felici per la crescita dell'informazione e del suo uso, e per le conseguenti possibilità di creare una società sempre più informata, sempre più rappresentativa di un popolo attivo, che fa sentire la sua voce.

Il problema, invece, è che le biblioteche sono lontane dalla principale fonte odierna di informazioni, cioè dal Web. La spinta a muovere le biblioteche verso i linked data non risponde semplicemente al desiderio di modernizzare i cataloghi, ma rappresenta la necessità di trasformare i cataloghi delle biblioteche da database chiuso e separato in un sistema integrato con la tecnologia che la gente utilizza per la ricerca e per la creazione di nuove idee. Insomma, il catalogo della biblioteca, per essere visibile all'utente di oggi, deve cessare di essere un'entità staccata, un database separato e divenire un insieme di dati, milioni di dati tra gli innumerevoli dati del web.

Il nostro compito oggi, come bibliotecari e informatici, non è di tradurre gli stessi dati bibliografici di oggi in linked data; il nostro compito è di creare un nuovo sistema di accesso e d'uso dei dati bibliografici che sia compatibile con il funzionamento della rete. Due sono gli aspetti di questa evoluzione. Il primo è quello di rendere utilizzabili i dati bibliografici nel web. Ogni persona che fa ricerca, che studia, che scrive e che cita, ha bisogno di dati bibliografici, alcuni dei quali possono essere forniti dai cataloghi delle biblioteche. E con la disponibilità dei dati bibliografici sul web, ogni persona connessa in rete diviene potenzialmente un "utente della biblioteca".

Il secondo aspetto riguarda l'uso di dati presenti in rete per migliorare i servizi della biblioteca. Collegando dati bibliografici e risorse web si può, ad esempio, collocare un libro nel suo contesto storico o dimostrare l'influenza di un autore sulla cultura del tempo.

A che punto siamo nella nostra esplorazione dei linked data per le biblioteche?

Progressi sono già stati fatti in alcune aree, come avrete modo di apprendere dai relatori a questo Seminario.

Ci sono due attività principali nella creazione di linked data: la prima è lo sviluppo degli elementi di metadati che si utilizzeranno per i dati, come ad esempio "autore" o "titolo"; la seconda, è la raccolta di liste controllate di termini che saranno usati come valori, come le lingue, i luoghi geografici, i nomi di persona.

Poiché nell'ambito bibliotecario gli standard di metadati definiscono già un numero di liste controllate di termini, questi sono stati abbastanza facilmente convertiti; insomma, il lavoro per creare sia degli schemi di metadati per i dati bibliografici che elenchi di vocaboli da usare in quei dati, è già piuttosto avanzato.

Il soggettario della Library of Congress è disponibile nella forma di linked data, cioè del semantic web; quelli delle biblioteche nazionali di Francia, Germania e Giappone sono analogamente disponibili. E tra loro hanno creato collegamenti che permettono di trattare lo stesso argomento, gettando così le basi per la creazione di un soggettario multi-lingue e internazionale.

I dati di autorità relativi ai nomi si trovano in un sistema chiamato il Virtual International Authority File, o VIAF. VIAF, che ha sede a OCLC, riceve i records da circa venti biblioteche nel mondo e raggruppa tutti i record relativi alla stessa persona, creando un'unica identità per quel gruppo di nomi collegati, quando è possibile, alla pagina Wikipedia intestata alla persona; in alcuni casi, il link è reciproco, da Wikipedia a VIAF. E' in corso un certo numero di esperimenti nella traduzione di schemi bibliografici tradizionali da biblioteche in linked data: in particolare, sono stati tradotti in standard semantic web ISBD, FRBR, FRAD, e RDA. Questi, però, non hanno ancora nessun collegamento con i dati del web semantico, e questo è un aspetto molto importante da sviluppare.

Una parte fondamentale del web semantico, che differisce in modo significativo dalle pratiche d'uso dei metadati del passato, è il collegamento, in particolare, tra elementi di metadati provenienti da comunità differenti. E' solo in questo modo, attraverso questo collegamento, che renderemo possibile la trasformazione dal mondo chiuso dei dati bibliografici provenienti dalle biblioteche al mondo aperto del semantic web.



Questo significa che abbiamo bisogno di creare connessioni tra i dati delle biblioteche e i dati che hanno origine in altre comunità e risorse, provengano esse dalla ricerca scientifica, da dati governativi, informazioni commerciali, o siano dati provenienti crowd-sourced. Dobbiamo capire una cosa fondamentale del web semantico: è un ambiente informativo altamente eterogeneo, sia nella sua estensione che nella sua qualità. Il mondo chiuso del controllo bibliografico, del quale abbiamo beneficiato fino ad oggi, non avrà spazio nel nostro futuro.

Per concludere...

Ovviamente bisogna chiedersi se i linked data risolveranno tutti i problemi che le biblioteche di oggi devono affrontare, e ovviamente la risposta è: no, naturalmente no. Ma non possiamo entrare nell'ambiente informativo, ricco e dinamico, del Web del XXI secolo con dati basati sui principi del XIX secolo.

È possibile – anzi, è *probabile* - che questo richieda cambiamenti profondi nei dati delle biblioteche per andare incontro alle esigenze di oggi, e che alla fine emerga una netta differenza tra il catalogo di oggi e l'accesso e la visualizzazione dei dati della biblioteca integrati con la rete.

Non ci serve più creare dati solo per le biblioteche; i nostri utenti non sono solo gli utenti del catalogo, ma tutti quelli che cercano informazioni e che creano risorse, ovunque siano, chiunque siano.

Dobbiamo essere non solo sul web, ma del web, utilizzando gli standard del web, la struttura del web, e i servizi e le applicazioni del web.

Il rischio è che cambieremo, ma non abbastanza.

L'intenzione originale di RDA era radicale: rompere con gli standard catalografici del passato e creare un nuovo modo di visualizzare i dati bibliografici, che fosse aperto, flessibile ed estensibile. Tuttavia, man mano che il lavoro sullo standard andava avanti molti nel settore cominciarono a mettere in discussione la nostra capacità di operare il cambiamento, e la commissione ha dovuto arretrare dalle sue posizioni per garantire che RDA si integrasse con i cataloghi esistenti. Sfortunatamente non fu fatta alcuna analisi su sistemi possibili per la trasformazione dei dati. Il nostro passato ci tiene ancorati sul posto, e stentiamo ad andare avanti, con la conseguenza che quando adotteremo RDA nel 2013, i nostri dati risulteranno pressoché indistinguibili dai nostri attuali cataloghi.

Non è solo il formato dei nostri dati leggibile dalla macchina che deve cambiare, ma anche il loro contenuto. Ricreare ISBD o MARC in linked data non ci darà rilevanza nel web.

Oggi, il record di un catalogo è essenzialmente un testo marcato che usa il linguaggio naturale per descrivere le risorse e che non è adatto all'elaborazione delle macchine.

Continuiamo a creare punti di accesso la cui funzione è rilevante per il catalogo lineare e

per l'ordinamento alfabetico – e questo non solo non è necessario, ma rende più difficile lo scambio di dati con altre comunità – comunità che organizzano i loro dati in modo da essere elaborabili dalle macchine.

Non possiamo continuare a vedere l'esito del nostro lavoro come un catalogo di biblioteca che assomiglia a quello di oggi. Non possiamo continuare a vedere il nostro catalogo come una meta che è separata dal web aperto. L'epoca del catalogo è ormai superata, come l'epoca della carrozza e del cavallo. Invece di insistere che i nostri dati non possono cambiare perché sono sempre stati così, dobbiamo rivolgere la nostra attenzione ai modi in cui possiamo riutilizzare questi dati, trasformandoli, usando sia il potere di calcolo che abbiamo oggi a disposizione che la potenza combinatoria del web.

Detto questo, concludo con un'esortazione a confrontarvi, nei prossimi giorni, con l'idea della biblioteca nel Web come un'ipotesi da esplorare, un futuro possibile, ma non l'unico, e di essere disponibili a considerare che i dati bibliografici avranno una forma completamente diversa da quella che hanno oggi, e che questo non porterà alla distruzione della biblioteca come la conosciamo, ma alla sua evoluzione.