

OSSERVATORIO

Old wine, new bottle? Principi e metodi per una reale innovazione nelle prospettive LIS.

Il parere di Marshall Breeding,
Mauro Guerrini, David Weinberger,
Paul Gabriele Weston, Maja Žumer

a cura del Gruppo di studio AIB
Catalogazione, indicizzazione, linked open data e web semantico (CILW)

Premessa

Questo quaderno è costituito da interviste a personalità che appartengono all'area della Library and information science (LIS), realizzate ponendo tre domande a cinque figure riconosciute a livello internazionale. Tramite questi saggi il Gruppo di studio AIB Catalogazione, indicizzazione, linked open data e web semantico (CILW) vuole avviare la propria attività partendo da un'attenta valutazione di teorie e principi sia 'classici' sia 'innovativi', e dei pareri degli studiosi su di essi. In tal modo, è possibile fare il punto sull'evoluzione delle attività di descrizione e organizzazione delle informazioni e delle risorse, per avviare una proficua discussione e un necessario confronto anche con discipline vicine alla LIS, partendo dagli studiosi e dai professionisti che più hanno presente la problematica del cambiamento, le cui attività di ricerca sono anch'esse legate al 'mutamento di paradigma' avvertito nelle nostre discipline.

Le domande messe a punto dal gruppo, dunque, hanno l'obiettivo di richiamare l'attenzione su alcune questioni fondamentali. Anzitutto, sarebbe utile chiarirsi le idee sul ruolo che può ancora avere il catalogo, nelle biblioteche e per la società: se continuerà a essere strumento centrale per le attività di ricerca, o se la sua posizione continuerà a scendere gerarchicamente fino ad allinearsi o passare in secon-

GRUPPO DI STUDIO AIB CILW (Catalogazione, indicizzazione, linked open data e web semantico), composto da Roberto Raieli, Marinella Cisternino, Elena Corradini, Enrico Francese, Antonella Iacono, Roberto Morelato, Antonella Trombone, sito web <www.aib.it/struttura/commissioni-e-gruppi/cnc>, e-mail cnc@aib.it.



do piano rispetto agli altri strumenti attuali di ricerca e scoperta dell'informazione. Il catalogo, dunque, si trova dinanzi al Web semantico, alla metodologia dei linked data, alla necessità di aprirsi ad altre comunità, per poter essere lo strumento preferito dagli utenti.

La poca efficacia dei cataloghi nella ricerca di informazioni è spesso dovuta alle trasformazioni degli oggetti della ricerca stessa, cioè le risorse della conoscenza: serve quindi interrogarsi anche su quale possa essere una definizione del nuovo oggetto del catalogo, che cosa sono di fatto le risorse della conoscenza, e quali strumenti sono più adatti a trattarle. Le risorse oggi utili alla conoscenza sono molto diverse tra loro e da quello che finora è stato inteso come 'documento', la loro qualità principale è l'interesse che ne hanno gli utenti, si inseriscono in molteplici reti di significato e appartengono a molteplici contesti.

Infine, indipendentemente dagli strumenti tecnologici più adeguati alle esigenze della società, è necessario chiarire le politiche e le metodologie auspicabili per il trattamento delle risorse, le informazioni e i dati, che restano l'oggetto primario di ogni strumento di organizzazione e gestione delle conoscenze: è sempre più determinate che i dati siano corretti e affidabili, e questa necessità cresce tanto quanto aumenta la loro libertà di crescita e diffusione. La descrizione dovrà essere decentrata, nutrendo fiducia nell'apporto di altri, favorendo l'integrazione nel Web semantico, ma senza rinunciare alla credibilità dei dati.

Nell'ottica delle possibili risposte, è sempre fondamentale mantenere un giusto equilibrio tra 'conservazione' e 'rivoluzione'. Visti da prospettive diverse, i principi di base, più o meno classici che siano, possono essere riconsiderati a livello teorico, nei punti di forza o nelle criticità, al fine di impostare un progetto che possa dirsi realmente innovativo dal punto di vista teorico e metodologico per le discipline che ci riguardano. Il vero progresso e la vera innovazione si ottengono non solo 'superando' i principi e le pratiche realmente obsoleti o allineati a realtà non più attuali – il vino diventato vecchio guastandosi –, ma anche sviluppando e 'ricomprendendo' quei principi stabili, estranei a condizionamenti temporali ed epocali, che restano sempre validi per il mantenimento e lo sviluppo delle discipline biblioteconomiche e della LIS – il buon vino invecchiato in solide botti.

Se i principi di fondo non sono adeguati, aggiornati e contemporaneamente durevoli, a nulla serve un nuovo contenitore, tecnologico o concettuale che sia, per tentare di proporli come vera innovazione metodologica.

Domande poste agli intervistati

Il catalogo e gli 'altri'

Si può affermare che il catalogo – almeno come lo conosciamo oggi – è destinato a perdere centralità tra gli altri strumenti di ricerca e conoscenza a disposizione delle biblioteche?

I nuovi strumenti di ricerca e recupero dell'informazione consentono di allargare la selezione e la mediazione ben oltre i classici confini del 'posseduto', verso la rete sterminata di informazioni e risorse accessibili tramite il Web. Se è vero, però, che non sempre è facile stabilire l'attendibilità e l'autorevolezza di molti dei suoi contenuti, è altrettanto vero che l'idea ispiratrice del Web semantico è legata alla creazione di repertori dai quali si possono attingere dati che provengono da fonti controllate e affidabili. In questa nuova dimensione, si può pensare a una sorta di catalogo 'unico' condiviso e partecipato, sviluppato *open source*, popolato e mantenuto da specialisti con nuove competenze, dove catalogazione e creazione di ontologie e dizionari si avvicineranno sempre più?

Un concetto attuale di risorsa

Al concetto di 'documento' si è fatto riferimento a lungo in ambito biblioteconomico, oggi gli *International cataloguing principles* (ICP) restringono questa definizione ad alcune risorse archivistiche. Inoltre, nell'attuale panorama culturale è necessario includere diverse tipologie di oggetti che possono essere registrati e, quindi, descritti e utilizzati come informazioni affidabili, possibili fonti di conoscenze. In questa prospettiva, quale potrebbe essere una valida e funzionale definizione del nuovo concetto di 'risorsa', e quale la sua identità?

In breve: qual è l'oggetto che dobbiamo trattare? Come possiamo prendere qualcosa che per natura è 'fluida' e oggettivarla in un record, un catalogo, un sistema di riferimento? Che cosa consideriamo come unità informativa da descrivere e collegare? Il *thread* di un forum, o un suo singolo post? Una sinfonia, o ciascuno dei suoi movimenti?

'Autorevolezza' dei dati

Con l'applicazione dei *linked open data* (LOD) ai cataloghi e agli altri strumenti di ricerca dell'informazione, le biblioteche e le altre istituzioni del patrimonio culturale dovranno abituarsi a non avere più il 'possesso' dei dati con cui descrivere e mediare le risorse. Ciò comporta un vero mutamento di paradigma che innova la tradizione documentale, e quella digitale nel complesso, a partire dalla definizione del concetto di informazione. In questo scenario, come potranno le diverse istituzioni culturali mantenere il ruolo di autorità nel processo di produzione, diffusione, conservazione e aggiornamento dei dati, nonché essere garanti della loro qualità?

Le diverse istituzioni, inoltre, riusciranno a ben integrarsi nel progetto del Web semantico sviluppando una più precisa interpretazione della propria identità?

Il parere di Marshall Breeding

Attraverso i linked data per la scoperta delle risorse

Per molti versi il catalogo ha già attraversato un grande cambiamento dalla sua forma tradizionale di strumento online focalizzato sul posseduto a stampa della biblioteca. Oggi le biblioteche si adoperano per fornire alle loro comunità servizi per una scoperta sempre più ampia delle risorse che riguardano l'interesse delle loro collezioni, e comprendono ogni tipologia di materiale. Diversamente dai cataloghi elettronici che restituiscono solo titoli di libri o di periodici, è fondamentale ora fornire anche l'accesso ai singoli articoli, ai capitoli dei libri, o agli oggetti digitali, e mostrare il full text o le rappresentazioni visuali quando è possibile. Questi *discovery service* devono essere più attenti alla pratica degli utenti, fornendo l'accesso al contenuto della biblioteca utilizzando nell'interfaccia alcune convenzioni e tecniche di ricerca alle quali il pubblico si è ormai abituato nella sua esperienza quotidiana di ricerca nel Web. Queste interfacce devono anche essere ottimizzate per i dispositivi mobili e i *tablet*, che rappresentano ormai la maggioranza degli utilizzi, più che gli schermi grandi dei *laptop* o dei computer *desktop*. La tipologia dei *discovery service* basati sugli indici ha trovato una grande accoglienza nelle biblioteche accademiche, dove la maggioranza li adotta in aggiunta – o in sostituzione – dei tradizionali cataloghi elettronici. Le biblioteche pubbliche richiedono *discovery service* o cataloghi che forniscano un accesso analogo sia agli e-book sia ai libri a stampa. Attraverso iniziative come ReadersFirst, le biblioteche pubbliche hanno richiesto – e i fornitori hanno offerto – metodi molto semplificati per scoprire, selezionare, e scaricare gli e-book nei dispositivi di lettura.

Queste nuove tipologie di interfacce per la scoperta rappresentano solo un versante dell'attività bibliotecaria. I linked data e le tecnologie del Web semantico hanno riscosso un interesse crescente. L'universo delle risorse di interesse per le biblioteche esposto sotto forma di linked data è cresciuto considerevolmente negli ultimi anni, a causa del fatto che la maggior parte del *corpus* dei contenuti scientifici rimane bloccato da editori proprietari. Le biblioteche hanno fatto importanti progressi verso il cambiamento dalla descrizione bibliografica orientata al record, primariamente attraverso i vari formati MARC, verso strutture basate sui linked data, specialmente BIBFRAME. Gli sviluppatori di molti sistemi di gestione per le biblioteche e *discovery service* hanno cominciato a lavorare per incorporare BIBFRAME, nonostante si rimanga in un periodo di sperimentazione di prototipi più che di prodotti operativi. Prevedo sistemi ibridi che useranno i database relazionali e gli indici per processi di *business* orientati alle transazioni, ma che attingeranno sempre più dall'universo dei dati collegati per completare la scoperta, l'accesso ai contenuti e la visualizzazione dei risultati.

Le tecnologie del Web semantico inoltre aiuteranno le biblioteche a migliorare la rintracciabilità delle risorse di interesse delle loro comunità di utenti. Dobbiamo riconoscere la realtà di molti utenti che non si rivolgono ai cataloghi o alle interfacce che le biblioteche offrono, ma si affidano invece ai motori di ricerca. L'incor-

porazione dei codici semantici nella presentazione delle risorse delle biblioteche, così come definito in *schema.org*, migliora fortemente la scoperta di risorse al di fuori delle interfacce delle biblioteche. Dal momento che gli utenti non si rivolgono alla biblioteca, la biblioteca deve lavorare duramente per rendere le risorse delle biblioteche scopribili e disponibili nei luoghi che essi frequentano nel Web.

Questa transizione dai cataloghi orientati al record verso le tecnologie del Web semantico non necessariamente significa un minor controllo per i bibliotecari nell'ecosistema dell'informazione. Le biblioteche continueranno a curare e descrivere le loro collezioni locali, ma utilizzando maggiormente le strutture dei metadati basate sui *linked data*, RDF, XML e utilizzando meno record e protocolli specificamente bibliotecari, di minore interesse per il più ampio ecosistema dell'informazione. Mentre il MARC ha servito bene le biblioteche, in qualche modo esso ha anche condotto a un certo isolamento rispetto agli editori, al settore dell'*e-commerce*, e ad altri settori dell'informazione. L'adozione del Web semantico potrebbe risolversi in un miglioramento della rilevanza e del posizionamento delle biblioteche. Ciò significa anche lavorare più che mai nel creare descrizioni per le risorse che si basino su fonti autorevoli, ma espresse sempre di più sotto forma di collegamenti e relazioni e meno come record autoreferenziali. Le funzioni principali delle biblioteche sono perdurate attraverso molte fasi della società e cicli tecnologici. Il modo in cui i bibliotecari svolgeranno questo lavoro si evolverà in base ai cambiamenti del contesto più ampio.

L'incredibile diversità delle risorse

Le collezioni delle biblioteche includono diverse tipologie di materiali. Mentre documenti a stampa, monografie, fotografie e manoscritti ne costituiscono una parte, le biblioteche collezionano anche molte tipologie di contenuti digitali. Le biblioteche accademiche, per esempio, generalmente spendono la maggior parte dei loro fondi destinati alle collezioni per l'accesso ai contenuti scientifici e professionali forniti tramite sottoscrizione di pacchetti di contenuti offerti da fornitori commerciali e non profit, e traggono vantaggio dalla crescente massa di documenti ad accesso aperto. Gli oggetti digitali formano l'altra componente vitale delle collezioni bibliotecarie. Molte biblioteche hanno programmi a lungo termine per la digitalizzazione di libri, giornali, manoscritti, fotografie, e altri oggetti di valore storico o scientifico. Il patrimonio culturale è rappresentato sempre di più sotto forma digitale nativa. I professionisti di biblioteche, archivi e musei, quotidianamente descrivono i materiali in tutte le tipologie di supporti e formati, traendo vantaggio dalla varietà di schemi e standard di metadati che vi si possono applicare. Qualsiasi definizione formale di cosa costituisca un documento o un insieme di principi di catalogazione deve comprendere l'incredibile diversità di materiali e le loro corrispondenti convenzioni per i metadati. Nessuna definizione univoca e statica di ciò che costituisce una risorsa bibliotecaria può necessariamente comprendere la gamma in continua espansione dei materiali che sono conservati e descritti dalle biblioteche e dalle altre istituzioni culturali.

Decentramento della descrizione bibliografica

Il crescente coinvolgimento delle biblioteche e di altre simili istituzioni nella creazione di strumenti basati sui *linked open data* rappresenta un momento positivo, dato che ciò fornisce maggiori opportunità per la scoperta di risorse di tipo bibliotecario. Questa transizione può significare un modello decentralizzato di descrizione bibliografica. Gli elementi della descrizione bibliografica potranno trovarsi in triple RDF invece che in record MARC immagazzinati in database centralizzati, ma lo stesso

tipo di sforzo intellettuale è richiesto per supportare la gestione e la scoperta delle risorse. Le biblioteche nazionali, i consorzi regionali e le organizzazioni come OCLC continueranno a giocare un ruolo chiave seguendo la transizione già anticipata verso i *linked open data*. Ulteriori attori potrebbero emergere. Il concetto di proprietà dei dati bibliografici potrebbe essere ulteriormente sminuito. Stiamo già vedendo molta più libertà di scambio e sempre meno asserzioni di paternità man mano che i dati bibliografici diventano più che una merce e crescono le aspettative di uno scambio collaborativo. Progetti come Europeana richiedono l'apporto di record bibliografici sotto la più libera licenza Creative Commons (CCo), indipendentemente dal fatto che derivino da OCLC o da altre fonti che precedentemente hanno dichiarato qualche grado di proprietà. L'eventuale transizione verso un maggiore coinvolgimento nei *linked open data* non necessariamente cambierà del tutto il ruolo delle biblioteche nel creare e descrivere le collezioni, ma si spera fornirà una serie di vantaggi. La creazione di metadati di alta qualità continuerà ad aver luogo in modi che potranno essere collaborativamente distribuiti nell'intera comunità bibliotecaria globale. Anche se i contenitori che trasmetteranno la descrizione bibliografica cambieranno drasticamente, le operazioni, le strategie e i valori delle biblioteche sopravvivranno nel tempo.

Il parere di Mauro Guerrini

I migliori strumenti per gli utenti

La questione si estende dal tema della centralità del ‘catalogo’ alla necessità di cataloghi ‘open source’ e di catalogatori con nuove competenze in grado di creare ontologie e vocabolari. Il ruolo centrale di un catalogo è indubbio in quanto strumento per sapere cosa una struttura possiede e quale sia la collocazione delle sue risorse, in modo che i suoi utenti (nel senso più ampio della parola) possano sapere cosa è disponibile e come recuperarlo (trovare, identificare, selezionare, ottenere... – le funzioni utente di FRBR). È un catalogo ‘open source’ quello che possiamo aspettarci per il futuro? Non è, invece, l’occasione per esplorare gli strumenti migliori per gli utenti – quale potrebbe essere il mezzo più adatto per connettere gli utenti a tutte le informazioni di cui hanno bisogno? – oltre i limiti dei motori di ricerca come Google, verso la ricchezza delle risorse che si trovano nelle collezioni e nelle istituzioni che vogliono condividerle.

Ritengo che non si debba confondere l’OPAC locale tradizionale (che è stato un’evoluzione del catalogo a stampa) con gli strumenti di ricerca integrati e sociali di nuova generazione, che ora cominciano a essere realizzabili. Ciò che si spera per il futuro, invece, è un cambiamento a partire dai cataloghi ‘autistici’ di molte biblioteche verso un ambiente più collaborativo. Credo sia necessario avere il coraggio di abbandonare i vecchi paradigmi catalografici, e credo ci si debba veramente appropiare del modello di catalogazione condivisa prefigurato dai linked data.

Vedo che il lavoro sul Web semantico e i linked data procede, e penso sia importante vederne i risultati, ma è importantissimo che tali risultati non siano inferiori a quelli che già si ottengono tramite la catalogazione tradizionale riguardo l’aiuto agli utenti nello scegliere e identificare che cosa sia disponibile.

Credo, inoltre, che l’ipotesi di un catalogo ‘unico’, o ‘condiviso e partecipato’, sia alquanto impossibile e forse non corretta. Ritengo, invece, che sia preferibile concentrarsi sull’idea di avere vari cataloghi, diversi per natura, tipo e ambito, ma collegati e uniti tramite l’uso di coerenti strutturazioni dei dati, il modello RDF appunto, almeno per ora.

L’oggetto di interesse per gli utenti

Come è possibile notare, gli ICP non usano mai il termine ‘documento’, né nel testo né nel glossario. Credo ciò sia dovuto principalmente al rischio di equivocare tra il documento come tipicamente inteso nella LIS, e il documento delineato in archivistica. Lo spirito degli ICP era essere molto più ampi (e più precisi) per includere tutte le possibili risorse – oggetti raccolti non solo da archivi, ma anche da biblioteche, musei e da tutte le altre istituzioni culturali e della memoria registrata. La traduzione del termine ‘documento’ in lingue diverse dall’inglese è essa stessa proble-

MAURO GUERRINI, Università degli studi di Firenze, Dipartimento di studi sul medioevo e il rinascimento, piazza Brunelleschi 3-4, 50121 Firenze, e-mail mauro.guerrini@unifi.it.

Vorrei ringraziare Barbara Tillett per i suoi preziosi suggerimenti; mi sono confrontato, inoltre, con Carlo Bianchini, Andrea Marchitelli, Dorothy McGarry, Valdo Pasqui e Tiziana Possemato.

matica, dato che così come utilizzato in altre lingue ha diverse connotazioni – altro motivo per gli ICP di rifiutare il termine.

Il senso più appropriato – il termine più universale e, si spera, meno ambiguo usato negli ICP – è ‘risorsa bibliografica’, che sicuramente include il problematico termine ‘bibliografica’, ma si deve osservare il suo porsi oltre l’indicazione del semplice libro per denotare qualsiasi risorsa raccolta dalle istituzioni culturali. Le prossime generazioni saranno, magari, in grado di trovare un aggettivo più adatto, più onnicomprensivo, che sia meglio compreso in futuro.

Che sia ‘risorsa’, o ‘risorsa bibliografica’, il termine sarà chiaro a tutti in futuro. Credo che questi termini siano stati un tentativo di disporre di una parola neutra che potesse essere utilizzata per comprendere nel modo più chiaro possibile tutti i tipi di materiale presenti nelle biblioteche e nelle altre strutture. La risorsa bibliografica è definita come «un’entità nel dominio della biblioteca e di analoghe raccolte costituita dal prodotto del lavoro intellettuale o artistico» (ICP Glossary). Un’aggiunta importante è che «le risorse bibliografiche nel modello FRBR sono le entità del Gruppo 1: opera, espressione, manifestazione e item», perché questo è il tratto di unione tra i principi ICP e lo standard internazionale RDA, in cui il termine risorsa perde l’aggettivo ‘bibliografico’ ed è utilizzato per identificare «un’opera, espressione, manifestazione o item» (RDA Glossary). Il termine comprende una singola entità e gli aggregati e i componenti di tali entità: «(per esempio, tre fogli di carte geografiche, una diapositiva singola pubblicata come parte di un set di venti diapositive, l’articolo di un fascicolo di una rivista accademica)». Con questa impostazione, e in linea generale, risorsa significa l’oggetto d’interesse per gli utenti; nella misura in cui si applica a biblioteche, archivi e musei, e a tutte le altre istituzioni culturali, risorsa è un termine molto ampio che ha lo scopo di includere libri antichi e moderni, documenti, oggetti e reperti di ogni tipo.

La fiducia nel lavoro altrui

La questione riguarda i LOD e la perdita di ‘possesso’, o la perdita del potere di autenticare i dati forniti agli utenti, e in tal modo bibliotecari e utenti devono aver fiducia in ciò che viene riportato loro. A prima vista ciò ricorda la situazione degli anni Ottanta, quando certi catalogatori sembravano più preoccupati della scarsa qualità della catalogazione realizzata da altri che dalla riflessione sul valore del fornire agli utenti l’accesso alle raccolte. Le esperienze dei primi database partecipati hanno evidenziato le imprecisioni nei record quando questi sono stati condivisi al di fuori del proprio catalogo – all’inizio soprattutto in OCLC. La percentuale di sbagli ed errori era in realtà molto piccola – errori umani, giusto quelli che si possono trovare quando si esamina un catalogo creato da molti nel corso di un lungo arco di tempo. La soluzione e l’orientamento sono stati di ‘correggere e basta’ gli errori scoperti, e nel tempo la qualità dei database condivisi è migliorata – ma i cataloghi non potranno mai essere perfetti, dunque è necessario lavorare per realizzarli sempre al meglio. Fare un lavoro di questo tipo in un ambiente condiviso è un notevole vantaggio – in quanto più persone possono contribuire. Se non si ha mai fiducia o apprezzamento per il lavoro altrui, non si potrà mai essere soddisfatti di alcuna soluzione condivisa. Purtroppo, capita di trovare ancora tale mentalità o atteggiamento in alcuni catalogatori.

Quanto alla questione se l’attuale tecnologia del Web semantico e dei *linked open data* sia (o possa essere) la risposta, essa è attualmente la tecnologia ‘migliore’ per la condivisione dei dati a livello internazionale, in grado di ridurre i costi delle singole biblioteche e di altre istituzioni con un impegno comune. L’ostacolo principale, a mio avviso, è la mancanza di un sistema internazionale condiviso, liberamente

accessibile e gestibile da tutti; forse sono ottimista, pensando che tali sistemi condivisi siano quelli giusti, ma credo che lavorare con la comunità dell'informazione che sta intorno alle biblioteche sia veramente importante in questo momento, facendo così in modo che tutti possano far crescere la qualità del servizio agli utenti.

Per il futuro ritengo che l'identità delle biblioteche dovrebbe evolversi da quella legata ai metadati codificati nel catalogo a quella delle istituzioni che forniscono e promuovono servizi a valore aggiunto per l'ambito di copertura e per gli enti collegati. In ogni caso, dobbiamo garantire la qualità del contenuto dei dati bibliografici. Se viene fornito un accesso non corretto, questo non aiuta l'utente. Credo nei servizi a valore aggiunto, e questi devono essere costituiti con dati precisi.

Dovremmo prendere atto della grande innovazione rappresentata da RDA, *Resource description and access*; le linee guida danno molta attenzione alla qualità dei 'dati' e delle relazioni tra le entità, piuttosto che enfatizzare il record. Il *focus* di RDA sulla descrizione e sull'accesso permette un'elevata attenzione all'identificazione e descrizione della singola entità, e insieme consente al catalogatore di porre quell'entità nel contesto dell'universo bibliografico tramite le relazioni e i punti di accesso. Quindi, utilizzare RDA e seguire i principi ICP su cui si basa dovrebbe diminuire il rischio di avere record di cattiva qualità (descrizioni lunghe, descrizioni troppo legate all'uso locale e scritte in linguaggi complessi), a favore della qualità del dato individualmente preso e, quindi, del suo riutilizzo (interoperabilità) in altri contesti o domini. In breve, il principio di RDA «prendi ciò che vedi» dalla risorsa è collegato con l'obiettivo di fornire una 'descrizione accurata'. Ciò si completa con il *focus* di RDA nel fornire 'accesso' tramite le persone, gli enti, le famiglie e i concetti (termini di soggetto controllati) associati o collegati, e le relazioni (soprattutto con le opere collegate). Credo sia necessario raccogliere le nostre energie e la nostra attenzione per la creazione di un accesso autorizzato (e preferito); attenzione che dovrebbe innescare un ciclo virtuoso per la realizzazione della qualità dei dati, riutilizzabili da tutti in qualsiasi contesto, e continuare a evidenziare il grande valore della catalogazione (che può continuare a chiamarsi tale o, se si preferisce, chiamarsi 'metadattazione') nel permettere agli utenti di connettersi alle risorse che desiderano e scoprire risorse correlate che possono essere loro utili.

Il parere di David Weinberger

Un catalogo evoluto verso il Web

I cataloghi tradizionali sono sempre stati troppo limitati per quello che riguarda la quantità di informazioni che sono in grado di ospitare, e hanno pertanto richiesto decisioni uniformi e centralizzate. Oggi sembra che tutto sommato ci stiamo allontanando da sistemi simili, e stiamo andando verso reti di informazione decentralizzate.

Da tempo è in corso un cambiamento nella maniera in cui consideriamo i metadati: di solito li abbiamo trattati come semplice oggetti-ombra attaccati all'oggetto 'reale' che descrivono. Invece adesso, specialmente con i linked data, abbiamo a che fare con nuvole informi di molecole di metadati, all'interno delle quali non è data alcuna priorità a un tipo particolare di oggetto inteso come quello 'reale'. Per esempio, se a una ricercatrice interessa un libro particolare, quel libro diventa il centro di una rete di informazioni collegate. Se le interessano le opere pubblicate in una determinata città in un determinato periodo, quelle diventano il punto intorno a cui si aggregano i dati. Se le interessano le opere che citano Michelangelo, allora quelle diventano il centro... finché il suo interesse si sposta.

Anche in tutto questo c'è ancora spazio per qualcosa di simile a un catalogo. Queste nuvole di linked data funzionano grazie all'esistenza di identificatori persistenti. Questi identificatori consentono di indicare gli oggetti all'interno della nuvola in maniera univoca. Senza tutto ciò, questa zuppa è un misero brodino, mentre noi abbiamo voglia di un bel minestrone denso e nutriente. Una lista di autorità può funzionare come una specie di catalogo, offrendo identificatori persistenti e fornendo delle informazioni base che possono presumibilmente essere utile a molti utenti.

Grazie a questi identificatori persistenti e affidabili, una zuppa di linked data può essere trasformata in un 'grafo', nel quale viene data prevalenza ad alcune entità in modo che le relazioni fra esse possano essere determinate e rese navigabili. Un grafo di tipo bibliografico comprenderebbe entità come opere, autori, date, luoghi, editori ecc. Un grafo di questo tipo potrebbe esser considerato un catalogo che si è evoluto verso il Web, nel quale lo scopo non è ridurre l'informazione su un'opera a un insieme essenziale di elementi, bensì arricchire quell'informazione mettendola in relazione con un contesto più vasto.

Le risorse in quanto parte di molteplici reti di significato

Anche nel mondo fisico un oggetto è una rete. Un libro è un libro solo in quanto parte di molteplici reti di significato. Vale a dire che non possiamo comprendere che questo oggetto è un libro se non sappiamo anche che c'è un contenuto bidimensionale al suo interno, che ci sono pagine disposte in sequenza anche quando il contenuto non è sequenziale, che il contenuto è stato assemblato da un umano, che il contenuto viene prima del libro, che il libro è opera di un lavoro manuale, che è stato pubblicato da un'organizzazione, che quest'organizzazione ha un indirizzo, che il libro ha un peso, un valore economico, deve

essere messo da qualche parte ecc. Queste reti di significato si intersecano fra loro e possono essere collegate con qualcosa all'esterno di loro stesse nel mondo, perché il mondo stesso è la totalità di queste reti di significato (tutto questo è preso pari pari da *Essere e tempo* di Heidegger). Come dice David Lankes¹, potremmo persino considerare le conversazioni che avvengono intorno a questi oggetti come oggetti di una catalogazione.

I cataloghi, tradizionalmente, hanno dovuto fare una scelta su quale fosse il loro oggetto principale – i libri o altri oggetti fisici. Questa scelta riflette un tentativo di anticipare il modo in cui la maggior parte degli utenti vuole esplorare gli oggetti della collezione. Riflette anche le difficoltà logistiche che gli amministratori della collezione devono affrontare, dal momento che devono trovare un posto per ciascuno di questi oggetti. Riflette, inoltre, il modo in cui sono strutturate le transazioni economiche: puoi acquistare un libro, ma non un autore o la città in cui ha sede un editore. Queste sono tutte buone ragioni per cui i cataloghi si sono strutturati intorno a degli oggetti fisici.

Ma nell'era dei computer, noi possiamo fare quello che gli sviluppatori di software chiamano 'binding dinamico'. Nel nostro caso significa lasciare che siano gli utenti del catalogo a scegliere di volta in volta quali sono gli oggetti primari nel momento in cui utilizzano il catalogo avendo un determinato progetto in testa.

Una struttura di dati come quella dei linked data rappresenta lo stato del sistema molto meglio di qualsiasi altra cristallizzazione di quello stato, sia che si tratti di un catalogo tradizionale basato sugli oggetti, sia che si tratti dello specifico angolo di osservazione di un utente.

È meglio lasciare che sia l'utente a decidere da sé – meglio nel senso che permette un maggior numero di utilizzi, compresi quelli che non sono stati previsti in anticipo.

Ovviamente costruire un simile 'catalogo assoluto' non è cosa da poco. Farlo richiederebbe comunque di fare delle scelte su quali siano i tipi di informazione e relazione che potrebbero essere utili per un utente. Richiederebbe di fare meno scelte in anticipo, il che aumenta il numero di scelte che possono essere fatte direttamente dall'utente.

Guadagnare credibilità

In questo nuovo ecosistema, le istituzioni possono scegliere fra controllare l'accesso o essere visibili. Dando per scontato che vogliono avere un'autorità che conti qualcosa – vale a dire, un'autorità visibile nella rete – potranno acquisire questa autorità non facendo i guardiani dell'accesso all'informazione, ma cercando di guadagnare credibilità.

Vista dalla prospettiva opposta, c'è anche la questione di come coloro che usano l'informazione possano fidarsi che i dati siano credibili. Nel momento in cui le triple linked data si liberano nel Web, possono anche perdere informazioni sulla loro provenienza. Ma in un mondo *linked open data*, dovrebbe essere sempre possibile per i ricercatori verificare la validità dei dati presso istituzioni affidabili, se non prendere i dati direttamente da queste istituzioni.

Il *library knowledge graph* di OCLC sta adottando una strategia promettente. Gli elementi atomici dei linked data idealmente sono degli URL (o URI, per essere precisi) che puntano a qualche informazione pubblica e affidabile. Gli elementi che vengono dai progetti OCLC punteranno alle risorse OCLC. Questi URI avranno come indirizzo base qualcosa tipo 'www.oclc.org': in questo modo, ogni computer o essere umano che li vede saprà immediatamente qual è la loro provenienza. Sapere la provenienza dei dati che si trovano in rete continuerà a essere un bisogno molto urgente per ogni ricercatore che si rispetti.

1 R. David Lankes; Joanne Silverstein; Scott Nicholson, *Participatory networks: the library as conversation*. 2006, <<http://quartz.syr.edu/rdlankes/Publications/Others/ParticipatoryNetworks.pdf>>.

Il parere di Paul Gabriele Weston

La misura del successo del catalogo

Negli ultimi anni si è più volte fatto ricorso all'espressione 'catalogo 2.0', ma quasi mai, temo, nel modo più appropriato. Il termine ha designato, di volta in volta, il catalogo arricchito con le riproduzioni delle coperte e di parte dei contenuti, l'esistenza di faccette (ma sarebbe più corretto definirle filtri di ricerca) come dispositivi per aggregare in vario modo descrizioni aventi caratteristiche comuni (la presenza di una stessa persona, l'uso di un certo formato, la natura bibliografica della pubblicazione e così via), la possibilità per l'utente di aggiungere commenti e valutazioni, di condividere descrizioni sulle reti sociali e, nel caso di e-book, di accedere direttamente ai contenuti. *Catalogue 2.0*¹ è anche il titolo di una raccolta di contributi prodotti da esperti internazionali del calibro di Lorcan Dempsey, Emmanuelle Bermès, Marshall Breeding e Karen Calhoun. La pubblicazione, curata da Sally Chambers, prende le mosse dall'interrogativo: ci sarà ancora un catalogo di biblioteca nel nostro futuro e se sì, quale sarà il suo aspetto?

Dal catalogo a stampa a quello elettronico, passando per gli schedari, l'applicazione delle tecnologie ha di volta in volta offerto nuove opportunità di ricerca ai lettori, al prezzo di una riscrittura almeno parziale dei codici catalografici. Come, tuttavia, osserva Dempsey, l'ambiente del Web richiede che il catalogo venga riconfigurato in forme tali da renderlo una componente non più singolarmente identificabile del servizio bibliotecario. Ciò è senz'altro conseguente al passaggio da una distribuzione fisica di informazioni e documenti a una digitale (*workflow switch*), ma riflette anche il trasferimento dell'interesse degli utenti dall'ambito locale (la 'mia' biblioteca, il 'mio' negozio, la 'mia' città, il 'mio' Paese) alla rete nel suo complesso (ci rivolgiamo a Google, ad Amazon, a Expedia ecc.) (*attention switch*). Le due pulsioni, combinandosi, tendono a spingere continuamente l'utente del catalogo verso l'esterno, esercitano in altri termini una sollecitazione centrifuga che tende decisamente a sottrarre allo strumento catalografico quella visibilità di cui ha goduto nel passato e a minarne al tempo stesso le funzioni più consolidate. Già nel 2006, riflettendo sul ruolo del catalogo nell'economia dell'informazione, Dempsey scriveva: «Libraries have rich deep collections, and the aggregate library system is a major achievement. However, in our current network environment, libraries compete for scarce attention. This suggests that if the 'library long tail' is to be effectively prospected then the 'cost' of discovering and using library collections and services needs to be as low as possible»². Dempsey pone dunque l'accento sulla questione della logistica, una componente dei servizi che nella complessa organizzazione della società contem-

PAUL GABRIELE WESTON, Università degli studi di Pavia, Dipartimento di studi umanistici, p.zza del Lino, 27100 Pavia, e-mail paul.weston@unipv.it.

1 *Catalogue 2.0: the future of the library catalogue*, edited by Sally Chambers. London: Facet, 2013.2

2 Lorcan Dempsey, *Libraries and the long tail: some thoughts about libraries in a network age*, «D-Lib magazine», 12 (2006), n. 4, <<http://www.dlib.org/dlib/aprilo6/dempsey/04dempsey.html>>.

poranea acquista una crescente importanza per la necessità di combinare domanda e offerta all'interno di un tessuto di un numero potenzialmente infinito di soggetti. I *link resolver*, che rendono possibile la fornitura di dati, informazioni e documenti attraverso una catena che non presenta, sul lato dell'utente, soluzioni di continuità apparente tra il momento della ricerca e quello della fruizione, possono senz'altro configurarsi come un esempio di logistica applicata al mondo delle biblioteche. In questa prospettiva il successo del catalogo si misura con la capacità di coordinare e aggregare più efficacemente, a beneficio degli utenti (e del proprio futuro), la domanda e l'offerta di servizi, e di farlo tenendo presente la necessità di mantenere quello standard di qualità alto che viene riconosciuto alla mediazione catalografica (e finora anche premiato), anche nell'epoca in cui altre tipologie di risorse hanno decisamente preso il sopravvento come strumenti primari della ricerca nel Web.

La scelta strategica della qualità

Personalmente ritengo che quella della qualità non sia un'opzione, ma una scelta strategica necessaria. Possiamo ricorrere alla metafora della Blue ocean strategy (BOS), una teoria economica illustrata in un libro del 2005 di W. Chan Kim e Renée Mauborgne³, la cui applicazione consiste nell'individuazione di spazi di crescita e nell'offerta di servizi non ancora coperti, o non coperti adeguatamente, uno scenario che si contrappone all'altro, quello in cui «the excess of players leads to a ruthless competition that will turn the red ocean bloody». In concreto, questa strategia comporta la rinuncia a confrontarsi sul loro terreno con i grandi *players* del Web, che fanno dei grandi numeri la loro forza, e di privilegiare criteri come l'affidabilità dei dati, puntando al tempo stesso all'integrazione dei dati e dei servizi offerti nei percorsi di ricerca per i propri lettori con risorse digitali complementari liberamente accessibili nel Web.

Con l'avvento del catalogo elettronico, originariamente concepito come uno strumento finalizzato al controllo amministrativo del patrimonio librario e alla gestione dei prestiti, la biblioteca ha ridotto la propria capacità di offrire percorsi non soltanto all'interno del proprio universo documentario, ma anche verso l'esterno (e non mi riferisco soltanto ad altri cataloghi), una capacità che nella biblioteca cartacea era garantita dalla disponibilità dei repertori e, nelle biblioteche più attrezzate, dalle capacità dei bibliotecari addetti al reference. La capacità del catalogo di fungere non soltanto da punto di arrivo, ma anche da snodo intermedio, il cui valore aggiunto consista anche nella selettività e nell'affidabilità delle risorse connesse, risponde al comportamento generalizzato dell'utenza del Web, che tende a concepire gli strumenti di ricerca e le risorse informative come parte di un unico ipertesto, più che come un dispositivo individuale finalizzato a fornire risposte secche.

Ciò detto, occorre aggiungere che potenziare il catalogo nella direzione indicata non è di per sé strategia sufficiente ad assicurare allo strumento un'adeguata visibilità. Per superare il pericolo dell'opacità sul Web occorre disporre di una massa critica di dati (di qui l'aggregazione in strumenti bibliografici di sempre maggiori dimensioni in cui dati e risorse vengono integrati in profondità e/o la realizzazione di sistemi indicali che puntano a favorire la massima interoperabilità con le risorse esterne) e godere al tempo stesso di un profilo specificamente definito, che non implica sempre l'esistenza di un universo documentale sterminato, ma premia la scelta

³ W. Chan Kim; Renée Mauborgne, *Blue Ocean Strategy: how to create uncontested market space and make the competition irrelevant*. Boston: Harvard Business School Press, 2005 (trad. italiana: *Strategia oceano blu: vincere senza competere*. Milano: Rizzoli ETAS, 2015).

di scandire un orizzonte ben circoscritto (come fanno Edit16 e più in generale i cataloghi relativi ad ambiti documentari fortemente specializzati). L'opportunità di fare sistema presuppone una capacità organizzativa, programmazione e cabine di regia che negli anni passati il nostro Paese non ha mostrato di possedere, una lacuna ancora più esiziale in considerazione della grande dispersione del patrimonio sul territorio e fra istituzioni di differente appartenenza. Ciò vale, a maggior ragione, anche per le iniziative di biblioteca digitale, che, almeno in questa fase, non potranno rivaleggiare con analoghe iniziative di Google books e delle più grandi istituzioni ricche di patrimoni storici e da più lungo tempo attive nel popolare di contenuti culturali il Web, ma dovranno senz'altro privilegiare la selettività nella programmazione dei documenti da digitalizzare.

Integrazione del catalogo nell'ecosistema digitale

Tornando alla questione del rapporto tra il catalogo e il Web, occorre riflettere sul fatto che, nonostante gli sviluppi tecnologici e i nuovi servizi prima ricordati, il catalogo rimane tuttora una struttura modellata sulla filosofia del Web 1.0, ossia uno scenario all'interno del quale i dati si muovevano univocamente dal centro alla periferia, con una netta divisione tra coloro che producevano informazione e coloro che ne fruivano. Al di là di poche e purtuttavia significative eccezioni, i cataloghi ancora oggi riflettono quell'architettura stellare, come dimostra anche il fatto che si sia per tanto tempo continuato a puntare su formati proprietari di ambito bibliotecario, impenetrabili ai motori di ricerca e inintelligibili in ambienti diversi, che hanno relegato gli archivi catalografici, anche i più sterminati, nell'oscurità del Web profondo. L'idea che ogni record possa avere una propria storia, fatta di scelte catalografiche discusse tra gli specialisti, di modifiche, accorpamenti, aggregazioni, trasferimenti da un archivio all'altro da parte dei dispositivi elettronici, di osservazioni e integrazioni prodotte dai lettori, e che questa storia diventi parte integrante del record accompagnandolo nei suoi percorsi, non ha ricevuto sin qui sufficiente attenzione. La realizzazione che, al momento, si avvicina di più a questo modello di catalogo è Open Library⁴, alla cui progettazione ha collaborato Karen Coyle. Le soluzioni adottate in Open Library potrebbero in qualche modo prefigurare il funzionamento di un sistema catalografico basato su BIBFRAME⁵.

L'idea che i lettori possano arricchire di contenuti l'archivio, contribuendo a migliorare la granularità delle informazioni e a esplicitare la natura delle relazioni che collegano le entità in esso descritte, è alla base del progetto Linked Jazz diretto da Cristina Patuelli⁶. Il *crowdsourcing* viene utilizzato per consentire ai cultori del jazz, una comunità in genere fortemente motivata e competente, di condividere testimonianze di incontri con musicisti (sotto forma di locandine, autografi, fotografie, audioregistrazioni) che possano contribuire a far emergere o a meglio definire le relazioni esistenti tra i musicisti stessi. In questo modo il progetto consegue almeno due risultati: da un lato i collegamenti, per lo più espressi da formule gene-

4 <<https://openlibrary.org/about>>.

5 Antonella Trombone, *Il progetto BIBFRAME della Library of Congress: come stanno cambiando i modelli strutturali e comunicativi dei dati bibliografici*, «AIB studi», 55 (2015), n. 2, p. 215-26.

6 <<https://linkedjazz.org>>. Leanora Lange; Maria Cristina Patuelli, *Linked Jazz: building with linked open data*, «EDUCAUSE review online», 30 giugno 2014, <<http://www.educause.edu/ero/article/linked-jazz-building-linked-open-data>>.

riche quali 'ha suonato con' oppure 'è stato amico di', vengono valorizzati da informazioni fattuali precise (hanno suonato insieme in quell'occasione, quando un musicista accorgendosi della presenza dell'altro lo ha invitato a suonare un certo brano); dall'altro lato, radicandosi nella comunità del jazz, il progetto ne guadagna in termini di identità, autorevolezza e vitalità.

Un altro servizio che potrebbe venire alimentato dall'integrazione nel catalogo di informazioni prodotte nell'ambito di sistemi diversi è lo *storytelling*. Il riferimento a uno fra i numerosi spazi di cooperazione tra le biblioteche e Wikipedia, una cooperazione cui si guarda con crescente interesse, ci porta a considerare la possibilità di valorizzare le relazioni, ad esempio, tra opere o tra autori integrando i rispettivi record di autorità nelle voci di Wikipedia che illustrano diffusamente tali relazioni, così che il lettore non si trovi di fronte a una serie di collegamenti insufficientemente illustrati, ma possa beneficiare di una presentazione articolata, dalla quale ottenere i riferimenti alle opere e qualche rinvenimento serendipico di tanto in tanto⁷.

Gli esempi si potrebbero moltiplicare, ciascuno a dimostrazione di quanto possa evolvere il catalogo in direzione di una sua più soddisfacente integrazione nell'ecosistema digitale. Un ultimo accenno lo riserverò al potenziamento del servizio del reference digitale. Qui il riferimento obbligato è quello a data.bnf.fr, lo strumento di corredo digitale creato dalla Bibliothèque nationale de France aggregando le informazioni disperse nei diversi cataloghi e all'interno della biblioteca digitale Gallica⁸. Il progetto, che è valso ai suoi ideatori lo Stanford prize for innovation in research libraries (SPIRL), presuppone l'utilizzo degli *open data* e l'adesione agli standard del Web. Analogo servizio offre, seppure in un contesto completamente diverso, Wikipedia *personensuche*⁹, un dispositivo finalizzato alla valorizzazione dei nomi di persone presenti nelle pagine di Wikipedia. Lo strumento si appoggia sulle aggregazioni di VIAF per individuare, oltre alle risorse rilevate in Wikipedia in corrispondenza di pagine dedicate a quel soggetto, altre risorse informative presenti nel Web e nei cataloghi delle biblioteche e delle reti aderenti a VIAF. Per ciascun soggetto l'utente dispone in tal modo di una lista di altri soggetti in vario modo collegati (parentela, professione, collaborazioni ecc.), nonché di un elenco di risorse (biografie, periodici, raccolte di partiture, di illustrazioni, cataloghi specializzati ecc.) che viene dinamicamente aggiornato e che offre una miniera di spunti per proseguire la ricerca e approfondire la conoscenza del personaggio.

Non posso non sottolineare, concludendo, quanto questi sviluppi siano debitori del lavoro di qualità svolto dalle biblioteche nel corso del tempo e della disponibilità di una mole enorme di dati strutturati. Questi dati svolgono la propria funzione, hanno un significato, se vengono presi non come monadi, prive di contesto, ma come componenti di quella struttura articolata e complessa che è la descrizione bibliografica. È questa struttura, maturata nell'esperienza secolare del bibliotecario, il valore aggiunto dell'informazione catalografica nel Web.

7 Un gruppo di lavoro (Camilla Fusetti, Agnese Galeffi e Antonella Trombone) sta valutando l'applicazione dello *storytelling* alla valorizzazione delle opere complesse nell'applicazione del modello FRBRoo.

8 <<http://data.bnf.fr>>. Romain Wenz, *Linked open data for new library services: the example of data.bnf.fr*, «JLIS.it», 4 (2013), n. 1, <<http://leo.cineca.it/index.php/jlis/article/view/5509>>.

9 <<https://tools.wmflabs.org/persondata>>.

Il parere di Maja Žumer

Aprire i dati ad altre comunità

La maggioranza dei bibliotecari, a questa domanda, direbbe, senza esitazione, che naturalmente il catalogo resterà così come lo conosciamo oggi. Questa forte convinzione è basata su secoli di monopolio da parte delle biblioteche in quanto principali fonti di informazione e del catalogo come loro principale strumento. Ma d'altra parte, sappiamo che i cataloghi delle biblioteche sono percepiti come poco intuitivi, confusi e difficili da utilizzare da parte degli utenti. I primi moniti in tal senso furono resi noti già negli anni Ottanta (per esempio in Borgman¹ e Matthews *et al.*²), quando iniziò a intensificarsi l'implementazione dei cataloghi automatizzati. In quegli anni la situazione non era ancora allarmante, perché gli utenti non avevano una sicura alternativa rispetto alle fonti di informazione. L'avvento del Web, dei *browser* e di altri strumenti hanno completamente cambiato l'orizzonte informativo, e i *report* attuali ci dicono che un'ampia maggioranza di utenti inizia a informarsi partendo da un motore di ricerca (e nessuno dal sito di una biblioteca)³ e che gli utenti evitano di utilizzare il catalogo anche quando sanno di voler prendere in prestito un libro da una biblioteca.

Le biblioteche per tradizione producono e aggiornano metadati di alta qualità sulle proprie risorse. I record bibliografici e di autorità creati dalle biblioteche sono senza dubbio i metadati migliori disponibili sulle risorse informative pubbliche e su altre entità collegate, come per esempio gli 'agenti'. Dunque, la domanda principale è perché questi dati non vengano utilizzati nel loro pieno potenziale. O, come la mia collega Tanja Mer un ha detto recentemente: «I nostri dati sono in vacanza. Dobbiamo farli lavorare di più!».

Per utilizzare appieno la ricchezza dei loro dati, le biblioteche dovranno promuoverli e soprattutto aprire i dati ad altre comunità. La tecnologia del Web semantico è l'infrastruttura tecnica necessaria, ma per usarla al meglio è essenziale disporre di modelli concettuali chiari. I modelli concettuali non chiariscono soltanto il dominio che modellano, ma permettono anche la reciproca comprensione di domini diversi. Si è aperta una finestra di opportunità per le biblioteche. Implementando modelli concettuali condivisi e aprendo i loro dati (attualmente in formati MARC di dominio specifico) ad altre comunità, le biblioteche favoriranno l'uso e il riuso di dati bibliografici. In questo modo le biblioteche riacquisteranno e manterranno la loro posizione di attori principali nell'era del Web semantico.

MAJA ŽUMER, Univerza v Ljubljani, Filozofska fakulteta, Aškerčeva 2/V, SI-1000 Ljubljana, e-mail maja.zumer@ff.uni-lj.si.

1 Christine L. Borgman, *Why are online catalogs hard to use? Lessons learned from information retrieval studies*, «Journal of the American Society for Information Science», 37 (1986), n. 6, p. 387-400.

2 *Using online catalogs: a nationwide survey*, edited by Joseph R. Matthews, Gary S. Lawrence, Douglas K. Ferguson. New York: Neal-Schuman, 1983.

3 Online Computer Library Center, *Perceptions of libraries, 2010: context and community. A report to the OCLC membership*. Dublin: OCLC, 2011, <<http://oclc.org/reports/2010perceptions.en.html>>.

La nozione generale di documento e il contesto

Poiché la missione principale delle biblioteche è raccogliere, selezionare, organizzare e fornire accesso alla conoscenza registrata, il 'documento' ne è il costrutto teorico centrale. Nel senso più generale, un documento è la rappresentazione registrata del pensiero.

Sebbene io non voglia qui discutere sull'esempio di Suzanne Briet, ovvero l'antilope come documento⁴, è chiaro che la nozione di documento è sufficientemente generica per comprendere sia i formati tradizionali che le risorse digitali, incluse quelle emergenti.

I *Functional requirement for bibliographic records* (1998)⁵, ovvero il modello concettuale dell'universo bibliografico, definiscono quattro classi di entità in base al livello di astrazione (*item*, manifestazione, espressione e opera). Per utilizzare un esempio semplice di una risorsa testuale: il mio esemplare di una particolare edizione del *Romeo e Giulietta* di Shakespeare è un *item*. La serie di tutti gli *item* identici è una manifestazione. Ma un libro non è soltanto un oggetto fisico, lo valutiamo per il suo contenuto. I segni, le parole e le frasi nel libro (in questo esempio, il testo originale di Shakespeare) sono un'espressione, e le idee rappresentate da questi segni sono l'opera. FRBR definisce inoltre gli agenti e le relazioni che connettono fra loro queste entità.

Negli ultimi anni molti studi sull'utenza (per esempio Pisanski e Žumer⁶) hanno confermato il carattere intuitivo di FRBR, fornendo prove di quanto tale modello sia un appropriato meccanismo di descrizione delle risorse, o documenti, in maniera tale da favorire l'esplorazione e consentire accorpamenti significativi. Il maggior peso dato tradizionalmente al supporto è sostituito da una visione più sofisticata ma al tempo stesso più flessibile. Il che mi riporta alla famosa discussione tra Lubetzky e Verona riguardante l'importanza relativa dell'unità bibliografica e letteraria, come vennero definite allora, quando l'utente aveva ancora a che fare con il catalogo cartaceo e le sue limitazioni. Poiché, in base al contesto, un particolare utente può essere interessato all'uno o all'altro aspetto, è importante considerare entrambi i punti di vista e presentare quello più appropriato in ogni situazione specifica. La tecnologia non è più un ostacolo e le biblioteche dovrebbero accettare FRBR quale base concettuale per la gestione delle risorse. Gli *International cataloguing principles* devono prendere in considerazione questa nuova concettualizzazione dell'universo bibliografico.

Integrarsi nel Web semantico

Esiste una sovrapposizione sostanziale sia nell'insieme degli utenti che nelle tipologie di risorse delle istituzioni culturali (biblioteche, archivi, musei), ma i tre ambiti, per ragioni storiche, hanno sviluppato ciascuno le proprie concettualizzazioni, strumenti e pratiche. Le tecnologie del Web semantico, per la prima volta, offrono l'infrastruttura che supera le differenze e che crea una visione comune delle rispettive risorse. Con modelli concettuali definiti in modo chiaro e standard non legati a un settore specifico (come i LOD), le implementazioni fra settori sono possibili

4 Suzanne Briet, *Qu'est-ce que la documentation?* Paris: EDIT, 1951.

5 International Federation of Library Associations and Institutions, *Functional requirements for bibliographic records: final report*. München: Saur, 1998.

6 Jan Pisanski; Maja Žumer, *Mental models of the bibliographic universe*, «Journal of documentation», 66 (2010), n. 5, p. 643-667, 668-680. *Id.*, *User verification of the FRBR conceptual model*, «Journal of documentation», 68 (2012), n. 4, p. 582-592.

senza costringere ciascuno di essi ad adottare un paradigma diverso. Anche se chiaramente cerco di fare il possibile affinché i dati delle biblioteche siano riutilizzati in altri ambiti, esiste pure un vantaggio evidente nel riutilizzare e integrare nell'ecosistema bibliografico i dati prodotti da altri settori. Per esempio, vari sistemi recenti di catalogazione bibliografica stanno già importando e integrando informazioni sugli autori da DBpedia, recensioni da Amazon, o aggiungendo dati forniti dagli utenti, come valutazioni e *tag*.

La domanda pertanto non è se le istituzioni culturali saranno capaci di integrarsi nel Web semantico – devono farlo. L'unica alternativa, temo, sarebbe l'oblio.

Articolo proposto il 17 luglio 2015 e accettato il 31 luglio 2015.

ABSTRACT AIB studi, vol. 55 n. 3 (settembre/dicembre 2015), p. 385-403. DOI 10.2426/aibstudi-11384

GRUPPO DI STUDIO AIB CILW (Catalogazione, indicizzazione, linked open data e web semantico), composto da Roberto Raieli, Marinella Cisternino, Elena Corradini, Enrico Francese, Antonella Iacono, Roberto Morelato, Antonella Trombone, sito web <www.aib.it/struttura/commissioni-e-gruppi/cnc>, e-mail cnc@aib.it.

Old wine, new bottle? Principi e metodi per una reale innovazione nelle prospettive LIS. Il parere di Marshall Breeding, Mauro Guerrini, David Weinberger, Paul Gabriele Weston, Maja Žumer

Il saggio è costituito da interviste a Marshall Breeding, Mauro Guerrini, David Weinberger, Paul Gabriele Weston e Maja Žumer, realizzate ponendo tre domande a queste figure riconosciute a livello internazionale. Tramite questo saggio il Gruppo di studio AIB Catalogazione, indicizzazione, linked open data e web semantico (CILW) vuole fare il punto sull'evoluzione delle attività di descrizione e organizzazione delle informazioni e delle risorse, per avviare una proficua discussione e un necessario confronto anche con discipline vicine alla LIS.

Le domande proposte dal gruppo hanno l'obiettivo di richiamare l'attenzione su alcune questioni fondamentali: anzitutto, il ruolo che può ancora avere il catalogo dinanzi al Web semantico, alla metodologia dei linked data, alla necessità di aprirsi ad altre comunità, per poter essere lo strumento preferito dagli utenti; di seguito, quale può essere una definizione del nuovo oggetto del catalogo, le risorse della conoscenza, molto diverse tra loro e dal classico 'documento', identificate in base all'interesse degli utenti, e che si inseriscono in molteplici reti di significato e in molteplici contesti; infine, quali sono le politiche e le metodologie auspicabili per il trattamento delle risorse, le informazioni e i dati, se la descrizione dovrà essere decentrata e condivisa, favorendo l'integrazione nel Web semantico, ma senza rinunciare alla credibilità dei dati.

Se i principi di fondo non sono adeguati, aggiornati e contemporaneamente durevoli, a nulla serve una nuova metodologia per tentare di proporli come vera innovazione.

Old wine, new bottle? Principles and methods for a true innovation in LIS perspectives. The view of Marshall Breeding, Mauro Guerrini, David Weinberger, Paul Gabriele Weston, Maja Žumer

This article consists of interviews with Marshall Breeding, Mauro Guerrini, David Weinberger, Paul Gabriele Weston e Maja Žumer, realized asking three questions to these internationally recognized scholars. Through this essay, the AIB Study group on Cataloging, indexing, linked open data and semantic web (CILW) wants to review current development of information and resources description and organization, in order to generate a fruitful debate and a necessary comparison also with disciplines neighbor to LIS.

Questions developed by the group are intended to call attention to some fundamental questions: first, the role a catalog may still have in front of the semantic Web, the linked data methodology, and the need to

open up to other communities, in order to be the preferred tool for users; second, what might be a definition for the new object of the catalog, the resources of knowledge, very different from each other and from the classical 'document', identified according to users' interests, and that are placed in multiple webs of meaning and multiple contexts; finally, what are desirable policies and methodologies for the treatment of resources, information and data, if description should be decentralized and shared, fostering integration in the semantic Web, but without losing the data credibility.

If the underlying principles are not adequate, up-to-date and at the same time durable, it's no use proposing a new methodology to try to present them as a real innovation.