

# La metadatazione: cos'è?

DOI: 10.3302/0392-8586-202203-021-1

A cura di  
**MAURO GUERRINI**  
mauro.guerrini@unifi.it

*Una discussione con Stefano Bargioni e Alberto Gambardella (Pontificia Università della Santa Croce), Giovanni Bergamin (AIB), Carlo Bianchini (Università di Pavia), Roberto delle Donne (Università di Napoli Federico II), Pierluigi Feliciati (Università di Macerata), Claudio Forziati (Università di Napoli Federico II), Maurizio Lana (Università del Piemonte Orientale), Rossana Morriello (Politecnico di Torino), Valdo Pasqui (Università di Firenze), Tiziana Possemato (Università di Firenze), Roberto Raieli (Sapienza Università di Roma), Maurizio Vivarelli (Università di Torino). Le risposte di alcuni riguardano solo alcune domande.*

La metadatazione, in ambito biblioteconomico, è la dimensione assunta dalla catalogazione nell'era digitale e, come la catalogazione, si riferisce al processo di registrazione dei metadati, ovvero dei dati funzionali all'identificazione e al reperimento di una risorsa. Essa costituisce una realtà da almeno quindici anni nel contesto bibliotecario nordeuropeo, statunitense, canadese, australiano e non solo. La nostra discussione desidera chiarire le finalità principali di questa prospettiva dando un contributo rivolto a rimuovere le tentazioni

di facili luoghi comuni che ne falsificano il significato. I metadati sono usati in ogni campo della conoscenza e delle attività umane; tendono a descrivere l'insieme dei *real world objects*, degli oggetti del mondo reale che possono far parte di o essere riferibili a una collezione analogica e digitale: una persona, un'opera, un oggetto d'arte, un concetto ecc. Da qui discende una visione estesa dell'ambito d'applicazione della metadatazione, che abbraccia tutti i contesti (*cross-domain*) necessari ad accompagnare l'esperienza della ricerca. Nel web semantico, infatti, non esiste distinzione tra dati bibliografici e dati di altra natura, ma tra dati collegati (o connessi) e dati non collegati; nel web dei dati, infatti, esistono solo dati condivisibili, modulari e riutilizzabili, che possono essere creati da vari agenti. Il contesto digitale favorisce la collaborazione tra molteplici soggetti, è partecipativa, si basa su dati riutilizzabili in domini diversi: biblioteche, archivi, musei, ambiti amministrativi, medici, legali ecc. I dati o insiemi di dati possono essere catturati e riutilizzati. Essi, tramite procedure differenti, si arricchiscono e si modellano in forme articolate, mantenendo una propria identità che prescinde dallo speci-

fico dominio di utilizzo. Ciò significa che lo stesso dato, la stessa informazione, se ben costruita, può essere utilizzata indifferentemente sia nel contesto bibliografico, sia in altri contesti apparentemente molto distanti; per esempio: il luogo geografico desunto da *GeoNames*, che nasce in un ambito geografico, ed è utilizzato come luogo di pubblicazione di una risorsa bibliografica.

Il web, e in particolare il progetto del semantic web, ha un'essenza democratica corroborata dalla possibilità di scambio di dati, includendo nel processo di metadattazione la pubblicazione e il riuso dei dati con la tecnologia LOD (Linked Open Data) indirizzata agli utenti e agli agenti software. Il passaggio dal record ai dati è segnato proprio dall'introduzione del paradigma dei linked open data, che ha generato un notevole cambiamento di prospettiva e che sta ridisegnando i confini del concetto di catalogazione. Dalla centralità del record si è passati alla centralità del dato e, pertanto, dalla gestione dei record si è passati alla gestione dei dati. Non si ha più la redazione di record strutturati secondo aree prestabilite con dati contrassegnati da una punteggiatura convenzionale, come previsto da ISBD (lo standard più longevo nella storia della catalogazione), bensì si assiste a un processo d'identificazione e collegamento di dati relativi a un'opera, a un autore, a un soggetto ecc. Si tratta di un procedimento che privilegia le relazioni tra entità tramite i loro attributi. Entità, attributi e relazioni sono al centro dei modelli concettuali, da FRBR a IFLA LRM. Il procedimento di metadattazione, dal punto di vista dei principi e degli standard, non cambia molto. RDA ne è un esempio: le sue istruzioni non differiscono sostanzialmente da AACR2. La metadattazione, piuttosto, corrobora il rigore e l'analisi bibliografica necessari per una descrizione di qualità, nel rispetto filologico dei dati così come si trovano sulla fonte d'informazione, senza alcuna manipolazione. Contemporaneamente la metadattazione conferma la necessità del possesso da parte del bibliotecario di un'adeguata preparazione culturale, nonché di abilità diversificate legate alle tecnologie dell'informazione. La metadattazione, infatti, è sia un'attività intellettuale, sia un processo altamente tecnologico, tipico dell'era digitale in cui viviamo.

***Ripensare criticamente la tradizione catalografica serve a capire cosa nelle procedure tradizionali vale la pena di continuare e di sviluppare e cos'è stato a suo tempo rivoluzionario, ma ha esaurito la sua "spinta propulsiva". Come motivereste, in poche parole, il passaggio dalla catalogazione alla metadattazione? Quali sono i vantaggi della metadattazione per il lettore affinché trovi ciò di cui ha bisogno?***

*Stefano Bargioni e Alberto Gambardella*

Indubbiamente nel corso degli ultimi vent'anni il mondo dell'informazione, veicolato dal web, è stato investito da cambiamenti epocali: lo spazio informativo si è dilatato a dismisura e il tempo, quello delle condivisioni plurali e simultanee dei collegamenti alla rete, si è accorciato. Il cassetto dello schedario cartaceo, un tempo fruibile da un solo utente alla volta e che costituiva l'immagine emblematica della biblioteca, oggi si è allungato, dilatato, frammentato ed è fruibile da più utenti contemporaneamente. La quantità di informazioni è aumentata e si è distribuita spesso in maniera caotica e disordinata. In questo contesto la metadattazione fornisce un nuovo modello che, nel solco della tradizione bibliotecaria, guida gli utenti nell'universo dei dati e rende possibile l'individuazione della risorsa ricercata. Per attuare tale modello occorre creare relazioni (connessioni) attraverso una nuova registrazione delle informazioni: metadattare, appunto. La metadattazione è pertanto figlia dell'onnipresenza del web e la risorsa catalografica descritta si "affaccia" sul web e si "aggancia" al web proprio tramite la rete delle relazioni. La rilevanza assunta dalle entità/attributi/relazioni genera la possibilità di offrire non solo punti di accesso ma anche nuovi percorsi di accesso. Se per esempio creiamo un collegamento all'interno di un record di autorità persona, tramite un URI, al VIAF (campo 024 del formato MARC) rendiamo disponibile all'utente, in maniera potenzialmente esaustiva, la produzione editoriale della persona.

*Giovanni Bergamin*

Per rispondere si può partire da una definizione di metadati. Mettendo insieme due definizioni – quella di Priscilla Caplan (*Metadata fundamentals for all librarians*, 2003) e quella di Karen Coyle (*Metadata: data with a purpose*, 2004) – possiamo definire i metadati come informazione strutturata secondo un determinato modello (o schema) con l'obiettivo di facilitare una determinata attività umana. Parlare di metadattazione da un lato valorizza tutta la nostra tradizione e dall'altro ci porta a lavorare in maniera più efficiente in un territorio (l'informazione in rete) condiviso con altri domini.

*Carlo Bianchini*

Il prodotto materiale che risulta dai due processi della catalogazione e della metadattazione è diverso. Nel primo caso il risultato è un catalogo, cioè un elenco ordinato secondo un preciso principio, il principio in base al quale il catalogo viene costruito: per autore, per titolo, per classe, per soggetto, topografico e che presenta informazioni strutturate in una sequenza predefinita, secondo gli scopi della descrizione stessa. In questo

approccio, il catalogo a libro e quello a schede doveva essere costruito secondo una precisa – e di volta in volta unica – modalità di accesso e di presentazione delle informazioni. Tale modalità rifletteva in modo molto evidente le scelte del catalogatore. Con il passaggio al catalogo elettronico, le modalità di accesso sono poi confluite in un unico strumento, in grado di presentarci le descrizioni bibliografiche prestrutturate, ma secondo un ordine determinato con maggiore flessibilità. Anche in questo caso, comunque, il numero delle modalità con cui si accede è sempre stabilito a priori dal catalogatore che progetta il catalogo e, per quanto ampio, è sempre limitato. Nella metadattazione, il prodotto sono i metadati sui medesimi oggetti descritti dal catalogo, ma le modalità di accesso alle descrizioni e la struttura della presentazione dei dati sono aperte, possono essere definite a posteriori. Sicché dai metadati possiamo ottenere i medesimi ordinamenti e la medesima presentazione strutturata delle informazioni che si avevano nel catalogo a stampa, a schede o elettronico, ma anche molto, molto di più. Per esempio, nel catalogo elettronico sono registrati spesso i dati di nascita e di morte e la professione di un autore, o il luogo di stampa delle risorse, ma è impossibile, sul piano pratico, sia per il lettore che per il bibliotecario, rispondere a domande ormai legittime come “Posso avere un elenco dei giornalisti nati a Parigi e morti a Mosca?” oppure “Quali opere di autori nati a Venezia nella prima metà dell’Ottocento sono state pubblicate in Germania?”. Con la metadattazione, gli stessi dati che si utilizzano oggi per produrre il catalogo elettronico consentono di fornire gli accessi “tradizionali” ma anche di *esplorare* gli stessi dati in modo molto più ampio e flessibile. Bisogna sottolineare che in questa risposta si sono evidenziati soltanto i vantaggi dei metadati catalografici “a uso interno” dell’istituzione ma non di molti altri vantaggi dei metadati quali identificazione, arricchimento e integrazione, qualità del dato, indipendenza dal software, accessibilità, condivisione e interoperabilità, decentramento, riuso, visualizzazione e analisi.

#### *Pierluigi Feliciati*

Credo, da non specialista di catalogazione bibliografica, che l’intuizione davvero rivoluzionaria sia stata quella di Ranganathan quando ha affermato che la biblioteca debba essere un *growing organism*. Se il libro – la risorsa, diremmo oggi – deve essere a disposizione del lettore è necessario che il catalogo risponda alla *metanoia*, ovvero all’evolversi tanto delle domande e dei comportamenti informativi dei lettori quanto delle tecnologie di registrazione, di ricerca e di accesso della conoscenza. La tradizione catalografica ha contribuito a definire il catalogo come strumento di mediazione tra libro, bi-

bliotecario e lettore. Il catalogo è diventato man mano anche uno strumento di gestione della biblioteca, occasione di cooperazione tra bibliotecari, ha aperto le finestre delle sale di lettura sul paesaggio variegato e globale della conoscenza. L’esplosione della *scheda*, della registrazione bibliografica in entità-in-relazione è stata motivata dalla provocazione intellettuale che il World Wide Web ci ha posto e ancora ci pone: uscire dalla descrizione/narrazione come opera chiusa e accettare la sfida dalle connessioni significative tra dati aperti. Il catalogo, dunque, deve costituire sempre di più un punto di partenza verso altri cataloghi, altri registri d’autorità, altri dati e altri ambiti informativi, offrendo al lettore non solo il suo libro, le edizioni dei libri di un autore, ma il rapporto tra libro, autore, soggetto, editore, luoghi, tempi con tutti gli altri possibili contesti.

#### *Maurizio Lana*

Il passaggio alla metadattazione è correlato alla comparsa di risorse informative digitali. Catalogare le risorse informative digitali significa concepirle come corrispondenti, equivalenti, alle risorse a stampa. Cercare nel digitale, che è nuovo, la corrispondenza con l’analogico, con il fisico, che è noto, è un modo per sterilizzare la novità che il digitale può esplicitare; ma è anche un modo per cercare di rendere fruibile il digitale anche nei contesti in cui non potrebbe esplicitare in pieno le sue potenzialità. Un segno del cambiamento è la metadattazione. La metadattazione è vantaggiosa per il lettore perché il suo orizzonte applicativo è più ampio: complicato *catalogare* un post in un blog; un post in un blog scientifico è una risorsa informativa che può avere lo stesso valore informativo di un articolo scientifico (anche se con differenti percorsi di convalida) ma catalogarlo come oggetto digitale non è agevole. Metadattarlo, invece, è possibile e soprattutto la metadattazione permette di inserirlo in una rete di riferimenti e di rimandi *in primis* attraverso il nome/i nomi degli autori, le keyword, le fonti citate. Per la conoscenza scientifica il senso della rete degli studiosi è sempre stato presente (la *res publica litterarum*, gli epistolari degli studiosi, le lettere pubblicate dalle riviste scientifiche, sono solo alcuni spunti sul tema) ma questa rete non è mai stata pienamente e agevolmente fruibile prima che le risorse informative e la rete di rimandi che esse contengono entrassero nel mondo digitale: nel mondo digitale leggo un articolo A, esso cita e linka un altro articolo B, interrompo la lettura di A, vado all’articolo B, lo apro, lo leggo, esso linka C che decido di leggere, torno ad A, riprendo la lettura, e così via. In un articolo a stampa di cinquant’anni fa questa rete di connessioni citazionali era (è) ugualmente presente, esplicita e percorribile, ma richiede azioni esplorative (la ricerca nel web con

un motore di ricerca). La metadattazione permette operazioni simili ma senza l'incertezza dell'esplorazione, un'incertezza che può apparire indesiderabile ma che forse potrebbe non esserlo così tanto se in essa è insita la possibilità della scoperta imprevista.

#### Rossana Morriello

Non credo si tratti di un passaggio rivoluzionario quanto di una trasformazione piuttosto lineare e naturale. La descrizione catalografica è di per sé una forma di metadattazione, ovvero i dati catalografici sono "dati funzionali all'identificazione e al reperimento di una risorsa". Come si potrebbe trovare un libro in biblioteca senza i metadati forniti dal catalogo? La funzione della metadattazione in ambiente digitale è analoga e ancora più importante, con le ovvie differenze indotte dai cambiamenti del contesto documentario. In primo luogo, la struttura dei metadati catalografici nella catalogazione tradizionale segue regole rigide e concettualmente nate per l'ordinamento di un catalogo a schede cartaceo. Ma già con il web, ancor prima che con il web semantico, la struttura della scheda catalografica risultava inadeguata, o meglio non in grado di cogliere le potenzialità delle macchine alle quali poco importano l'ordinamento pre-stabilito dalle regole di catalogazione per determinare il punto di accesso "fisico" o la punteggiatura specifica per separare i campi secondo gli standard biblioteconomici. Il web semantico ha indotto un ulteriore passaggio nel livello di granularità e fluidità delle risorse documentarie. Sono l'evoluzione del web e il numero enorme e crescente di dati di cui disponiamo oggi a rendere indispensabile la metadattazione. I vantaggi per il lettore sono molteplici come l'accresciuta possibilità di ottenere il risultato cercato anche da fonti che non aveva considerato, la possibilità facilitata di uso e riuso dei dati che le strutture di metadattazione consentono. Tutto ciò a condizione che i metadati siano di qualità. La qualità dei metadati aumenta richiamo e precisione della ricerca e le possibilità di riuso del suo risultato. Ed è proprio questa la sfida per i bibliotecari, ovvero creare metadati di qualità.

#### Valdo Pasqui

La descrizione di opere prodotte dall'ingegno umano attraverso dati che ne colgono e codificano alcune proprietà e la creazione di cataloghi per agevolarne il reperimento sono una delle attività più antiche delle biblioteche (cfr. per esempio, Irene Vallejo *Papyrus. L'infinito in un giunco*, Milano, Bompiani, 2001). Il concetto di metadati (dati sui dati) e la metadattazione possono essere ricondotti all'aspirazione umana di descrivere e ordinare i prodotti dello scibile. Questa attività ha avuto un impulso significativo grazie allo sviluppo della Rete

e del web; infatti, ogni oggetto digitale e ogni soggetto (umano o no) a esso collegato necessitano di essere opportunamente descritti per consentirne il reperimento. Dunque, a partire dalla fine degli anni Novanta del secolo scorso si è assistito alla proliferazione dei set di metadati. Spesso uno dei limiti di queste iniziative, nate in contesti culturali, scientifici ma anche aziendali e amministrativi molto eterogenei, è stata la reciproca incomunicabilità e la mancanza di condivisione. Il web semantico e i LOD sono volti ad assicurare l'interoperabilità tra domini diversi e prevedono la condivisione di set di metadati riusabili in contesti diversi senza dover ogni volta reinventare la ruota. In questa visione la "tradizionale" catalogazione delle biblioteche (ma le stesse considerazioni potrebbero valere per l'ambito archivistico e museale) va vista ormai come una parte significativa del più ampio/trasversale concetto di metadattazione (concetto che potremmo aggettivare come "universalistico") al quale contribuisce con standard e buone pratiche ormai consolidate ma che deve anche aprirsi ai contributi migliorativi ed evolutivi provenienti dagli altri contesti (per esempio, set di metadati) senza preclusioni e condizionamenti ideologici o elitari ed evitando di chiudersi in una nicchia.

#### Tiziana Possemato

Catalogare significa metadattare: significa individuare, selezionare, strutturare dati (attributi) sui dati (titolo, autore ecc.) al fine di rappresentare una risorsa, esprimendone le caratteristiche fisiche, contenutistiche e di relazione in modo da renderla identificabile e, dunque, fruibile, a chi la cerchi. Ma il termine metadattazione assume una propria fisionomia fino al punto da sostituire lo stesso termine catalogazione quando il contesto di applicazione si espande, sia in termini di *domini* e *comunità di fruitori*, sia in termini di *funzioni* e *scopi* dei metadati stessi:

- l'inclusione nel catalogo di materiali differenti rispetto a quelli tradizionali, dai video alla musica, dalle mappe ai file digitali ecc.;
- i progetti di conversione retrospettiva e di migrazione di dati tra sistemi differenti, con l'applicazione o la creazione ex-novo di mappature tra formati di dati diversi;
- l'esplosione del materiale digitale con la moltiplicazione di dati e metadati extra-bibliografici, tra cui i metadati gestionali-amministrativi;
- il consolidamento di pratiche e servizi collaborativi tra istituzioni diverse, tra cui il copy cataloging, la creazione di cataloghi unici, i discovery system, il prestito interbibliotecario, il document delivery, tutti fattori che hanno moltiplicato i formati di scambio e i protocolli di dialogo;
- l'evoluzione dei gestionali di biblioteca verso siste-

mi integrati (LMS, ILS, LSP), che hanno imposto una gestione dei metadati che va ben oltre quelli catalografici, includendo dati sulla conservazione, sull'acquisizione, sulla circolazione ecc.

Per una panoramica sull'universo dei metadati nel solo ambito culturale si veda l'ottima rappresentazione di Jenn Riley, *Seeing Standards: A Visualization of the Metadata Universe*, del 2018. L'estensione verso il web semantico, con l'adozione dei linked open data, ha amplificato questo fenomeno già in atto, portando il catalogo in una dimensione non più circoscrivibile a quanto tradizionalmente identificabile con i dati bibliografici o catalografici.

#### Roberto Raieli

La convergenza delle risorse della conoscenza verso il digitale e il multimediale è oramai realtà insostituibile del mondo dell'informazione, e proprio tale sviluppo tecnologico recupera e conferma il ruolo delle biblioteche in quanto luogo dove la convergenza può avvenire pienamente e può essere ben gestita. Le biblioteche sono da sempre aperte alle innovazioni e alla gestione dello sviluppo dei diversi supporti dell'informazione, sempre attente alle novità prospettate e alla ricerca del modo migliore per sviluppare la propria missione adeguandola all'epoca in cui vivono. Nell'epoca delle risorse convergenti, dove la struttura digitale e la forma multimediale diventano caratteristiche comuni a ogni genere di risorsa informativa – e sbiadiscono anche basilari distinzioni quali quelle tra testo e immagine, o tra supporto cartaceo e supporto magnetico –, ogni occasione è buona per ribadire la necessità della metadattazione come competente attività di mediazione, differenziazione, selezione e scelta tra tutto quanto “scorre” nei flussi digitali dei dati. Le biblioteche possono “riposizionarsi” nel nuovo scenario dell'informazione senza forzature, ridiscutendo la forma e la funzione dei propri strumenti principali di mediazione, e dei servizi collegati. Attraverso una “sana” evoluzione, gli strumenti di mediazione della biblioteca potranno efficacemente collocarsi nell'ambito del web semantico, e mantenere la collocazione autorevole che hanno sempre avuto nella società, anzi, potranno contribuire al nuovo web con l'indiscussa autorevolezza dei dati e delle informazioni di stampo bibliografico-bibliotecario. Per il lettore, poi, tutto il processo è trasparente, e non c'è bisogno che si renda interamente conto di quanto sia cambiato quello che sta dietro. L'informazione è costruita, anche al momento, dall'aggregazione corretta di tutti gli elementi di qualità che possono comporla e ricomporla a seconda degli input di ricerca. Al fruitore delle risorse, inteso in senso ampio, interessa trovare o scoprire l'informa-



zione che cercava, qualunque sia l'ambito, e trovarne giovamento e conoscenza: come quando vede un film, ignaro della costruzione registica di migliaia di elementi che appartengono ognuno alla propria realtà, concentrato su quanto appare completo e ben congegnato nella finestra della visione.

#### Maurizio Vivarelli

Credo che il termine “metadattazione” possa essere utile per descrivere alcuni tratti rilevanti della attuale realtà extradocumentaria e documentaria, sempre più fortemente caratterizzati dalla diffusione e trasformazione del Web. Rimane tuttavia aperto, mi pare, il problema di quale sia il risultato dell'attività, comunque la si denomini. Ciò che si ottiene con questa attività può essere chiamato ancora “catalogo”? Questa, come è noto, è la scelta proposta all'interno del Profilo di metadattazione DCAT-AP\_IT, con queste parole: “Il catalogo che contiene i dati (aperti e non) pubblicato da un'amministrazione”. Nella tradizione documentaria *catalogo*, *catalogazione*, *catalogatore* erano tra di loro strettamente integrati, anche sul piano etimologico e lessicale, all'interno di un quadro di riferimento concettuale sostanzialmente omogeneo, entro il quale ci imbattiamo in una parola particolarmente importante per la nostra tradizione culturale, che è quella di *λόγος*. Ora il termine “metadattazione” può essere tranquillamente correlato a “metadattore”, mentre vedo più difficile individuare il termine che designa il risultato finale; in altre parole, ciò che si colloca all'interno dell'universo bibliografico. Si tratta forse di un *meta-catalogo*? Per questo potrebbe essere utile mantenere, almeno in questa fase, questa divergenza lessicale (*metadattazione* e *catalogo*), che esprimerebbe anche la lunga durata delle connessioni storiche e culturali tra i diversi modi secondo cui possono essere intese le tradizioni disciplinari. Per quanto riguarda i vantaggi della metadattazione direi che possano consistere essenzialmente nella capacità di *navigare*, con un linguaggio e strumenti omogenei, ovvero nel web dei dati. Ma, come tutti sappiamo, molto c'è da studiare ancora

per capire e modellizzare i comportamenti informativi delle persone, documentari e non documentari, entro i diversi ecosistemi.

**Il cambiamento concettuale ha avuto conseguenze terminologiche per le biblioteche, i cui uffici non si chiamano più Dipartimento catalogazione, bensì, per esempio: alla Bibliothèque nationale de France (BnF), Département des Métadonnées e alla British Library semplicemente Metadata; dal 2015 esce "IFLA Metadata Newsletter". Come valuti la nuova terminologia che si è affermata nei modelli concettuali emanati dall'IFLA: risorse, metadati, metadattazione, agente ecc.?**

*Stefano Bargioni e Alberto Gambardella*

Un termine apposito e inusuale per la tradizione catalografica quale "metadattazione" appare particolarmente adatto a mostrare il forte e radicale impatto dell'introduzione delle entità/attributi/relazioni. La rete delle relazioni costruite, nel nuovo modello, si adattano flessibilmente a essere ri-utilizzate anche in contesti esterni alla biblioteca. Ma più che la valutazione terminologica derivante dai nuovi concetti – compito che sembra più attinente al campo della linguistica – preme valutare l'orizzonte percettivo del cambiamento apportato. Spesso le trasformazioni, comprese quelle terminologiche, vengono vissute come un ulteriore aggravio di lavoro piuttosto che come un cambio di focus e di opportunità. Occorre, almeno in ambito italiano, diffondere più capillarmente la portata della nuova visione: se accanto alla tradizionale descrizione della risorsa (operazione finalizzata sostanzialmente a una fruizione locale) si aggiunge la pratica della metadattazione, si amplia il raggio della fruibilità della risorsa stessa collocandola dinamicamente nel mondo – quello informatico dei "bits senza peso" – dell'informazione.

*Giovanni Bergamin*

Non si tratta solo di vino vecchio in botti nuove, ma le proposte di nuovi modelli concettuali (e le loro ricadute nel lessico professionale) danno conto di un lavoro di riflessione sui fondamenti in un contesto radicalmente nuovo. La trasformazione digitale in atto è qualcosa di irreversibile. Come ci ricorda Luciano Floridi, viviamo in società dove la capacità di trattare i dati con le tecnologie informatiche è non solo importante, ma essenziale per mantenere e sviluppare condizioni di vita soddisfacenti. Naturalmente i modelli concettuali vanno valutati dai risultati ma le tecnologie di oggi ci possono dare anche strumenti molto efficienti per farlo.

*Carlo Bianchini*

La nuova terminologia ha lo scopo di evidenziare il cambiamento avvenuto nell'approccio alla rappresentazione e alla valorizzazione delle collezioni delle istituzioni culturali. I cataloghi dei musei e delle biblioteche e gli inventari degli archivi erano strumenti di reperimento degli oggetti conservati nelle collezioni. Con l'avvento del web semantico ci si è resi conto che le informazioni contenute negli strumenti di corredo sono in grado di fare non solo quello, ma molto di più e sono una risorsa informativa che ha un enorme valore culturale, sociale e anche economico. Per questo motivo, il passaggio ai metadati ha spostato l'attenzione dalle sole informazioni relative ai beni culturali (per esempio il libro, il documento, l'opera d'arte) a tutte le molteplici informazioni che si possono avere su qualsiasi entità rappresentata dai metadati contenuti negli strumenti di reperimento. In questa prospettiva, sono una risorsa tutti i dati forniti sulle collezioni, non solo quelli relativi a un particolare oggetto, per esempio una particolare manifestazione o a una sua specifica copia (risorsa bibliografica). Sono una ricchezza comune, pubblica, condivisibile, i dati relativi a un'opera o a una sua espressione, a una persona, a un luogo (che sia una città, una suddivisione amministrativa, un ambiente) o a un arco di tempo (che sia un giorno, un anno o un'era) e a prescindere dal fatto che quei dati siano stati creati da una biblioteca, un museo o un archivio o qualsiasi altra agenzia autorevole. Il termine *risorsa* nel contesto del web semantico indica tutte queste entità; implica una visione più ampia e inclusiva della funzione dei dati di uno strumento di reperimento, ed è utile e utilizzabile perché trasversale, cioè consente di travalicare il confine di una specifica istituzione culturale. È molto importante questo aspetto perché l'approccio del cittadino che ricerca informazioni su una persona o un'opera, per esempio, è (sempre stato) a tutto tondo e le informazioni che riceve in risposta non devono più essere limitate a ciò che una specifica istituzione – biblioteca, archivio o museo – possiede o rende accessibile. *Risorsa* è un termine che consente di interagire correttamente con tutti gli altri attori che contribuiscono alla crescita del web semantico; per dialogare, è necessario un linguaggio comune.

*Claudio Forziati*

Per esprimere una valutazione su alcuni dei termini citati bisogna considerare il tempo di assimilazione necessario nell'uso professionale: la loro efficacia non dipende solo dall'adozione nei modelli di riferimento IFLA ma da quanto essi ci aiutano concretamente a rappresentare l'evoluzione dell'universo bibliografico. Riguardo *metadattazione* abbiamo un riscontro pratico, relativo alla varietà delle informazioni che dobbiamo essere capaci di codifi-

care per renderle comprensibili nell'ambiente in cui le esponiamo. E se anche per *risorse* ci possiamo riferire a dei principi applicativi (basta consultare il manuale UNIMARC) che ci spingono a rappresentare le informazioni non solo a partire dall'esemplificazione di un oggetto chiuso e creato per uno scopo esclusivo, più complessa è l'adozione di termini che richiedono maggiore capacità di astrazione: ad esempio l'entità di livello apicale *Res* di IFLA LRM. Questo sviluppo può essere difficile da recepire, perché non riguarda semplicemente l'applicazione di regole tecniche ma l'assimilazione di un modello concettuale. A mio avviso dobbiamo considerarne invece le opportunità perché ci spinge a riflessioni e scelte. Non è stato semplice e rapido modellare strumenti che non fossero una trasposizione dei cataloghi cartacei e che sapessero rappresentare le risorse non a partire da stringhe testuali ma da relazioni ed entità, per superare quelle risposte alle nostre ricerche che Pisanski e Zümer nel 2010 definiscono efficacemente *puzzling catalogue displays*. Ci sarà bisogno di tempo e di pensiero critico affinché termini come *agente* si affermino completamente nell'uso: richiedono la comprensione dei vari livelli di interazione che esistono tra entità materiali e concetti e quanto dal modo di rendere esplicite le relazioni e dalla possibilità di estendere le relazioni fuori dal contesto in cui la loro interazione viene inizialmente definita, dipenda l'efficacia degli strumenti di mediazione che facciamo evolvere.

#### Rossana Morriello

La nuova terminologia semplicemente si adegua al cambiamento del contesto documentario. La codifica della lingua segue dinamiche descrittive e l'approccio prescrittivo è ormai inefficace poiché non è in grado di stare al passo con i cambiamenti linguistici. Il linguaggio catalografico, che è un linguaggio, appunto, segue gli stessi principi. Nella pratica, i bibliotecari chiamano da sempre i periodici e libri elettronici "risorse elettroniche" o "risorse digitali", ancora prima del cambiamento di nome degli uffici o delle indicazioni IFLA. Questo perché l'unità documentaria è cambiata con il digitale e non è sempre possibile ricondurre una tipologia digitale a un identico corrispettivo cartaceo. Oggi possiamo accedere a un singolo articolo di una rivista in anticipo sull'uscita dell'intero fascicolo, a un preprint in un repository che è cosa differente da un articolo di una rivista, ci sono libri a stampa che presentano contenuti digitali complementari, libri online aperti e fluidi che nascono dall'aggregazione di capitoli distribuiti anche nel corso di diversi anni e aggregati solo in una fase successiva come monografia, ci sono dataset e data paper, micropubblicazioni e nanopubblicazioni. Risorse è una definizione che permette di includere tutte le diverse tipologie documentarie digitali. Metadattazione è un termine generale riferito alle risorse

a stampa e digitali, mentre catalogazione fa più specificamente riferimento alle risorse a stampa, e difatti la catalogazione è perlopiù rimasta estranea alle risorse digitali anche nella pratica. Allo stesso modo, il termine "agente" permette di coprire i vari ruoli nel mondo documentario digitale che non sono sempre riconducibili al concetto tradizionale di autore.

#### Valdo Pasqui

Molto positiva. Si tratta veramente di un'evoluzione non solo terminologica ma anche metodologica che fornisce alle biblioteche e ai catalogatori una prospettiva più ampia, trasversale rispetto ai vari domini/contesti e che apre il domani a scenari di cooperazione e di integrazione molto interessanti e innovativi, quali, per esempio, metadattazione di dati scientifici, statistici, dati estratti dai log di applicazioni e siti web che ne tracciano l'uso multidimensionale (tempo, luogo, funzioni usate ecc.). Più in generale, tutto quanto viene generato e reso disponibile in rete necessita di appropriata metadattazione per essere recuperato, referenziato, combinato con altri dati, elaborato, trasformato e conservato.

#### Tiziana Possemato

In questa estensione dei confini della biblioteca verso domini e servizi diversi, il primo ufficio che più ha dovuto ripensarsi e ridefinirsi è stato proprio quello catalografico, che si è evoluto in "ufficio per la metadattazione", per rispondere a tutte le nuove esigenze di gestione dei metadati che, appunto, sono andati ben oltre quelli catalografici. "Today, it's not about the catalogue any more, it's about the data": così Emmanuelle Bermès riassume nell'articolo *Enabling your catalogue for the Semantic Web* questa evoluzione che porta le biblioteche a essere parte, con i propri metadati, di un contesto ben più ampio di quello in cui il catalogo in sé poteva soddisfare il suo ruolo e la sua funzione. In una dimensione diversa, quella del web, in cui i dati per essere globalmente disponibili, comprensibili e riutilizzabili, devono cambiare la propria fisionomia, anche il linguaggio necessariamente cambia: certa terminologia, radicalmente riferibile al catalogo tradizionale, mostra tutti i propri limiti nel momento in cui esca dal proprio confine e si confronti con un paesaggio molto più vasto ed eterogeneo. È come se si cercasse di trovare oggi un ruolo a un elenco telefonico cartaceo, in un mondo in cui i dati riferibili a una persona sono di tutt'altra ricchezza e girano su innumerevoli e diversi canali di comunicazione. E così per la terminologia del web: è chiaro che il termine *documento* mantenga una propria valenza, soprattutto sulla scorta del significato che storicamente (seppur con non poche difficoltà nel darne una precisa definizione) ha avuto. Ma lo stesso

termine risulta tremendamente limitativo nel momento in cui il mio orizzonte cognitivo e di azione si estende oltre i tradizionali confini del mio catalogo: il termine *risorsa*, in questo caso, aiuta a identificare qualcosa che non è più strettamente limitato allo specifico dominio, ma abbraccia e riguarda ogni “cosa” disponibile nel più ampio web.

*Roberto Raieli*

Proprio perché si punta alla realizzazione di servizi e di strumenti di informazione nuovi, tecnologicamente e concettualmente in interazione con la realtà odierna, è difficile continuare a denominare con terminologie mutate dal passato, per quanto “gloriose”, nuove tipologie di risorse, di servizi e di usi a essi riferiti. I diversi “oggetti” da denominare, però, non sono ancora del tutto concettualmente chiari, quindi, se molta terminologia è giustamente in fase di ridefinizione, ben vengano tutti i nuovi termini sui quali si è finora trovato un accordo mondiale. Grazie alla palpabile crescita della ricerca intorno ai nuovi metodi e sistemi, comunque, la terminologia deve continuare a svilupparsi, per consentire la stabilità dell’ambito di applicazione, attraverso l’approfondimento concettuale dei nuovi oggetti e risorse della conoscenza.

*Maurizio Vivarelli*

Tutti i linguaggi evolvono, da quelli disciplinari e gergali al linguaggio naturale. Penso che la questione più importante consista proprio nel rimanere aperti a questi cicli di vita delle parole, inserite all’interno di metamorfosi complesse e incessanti. Penso anche possa essere utile fare riferimento ai modelli concettuali IFLA e, nello stesso tempo, tener conto di come si modificano i modelli concettuali e linguistici che le persone utilizzano. Credo insomma che il problema consista nel comprendere quali sono i caratteri e le funzioni dei linguaggi disciplinari e prendere in esame le relazioni attraverso le quali si modifica la configurazione del linguaggio naturale.

***La metadattazione ha accentuato la direzione verso una maggiore granularità, processo iniziato col formato MARC alla metà degli anni Sessanta del secolo scorso. Che rapporto intercorre tra la descrizione come identificazione delle entità e il linguaggio tecnologico?***

*Stefano Bargioni e Alberto Gambardella*

Nel mondo delle scienze del libro è sempre più rilevante l’attenzione posta alle modalità di registrazione dei dati tali che essi siano facilmente leggibili dagli strumenti

informatici, come mostra la successione degli standard e formati ISBD, MARC, BIBFRAME. La leggibilità informatica rende i dati flessibili ampliando lo spettro del loro impiego. D’altra parte, per la registrazione dei dati, la scienza informatica deve approntare ambienti amichevoli, coerenti e duttili, capaci al contempo di nascondere complessità e tecniche specifiche e peculiari del proprio ambito. Un’identificazione delle entità che si sviluppa verso una maggiore granularità offre alle potenzialità dell’informatica una vasta disponibilità di dati che vanno poi debitamente combinati e incrociati per offrire una gamma di informazioni capillari e dedicate (trovare/identificare/selezionare/ottenere).

*Giovanni Bergamin*

La granularità associata al formato MARC ci porta al concetto di riusabilità. Oggi è possibile riusare in maniera più efficiente che in passato i metadati che vengono prodotti e la riusabilità può non essere limitata al dominio specifico all’interno del quale i metadati sono stati creati. Le tecnologie che si stanno affermando in questo campo si basano sulla visione del web semantico e sulla tecnologia dei linked data. Si tratta – secondo la Library of Congress – di intraprendere un percorso di transizione con l’obiettivo di ottenere i vantaggi della tecnologia più recente senza rinunciare a tutto quello che il MARC è già in grado di garantire.

*Pierluigi Feliciati*

L’evoluzione dell’universo informativo verso la granularità è conseguenza diretta del superamento dell’idea di opera chiusa, di “scheda”, di rappresentazione definitiva. Tale forma di registrazione bibliografica era funzionale soprattutto alla rappresentazione di un oggetto fisico, dell’esemplare di una particolare edizione come di un fondo o di un documento d’archivio, oppure di un qualunque altro oggetto o idea “testimonianza di civiltà”. Ogni “cosa” necessitava di una narrazione originale, specifica, specializzata pur se standardizzata, che si risolveva in sé stessa e poteva al massimo essere aggregata ad altre narrazioni simili secondo logiche ordinali funzionali al loro ritrovamento, affinché fosse meno fortuito possibile. Siamo ancora assestandoci, rispondendo alla profonda spinta innovativa dell’universo informativo basato sull’ipertesto, sulle connessioni, il successo della esposizione dei dati in modo “aperto”, sia dal punto di vista tecnico-formale – adottando un metalinguaggio testuale leggero e universale come XML – sia per la necessità di trasparenza da parte degli enti pubblici inaugurata dalla presidenza statunitense Obama con il rilascio della piattaforma data.gov. Queste dinamiche hanno messo in evidenza come siano le relazioni lo strumento per rendere pienamente e con

efficacia semantica la complessità dell'agire umano, specie quando questo prende la forma di oggetti di conoscenza. Dunque, le unità informative si sono sempre più granularizzate, dalla pagina web all'oggetto digitale in una pagina web, al post, al commento, alla scheda d'autorità, infine all'entità semanticamente interpretabile dalle macchine in base alla formalizzazione delle sue proprietà di connessione con le altre entità, ancora con XML e i suoi derivati.

#### Rossana Morriello

La descrizione e il linguaggio tecnologico sono strettamente collegati. Se ci pensiamo bene, anche la catalogazione tradizionale usa un linguaggio tecnologico, solo che è riferito alla tecnologia della stampa, oppure alle tecnologie dei materiali non librari (*non book materials*). Quando, nella catalogazione tradizionale, nell'area della descrizione fisica usiamo l'indicazione che ci permette di identificare un CD-ROM, stiamo usando un linguaggio tecnologico. Il formato MARC nasce negli anni Sessanta perché è in quegli anni che si avvia l'automazione nelle biblioteche e, come dice il nome, sorge la necessità di avere un formato di catalogazione leggibile dalle macchine. L'universo documentario ha subito varie trasformazioni e i formati catalografici si sono di volta in volta adeguati alle trasformazioni per potervi far fronte efficacemente. Come fanno tutti i linguaggi, le lingue vive. È difficile identificare e descrivere un'entità se non possiedi la terminologia adatta per farlo. Oggi siamo nella società dei dati e dunque parliamo di metadattazione.

#### Valdo Pasqui

Il MARC, nelle sue varianti, ha avuto il merito di consentire una descrizione strutturata, dettagliata e interpretabile dai computer (dalla "macchina"), ma quando le risorse di calcolo, di memoria e di archiviazione (storage) sono diventate ampiamente disponibili a basso costo quello che era uno dei vantaggi iniziali, ovvero sfruttare ogni singolo bit per codificare informazioni descrittive utili, è diventato anche un grande limite. La sua strutturazione a record mischia sintassi rappresentativa e semantica (tag, intestazioni, campi di controllo ecc.) A partire dalla formalizzazione del linguaggio XML si è sempre più delineata una chiara separazione tra sintassi e semantica, uno o più set di metadati, concepiti anche in ambiti diversi, possono coesistere per descrivere un oggetto/opera/risorsa grazie alla sintassi XML che consente sia l'interpretazione manuale da parte di un essere umano sia quella automatica da parte delle applicazioni. La sostituzione dei valori degli attributi e delle entità con i loro identificativi che fanno riferimento (tramite URI) a set standardizzati di meta-

dati e di valori appartenenti a domini standardizzati e controllati migliora l'interpretazione automatica tramite applicazione e consente di costruire e scambiare metadattazioni più corrette.

#### Tiziana Possemato

La trasformazione delle descrizioni bibliografiche in elementi sempre più granulari ha l'obiettivo di rendere meglio identificabili e riusabili i dati. Il MARC, con la sua lunghissima storia di oltre mezzo secolo, è arrivato a un livello di maturazione al momento difficilmente eguagliabile: le attuali conversioni da MARC a qualsiasi altro formato (incluso l'RDF specificato in una delle tante ontologie di dominio) ancora non ottengono lo stesso livello di granularità e specificità del MARC, con i suoi tag e i suoi sottocampi. Bisognerà che le varie ontologie crescano, vengano applicate e maturino perché questo davvero avvenga. Il linguaggio tecnologico ha supportato questa evoluzione fino a divenire parte stessa della disciplina catalografica: difficile, oggi, parlare di competenza catalografica o di specialisti di metadati senza un minimo di conoscenza dei linguaggi tecnologici. La domanda introduce un ulteriore elemento, quello della *identificazione delle entità*, che ci conduce oltre lo stesso concetto di metadattazione verso quello dell'*entity modeling*: in un web fatto di cose reali del mondo (*real world objects*) descrivere risorse significa sempre più identificare e modellare entità, a prescindere dallo specifico contesto di utilizzo: descrivere Dante Alighieri significa selezionare e collegare insieme quella serie di attributi che definiscono la persona Dante Alighieri, e le relazioni che aiutano a meglio identificarlo in un tempo, in uno spazio, in un contesto operativo. In questo ulteriore passaggio di livello teorico e pratico, il ruolo del linguaggio tecnologico diventa meno ancillare rispetto alla tradizione: l'*entity modeling* è il risultato di un contributo quasi paritetico tra competenza catalografica e linguaggio tecnologico, in una sorta di rapporto di mutua assistenza.

#### Roberto Raieli

La granularità dell'informazione è il presupposto principale della rivoluzione che porta, per esempio, dalla scheda bibliografica monolitica al set di dati non necessariamente di natura bibliografica, espandibile, intercollegabile, riutilizzabile "anche" in ambito bibliografico. Il dato a sé stante, inteso nel proprio valore individuale e assoluto, atomo di informazione identificato da un URI, è a fondamento del sistema di stringhe XML-RDF tra di loro intercollegabili. Tale informazione "isolata", però, può appartenere a tutti i contesti possibili, il che equivale a non appartenere ad alcun contesto, e questa condizione può portare al rischio di dispersione del contesto originario di appartenenza di una risorsa.

sa che è rappresentata volta per volta dai singoli dati. L'esigenza della contestualizzazione è fortemente sentita nella comunità archivistica, ma non meno in quella bibliografico-bibliotecaria, e sicuramente può generare perplessità in vari altri ambiti, ogni volta che i diversi dati si allontanano troppo dall'aggregazione che in un certo settore può meglio rappresentare una risorsa organica, per essere variamente combinati, e interpretati, in altri generi di set. Del resto, l'architettura del *semantic web* implica comunque una scelta organizzativa dei dati grezzi, e quindi una specifica scelta informativa, per quanto neutrale o "agnostica" da un punto di vista conoscitivo. La granularità è, dunque, un modo di vedere il mondo dell'informazione, un paradigma per la conoscenza, e condiziona le visioni successive dei dati informativi tramite il sistema tecnologico dei linked data che da questa base trae origine, nonostante si sia generalmente propensi a considerare i LOD, in ogni caso, una delle soluzioni più neutre e generali tra quelle disponibili. È necessaria, quindi, un'effettiva riconcettualizzazione del modello di descrizione bibliografica, intendendola nella prospettiva della granularità e interoperabilità, che porti a una trasformazione dell'attività di catalogazione, aprendo realmente lo sviluppo del catalogo all'attuale progresso dell'informazione e della conoscenza, per usi e riusi di tipo non solo bibliografico e bibliotecario.

#### *Maurizio Vivarelli*

La mia impressione è che in primo luogo si debba comprendere la complessità delle procedure di descrizione delle diverse entità, e della loro rappresentazione linguistica. Credo che, contestualmente, e almeno a livello di ricerca, potrebbe essere utile riflettere approfonditamente sui modelli che proprio i "linguaggi tecnologici" utilizzano, sulla struttura concettuale – il database – che ne ha contraddistinto la fisionomia fin dalle origini. Come ha spiegato efficacemente Lev Manovich la logica del database mal si armonizza con la fisionomia delle "narrazioni", un tema di forte attualità in relazione al riuso dei dati. E sono convinto che questo costituisca un problema serio su cui riflettere, quando ci si pone la questione della comunicazione dei contenuti verso le persone. La tecnologia, insomma, non è un qualcosa d'altro rispetto al linguaggio, ma è piuttosto una tra le molte "tecnologie della parola", come ha scritto Walter J. Ong, che concorrono alla evoluzione, linguistica e antropologica, dei modelli di rappresentazione della realtà, dentro e fuori l'universo bibliografico.

***Nel contesto del web collaborativo, potenzialmente chiunque può creare, gestire e aggiornare dati, dando***

***il proprio contributo (per esempio, in agenzie bibliografiche, piccole biblioteche, Wikidata). Come valutare questa opportunità? Com'è possibile garantire la qualità dei metadati?***

#### *Stefano Bargioni e Alberto Gambardella*

La metadattazione richiede lo sviluppo di una rete di collaborazioni – a volte implicite, altre esplicite – tra agenzie: sotto l'aspetto tecnologico si manifesta nel mantenimento dei permalink e degli URI Registry; sotto quello umano, nella cooperazione alla definizione di ontologie, al mantenimento di thesauri e liste controllate, fino al confronto diretto di opinioni, al fine di assicurare registrazioni identificanti e uniformi. La qualità dei dati va assicurata curando sia la specificità (granularità) che l'uniformità. L'utilizzo di GeoNames o di Wikidata per la registrazione dei luoghi nei record di autorità (campo 370 del formato MARC21 per record di autorità) permette, per esempio, di utilizzare Parigi anziché Paris mantenendo intatta l'interscambiabilità del dato. L'utilizzo dei menù a tendina forniti di *autosuggest* per la registrazione di dati relativi ad attività, gruppo di appartenenza, occupazione, genere, lingua (campi 372, 373, 374, 375, 377 del formato MARC21) oppure l'utilizzo di standard ISO per registrare data di nascita e di morte (campo 046 del formato MARC21), oppure l'impiego di identificatori, VIAF, ISNI, Wikidata (campo 024 del MARC21) sono alcune delle strategie di registrazione che vanno capillarmente condivise nel mondo delle biblioteche. In Italia manca, per esempio, una lista controllata per le occupazioni (quella presente nel Nuovo soggettario Thesaurus è al momento piuttosto carente). Wikidata possiede una lista per le occupazioni carente per uniformità ma flessibile, poiché permette facilmente di aggiungere alla lista occupazioni mancanti e di correggere eventuali errori.

#### *Giovanni Bergamin*

Fermo restando la possibilità per chi pubblica i metadati di decidere chi li può aggiornare, il web ha indubbiamente potenziato le possibilità di *crowdsourcing* o di partecipazione nella creazione dell'informazione. Il successo di queste iniziative sta nel "meccanismo che separa il segnale dal rumore" o "filtro" (Andrea Zanni). Wikidata (per restare nel campo dei metadati) funziona perché la comunità che la gestisce ha concordato procedure e strategie che permettono di filtrare appunto il segnale dal rumore. Anche Google aveva pensato alla necessità di avvalersi di una base dati di informazioni strutturate e nel 2010 aveva acquisito – dalla società Metaweb – Freebase: una raccolta di dati strutturati raccolti da molteplici fonti e alimentata anche con modalità di *crowdsourcing*. La notizia dell'avvio di Wikidata a opera

di Wikimedia Deutschland nel 2013 ha orientato Google a offrire un contributo di 1,3 milioni di euro all'iniziativa, a chiudere Freebase (nel 2014) e a finanziare la migrazione di Freebase in Wikidata (nel 2015). In altre parole, Google riteneva che il sistema di "separazione del segnale dal rumore" che Wikimedia Foundation aveva creato con Wikipedia poteva essere un sistema di successo anche per il mondo dei dati strutturati con Wikidata.

#### *Carlo Bianchini*

Si tratta di un'opportunità molto importante perché va in tre direzioni: la prima è quella della possibilità di arricchire i dati relativi alle proprie collezioni con dati provenienti da insiemi di dati prodotti da altre agenzie. È proprio l'istituzione che decide di farlo a scegliere l'insieme di dati da utilizzare per arricchire i propri strumenti di ricerca. Per esempio, la sezione Protagonisti di Alfabeta ha scelto di arricchire le proprie voci di autorità prive di dati di nascita e di morte con dati provenienti da Wikidata. La seconda direzione è quella in uscita. Chi espone i propri dati come LOD si può ragionevolmente aspettare che questi siano riutilizzati da altre istituzioni, come Wikidata fa con moltissime applicazioni che ne riusano i dati. L'esposizione dei dati per il loro riuso aumenta in modo massiccio la visibilità dei dati e dell'istituzione che li produce e, di conseguenza, la valorizzazione delle collezioni dell'istituzione stessa, come dimostrano le esperienze del Metropolitan Museum e della BEIC. La terza direzione è proprio quella della qualità dei dati. Quando si espongono i propri dati in forma di LOD, bisogna necessariamente confrontarli con quelli di altri insiemi di dati che hanno dati simili. Questo confronto (riconciliazione) è un potentissimo strumento di controllo della qualità dei dati. Per esempio, la riconciliazione degli ID dei luoghi di SBN con Wikidata svolta nel mese di marzo 2022 ha consentito, nel giro di pochi giorni, di stabilire che ben il 10% dei luoghi erano da correggere. Molti errori di identificazione del VIAF – come l'assegnazione di più identificatori "unici" alla stessa entità oppure l'assegnazione di entità diverse al medesimo identificatore – sono individuabili proprio attraverso la riconciliazione dei dati del VIAF su Wikidata. Infatti, nel rapporto tra VIAF e singola agenzia bibliografica nazionale, questi errori non emergono; si possono individuare solo se il confronto si allarga oltre i confini del proprio orticello, esattamente come avviene nell'ambito del dibattito scientifico, come dimostriamo Stefano Bargioni, Camillo Carlo Pellizzari di San Girolamo e io in *Beyond VIAF. Wikidata as a complementary tool for authority control in libraries*, "Information Technology and Libraries", 2021.

#### *Pierluigi Feliciati*

La qualità di un servizio, di un prodotto, di un'informazione non è valutabile adottando metriche assolute, ma considerando che essa dipende soltanto dalla risposta alle esigenze di determinati utenti, che agiscono in determinati contesti con efficacia, efficienza e soddisfazione. Dunque, il confronto con i fruitori, gli utenti, i lettori costituisce un passaggio imprescindibile per verificare l'effettiva qualità del catalogo. Il confronto si espande, nel panorama fortemente interattivo e partecipativo creato dal web sociale, dalla connessione a Internet pressoché universale, dalla disponibilità di dispositivi mobili, dalla preponderante produzione e condivisione di dati e informazioni da parte di chiunque facilitata dalla presenza di sensori. Diventa annotazione, collaborazione in grado di attivare un processo culturale partecipativo che insegna a interpretare l'informazione anche mentre la si produce. Sono molti i vantaggi educativi e sociali di questo fenomeno, uno degli argini rimasti per difenderci dall'assalto della visione della cultura come *shopping mall* digitale globale, come profetizzava Pierre Levy ai primi degli anni Novanta del secolo scorso. Il successo popolare di Wikipedia – inaspettato, per chi ne ha osservato pur con favore la nascita e lo sviluppo – che ha superato i 21 anni di vita e si è "naturalmente" assestata, senza il supporto di ingenti capitali, come la "più consultata opera di riferimento generalista su Internet", con 55 milioni di voci in più di 300 lingue e una media di 400 milioni di visualizzazioni al mese, costituisce un'esperienza da cui è impossibile prescindere. La registrazione delle attività collaborative come parte integrante dell'oggetto informativo – tramite il software MediaWiki – è essa stessa una novità importante, garanzia di trasparenza e diario dell'intelligenza collettiva. Una questione aperta è il ruolo degli esperti nel processo di creazione e aggiornamento dell'*infosfera*: solo attivatori della collaborazione, revisori continui dei dati, autori "principali"? La scelta radicale di Jimmy Wales, che rifiutò del tutto l'idea di una redazione esperta per Wikipedia sostenuta dal suo socio Larry Sanger, fondando invece l'accuratezza sulle dinamiche di controllo delle comunità, non può non farci riflettere. Nel settore delle *digital humanities*, d'altra parte, le metodiche e tecnologie per favorire l'interazione attiva e l'aggiunta di annotazioni ai contenuti si sono fatte sempre più avanzate.

#### *Claudio Forziati*

Una platea ampia di soggetti che interagiscono in un ambiente aperto è certamente problematica e pone criticità in termini di coerenza organizzativa dei dati, a seconda delle scelte tecnologiche, dei modelli di rappresentazione e della terminologia utilizzata. Ma, in un

sistema come i linked data, valutare la coerenza e la qualità dei dati non è più soltanto una prerogativa umana poiché entrano in gioco processi che considerano la presenza di meta-informazioni specifiche (per esempio la *provenance*). Se consideriamo come un vantaggio la possibilità di dereferenziare gli URI, non possiamo pensare di interrogare e mettere in relazione i dati che produciamo riproponendo un sistema a *enclave*, in cui le istituzioni della conoscenza escludono la capacità creativa e propositiva delle comunità informali che popolano il web: perderemmo il vantaggio pratico di evitare la duplicazione di informazioni già esistenti e correttamente strutturate. Osservando con attenzione Wikidata possiamo capire facilmente perché è un progetto ormai apprezzato diffusamente in ambito bibliotecario. Quattro caratteristiche sono a mio avviso significative: la possibilità di creare entità in maniera gratuita e intuitiva, in una *knowledge base* la cui ontologia può essere modificata attraverso la conversazione degli utilizzatori; la funzione di collettore di identificatori persistenti provenienti da migliaia di fonti, spesso istituzionali; la presenza di informazioni di contesto e fonti verificabili potenzialmente per ogni dichiarazione; il rilascio dei dati in CC0, che ne consente un riuso ampio e indeterminabile *a priori*.

#### Maurizio Lana

Che chiunque possa creare, gestire e aggiornare dati, dando il proprio contributo è la caratteristica del web nella sua forma base – introdurre filtri e paletti è ugualmente possibile ma cambia lo stato delle cose. Che tutti possano produrre dati, è potenzialmente vantaggioso per tutti perché se la metadattazione è tale perché tutti gli oggetti digitali possono essere metadattati, metadattare tutti gli oggetti digitali è un'impresa di enorme ampiezza e di enorme difficoltà che deve coinvolgere moltissimi soggetti differenti. Il problema è la qualità della metadattazione che questi soggetti indeterminati possono produrre. Le riflessioni sono due: la prima è che i metadati dovrebbero essere accompagnati da informazioni su chi (individuo+ente di appartenenza se c'è) ha prodotto quel tal metadato. Se in un contesto di risorse informative bibliografiche vedo che ci sono metadati prodotti da Mauro Guerrini e altri prodotti da Amedeo Montanese [inteso come nome generico di una persona che non è nota come un esperto di risorse informative bibliografiche] posso decidere di usare quelli prodotti da Guerrini e non quelli prodotti da Montanese; o viceversa; con pieno controllo sulle mie scelte e sul loro probabile esito. "Amedeo Montanese" che metadatta dei contenuti bibliografici potrebbe essere visto come manifestazione di quella sovrabbondanza informativa che viene usualmente deprecata. Ma se risultasse che Amedeo Montanese è un cognitivista che



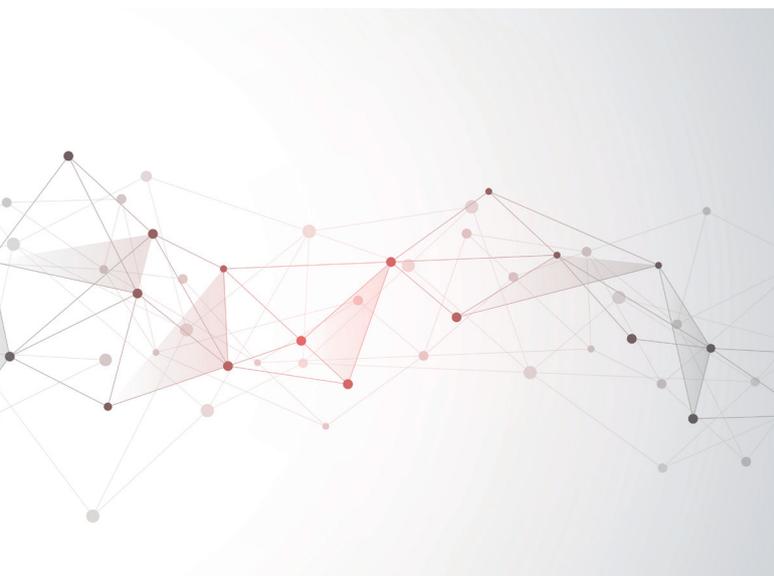
riflette sui meccanismi che gli esperti mettono in atto quando metadattano un contenuto informativo, allora la sovrabbondante (inutile, indesiderabile) informazione costituita dai metadati prodotti da Montanese si rivelerebbe in realtà informazione utile, necessaria. La seconda riflessione è che la metadattazione può apparire il campo in cui meglio possono esplicarsi procedure automatiche di individuazione di entità (NER Named Entity Recognition che si basa su NLP Natural Language Processing) con susseguente creazione e connessione di metadati. Ma non tutte le entità presenti in una risorsa informativa sono ugualmente rilevanti per il suo significato e, dunque, un soggetto umano che operi nella gestione e organizzazione dei metadati può definire quali debbano essere effettivamente attivati, con ciò aumentando il valore informativo.

#### Rossana Morriello

Si tratta di una realtà ormai consolidata, il web è collaborativo da molto tempo e non si può pensare di tornare indietro né far finta che non sia così. Le biblioteche hanno due opzioni: lasciare procedere liberamente il flusso di collaborazione degli utenti e, in questo modo, consegnare ad altri operatori il compito di organizzazione dei dati e della conoscenza, restandone fuori, oppure cercare di governare questi processi e indirizzarli verso principi e pratiche della cultura biblioteconomica. La qualità dei metadati può essere garantita dalla capacità di produrli professionalmente secondo standard aperti e interoperabili, proprio come si è sempre fatto con la catalogazione.

#### Valdo Pasqui

In ambito strettamente bibliotecario la domanda che



spesso mi sono posto e che crea attrito tra gli informatici e i bibliotecari catalogatori è perché sia necessario catalogare ricominciando da zero ogni volta. In realtà questo non è vero, perché come dimostrano l'Indice SBN o altri contesti di catalogazione condivisa e federata, spesso la catalogazione avviene in modo derivato recuperando (catturando nel linguaggio SBN) un record, cioè una descrizione, preesistente. Dunque, a ben vedere, la catalogazione tradizionale oramai potrebbe essere sostituita dal recupero automatico dei dati bibliografici alla fonte (editore) mentre il valore aggiunto dovrebbe consistere in fasi successive quali: soggettazione semantica, arricchimento con parole chiave e taggature, appropriata classificazione. Da ultimo, ma non ultimo, promuovere, tramite cataloghi web, la collaborazione degli utenti delle biblioteche secondo approcci di tipo social come l'aggiunta di recensioni e suggerimenti per altre letture pertinenti, la possibilità di taggare autonomamente, l'aggiunta di riferimenti del tipo "chi ha letto questo testo ha letto anche".

#### Tiziana Possemato

Ogni fonte informativa, sia essa un catalogo, un dataset, un'ontologia, un sito web o altro, è identificabile con un'istituzione o una comunità responsabile della sua creazione e gestione. La forza di una comunità, che garantisce l'autorevolezza e la certificabilità di una fonte, è data anche dall'estensione (aspetto quantitativo) e dalla tipologia (aspetto qualitativo) della comunità stessa, garante della fonte. Questi precetti dovrebbero in parte arginare i rischi inerenti al Principio AAA, che è la base fondante del Semantic Web: *Anyone can say Anything about Anything* (chiunque può dire qualsiasi cosa su qualsiasi cosa). Ma se può sembrare piuttosto semplice

inquadrare, verificare e certificare i dati quantitativi di una comunità che sostiene e produce una fonte, attraverso criteri obiettivi di misurazione, la valutazione dei dati qualitativi non è così semplice. E questo è tanto più vero se pensiamo a una dimensione globale, come quella del web, in cui una comunità può essere diffusa oltre ogni possibile confine misurabile. È qui che il concetto di *autorevolezza* rischia di dover cedere il passo al concetto di *consenso*, ed è qui che, forse, ancora di più, occorre ripensare e rafforzare il concetto di *autorevolezza certificata* di una fonte. I dati costituiscono lo scheletro su cui si costruisce la struttura della comunicazione: più i dati sono autentici, veritieri, autorevoli, certificati e verificabili, più difficile è alterare la realtà che descrivono. Il dato di *riutilizzabilità* è uno degli elementi e dei criteri di controllo della sua qualità: un largo riutilizzo del dato genera un diffuso controllo dello stesso e non è un caso che, nella mappa dei dataset in LOD pubblicata sul sito [lod-cloud.net](http://lod-cloud.net), le fonti più autorevoli siano anche quelle più ampiamente collegate (e, dunque, riutilizzate). Aprirsi significa esporsi al giudizio. E anche alle critiche, se utile o necessario.

#### Roberto Raieli

La nascita del web è ricca di istanze democratiche, sviluppate con coerenza in tre diversi periodi, ma non senza criticità. Il suo fondatore, Tim Berners-Lee, ha pensato a uno strumento per la condivisione di informazione accessibile a tutti, libero, dotato di tecnologie standard, per la crescita della libertà, del benessere, della collaborazione. In base a tali presupposti, il progetto è sviluppato dall'intera comunità. Il semantic web, nato dalle idee dello stesso Berners-Lee, arricchisce e magnifica le istanze democratiche originarie del web. Il meccanismo dei linked data, che lo struttura e lo sostiene, è *free*, nasce e si sviluppa collaborativamente, ha nell'essenza la granularità, l'integrazione, la condivisione. Il programma dei dati aperti, collegati e condivisi convince molte persone e molte organizzazioni, e pare essere quello che meglio può realizzare ogni progetto di diffusione delle conoscenze e di sviluppo del *welfare*. Lo scopo di queste attività non è nuovo per gli istituti e gli operatori culturali, ma si inserisce adesso in un panorama parzialmente nuovo, e in crescita, dove le istanze sociali e democratiche continuano a far breccia a tutti i livelli – politici, sociali, economici –, diventando teoria e pratica per lo sviluppo dei diversi paesi che aspirano veramente alla condivisione della conoscenza, al progresso, al benessere sociale. Proprio perché coscienti, insieme ai vantaggi, dei grandi rischi del produrre e pubblicare conoscenza in modo "diffuso" e aperto a tutti, Berners-Lee dota il web semantico di alcuni criteri di "autoprotezione" e controllo, che sono ancora in

fase di studio e sperimentazione. Già nel 2000 l'architettura del semantic web contempla come livelli costitutivi tali criteri di sviluppo, illustrati nell'ambito della rappresentazione grafica generale poi definita *semantic web stack*. Questa architettura pone tra i livelli più bassi quelli delle tecnologie fondamentali su cui si costruisce il web, prima HTTP, URI, poi XML, più in alto RDF, al centro OWL, SPARQL. I tre livelli più alti sono quelli di *unifying logic*, di *proof* e di *trust*, gli spazi della "logica unificante", della "dimostrazione" e della "credibilità", che completerebbero l'architettura del web semantico con strutture essenziali per dare affidabilità, certezza, esattezza ai dati e alla conoscenza da esso diffusi.

*Maurizio Vivarelli*

Il problema anche in questo caso è complesso. In questa fase, dal punto di vista culturale, storico e documentario è indispensabile tener conto della dimensione collaborativa e partecipativa, che non a caso costituisce il punto di approdo della storia sociale della conoscenza tracciata da Peter Burke nel suo volume *Dall'Encyclopédie a Wikipedia*. Tuttavia, è necessario anche essere consapevoli del fatto che non si può concretamente "metadattare" tutto, dando corpo nella sostanza alla riemersione di un sogno universalistico e totalizzante coesteso alla nostra tradizione culturale fin dalla sua configurazione originaria. Non si può "metadattare" tutto così come non si è potuto "catalogare" o "documentare" tutto, come ha cercato di fare Paul Otlet inseguendo il sogno del suo Mundaneum. L'oggetto della attività, dunque, un po' come il concetto di patrimonio culturale, è destinato a rimanere indefinitamente incerto, a non poter essere mai del tutto raggiunto, come nel paradosso di Achille e della tartaruga.

***È affascinante la prospettiva di un record che si crea sulla base di dati che provengono da vari contesti e che solo nel contesto bibliotecario diventano record bibliografici, mentre negli altri possono diventare di tutto? Ciò, secondo voi, arricchisce o snatura la tradizione catalografica?***

*Stefano Bargioni e Alberto Gambardella*

La valorizzazione del dato rispetto all'intero record bibliografico e la valenza assunta dalla sua modularità (flessibilità), implicita nei nuovi modelli di metadattazione, rappresenta un tema che spinge verso una condivisione più responsabile. Una condivisione allargata richiede una precisione maggiore, un approfondimento del dato che si registra. Il livello di profondità di strutturazione del dato, in termini di entità/attributi/relazioni, va modulato sia rispetto all'identità della biblioteca, sia

rispetto all'orizzonte degli utenti che la frequentano; va da sé che a una biblioteca di filosofia non spetta il ruolo di approfondire la registrazione di dati relativi al campo medico, ma certamente ha il dovere di accrescere la qualità dei dati relativi al mondo della filosofia. Il lavoro va sviluppato in senso verticale (approfondimento, precisione, granularità) ma sempre in relazione al proprio contesto lavorativo. Si crea così un circolo virtuoso di scambio: si possono offrire dati di qualità al web semantico e complementariamente si può usufruire delle competenze altrui. In questa ottica la metadattazione appare sia sostenere che rinnovare la tradizione catalografica.

*Giovanni Bergamin*

Quello che conta è che non si possono prendere scorciatoie. Mettere insieme differenti contesti non è facile. Recentemente Denny Vrandečić in un suo tweet (<https://tinyurl.com/hvbkpqt>) ha segnalato un curioso "assemblaggio di metadati" che Amazon aveva messo insieme per descrivere un libro con data di pubblicazione 22 marzo 2021 offerto in vendita su Amazon.com. I metadati mettevano in realtà insieme tre "libri" differenti: l'immagine della copertina era tratta da una edizione dei *Principia Mathematica* di Whitehead e Russell (la prima edizione è del 1910); il titolo della descrizione bibliografica era *The Principles of Mathematics* (che è invece un'opera di Russell del 1903); l'abstract era infine relativo all'opera del 1687 di Newton *Philosophiæ Naturalis Principia Mathematica*. In un tweet collegato Vrandečić invitava Amazon a considerare il precedente tweet come una "segnalazione amichevole di un problema fatta da qualcuno che sa quanto sia difficile trattare con i metadati" (Vrandečić è tra le altre cose uno dei fondatori di Wikidata). La pagina non è più disponibile su Amazon, ma è disponibile la copia su Internet Archive (<https://tinyurl.com/y7sjffyx>). In altre parole – per rispondere alla domanda – la tradizione catalografica non è assolutamente in discussione. Sono in discussione le applicazioni che semplificano la complessità.

*Calro Bianchini*

Ciò che fino a oggi abbiamo chiamato registrazione (record) è il risultato di un processo in cui i dati rilevanti – in base alle regole catalografiche adottate – per indicare qualcosa venivano appunto registrati in una forma definita e imm modificabile su un supporto; per esempio, una scheda cartacea o un record bibliografico in un certo formato (della famiglia del MARC). Con la metadattazione, continuiamo sempre a registrare tutti i dati rilevanti – preferibilmente secondo regole catalografiche adatte ai metadati – ma in un ambiente aperto. Quindi non cambia e non si snatura la tradizione catalografica, per-

ché continuiamo a produrre tutti i dati (attributi e relazioni) indispensabili alla realizzazione di un catalogo. Ciò che cambia è la presentazione, cioè il modo in cui restituiamo al lettore i dati che abbiamo creato. Ciò che il lettore vede è *una* delle possibili visualizzazioni dei dati che sono stati registrati; in quanto registrati come dati e non come 'registrazione' i dati possono essere presentati e visualizzati in molti modi diversi, e non tramite l'unica registrazione effettuata. Non si tratta soltanto di una questione estetica o grafica. In primo luogo, la presentazione può essere fatta in modo personalizzato per un particolare lettore. Per avere un esempio, si può guardare al Knowledge Panel di Google (il box che viene presentato a destra dei risultati di certe ricerche); se si fa una ricerca, il Knowledge panel presenta sia i dati che i metadati nella lingua del lettore (per una controprova, cambiare le impostazioni della lingua nel browser). In secondo luogo, la quantità, l'ordine e la provenienza dei dati che possiamo fornire nella presentazione dipendono dalle nostre scelte. Nulla vieta che si presentino i soliti dati nella forma tradizionale; o che si offrano dati in forma ridotta (come avveniva negli *short title catalogues*) o ancora dati in forma molto approfondita (come avviene nei repertori bibliografici). La differenza tra record e metadati e i vantaggi della metadatazione sono chiari: i metadati non offrono un prodotto solo predefinito nell'ordine e nello schema di presentazione delle informazioni, ma di dati che sono in grado di svolgere molte funzioni diverse, creando più presentazioni in relazione alle necessità del lettore e agli obiettivi dell'istituzione. Il progetto LOD del CoBiS e l'authority box del catalogo PUSC (come scrive Stefano Bargioni, in *From Authority Enrichment to AuthorityBox. Applying RDA in a Koha Environment*, "JLIS.It", 2020) sono inoltre un esempio di come, laddove i propri metadati siano pochi, possono essere arricchiti di metadati provenienti da altri insiemi di dati, che vengono scelti sempre dall'istituzione che progetta la presentazione dei dati ai propri lettori.

#### *Pierluigi Feliciati*

Le modalità con cui le persone, da sole o in forma organizzata, nella storia, producono segni materiali e immateriali delle proprie attività creative non corrispondono se non raramente, *ab origine*, a forme e a logiche basate sugli specialismi. Sono le attività a dettare le forme, non il contrario. Gli stessi *agents*, negli stessi luoghi e negli stessi momenti, hanno redatto e ricevuto documenti, letto o scritto opere letterarie o scientifiche, prodotto o acquistato oggetti artistici, di cui possono essere stati anche il soggetto, costruito e abitato mura, case e palazzi, vissuto e fatto evolvere le proprie tradizioni. Rappresentare le talvolta complesse dinamiche di tale sedimentazione e attributi speci-

fici dei sedimenti richiede competenze specialistiche, perché i contesti e i contenuti possano essere correttamente compresi, ricostruiti e formalizzati. Inevitabili allora i lessici disciplinari, i modelli concettuali diversificati, gli standard descrittivi di settore, le forme di composizione delle registrazioni. Ma tale diversificazione si è estremizzata, a mio parere. Abbiamo ecceduto negli specialismi fino al paradosso, trattando le stesse cose in modo diverso – dal punto di vista informativo – a seconda del contesto (talvolta fortuito) di descrizione o di conservazione, e sacrificando la potenza delle interferenze, delle multidimensionalità informative, delle trasversalità. Credo che un rovesciamento di tale tendenza, che moltiplica i punti di accesso – raramente completi – a discapito della chiarezza e della soddisfazione degli utenti, non debba snaturare le tradizioni disciplinari, ma possa trovare nel web semantico l'ambiente ideale per la ricomposizione delle distinzioni, per la scoperta delle coincidenze, per l'attivazione delle connessioni. Perché una persona può essere autrice di un libro, soggetto di un altro libro, soggetto produttore di un archivio, corrispondente in un carteggio, soggetto di un ritratto a olio, avere una casa museo dedicata, e tutto questo muovendosi in diversi luoghi e contesti nel tempo. Se consideriamo che tutte queste informazioni e molte altre siano già disponibili, spezzettate però tra inventari archivistici, cataloghi bibliografici, cataloghi di museo o di mostra, indici delle edizioni dei carteggi ecc., non è un obiettivo possibile contribuire a connetterle, in forma di dati, per consentire alle macchine e agli utenti di navigarne il grafo?

#### *Maurizio Lana*

Questo ("un record che si crea sulla base di dati che provengono da vari contesti e che solo nel contesto bibliotecario diventano record bibliografico") è inevitabile perché le risorse informative sono nel mondo digitale e nel mondo digitale tutto è un dato, tutto esiste in quanto è un dato. Il fatto che questo accada dice che le risorse bibliografiche sono pienamente parte del mondo dell'informazione digitale. Questo snatura la tradizione catalografica solo nella misura in cui essa sia un mondo specialistico chiuso (al limite: autoreferenziale); se invece vediamo la tradizione catalografica come espressione della gestione della descrizione delle risorse informative nel mondo fisico, quando le risorse informative si manifestano anche nel mondo digitale allora è ovvio che la loro descrizione assuma le specificità del contesto in cui esse prendono vita ed esistono; e che quindi (al limite) la metadatazione sia analoga nel mondo digitale a ciò che la catalogazione è stata ed è tutt'ora nel mondo fisico.

Rossana Morriello

Questa è la base della scienza dei dati ed è un aspetto fondamentale della loro gestione. I dati che provengono da vari contesti sono resi omogenei attraverso processi di modellazione, trasformazione, arricchimento e integrazione, a seconda dell'obiettivo che ci si pone e del contesto nel quale si opera. L'ambito bibliotecario non fa eccezione, trattandosi di dati. Nel processo di gestione dei dati si opera tramite le fasi di trasformazione di questi, di arricchimento semantico e di integrazione. La prima è una fase cruciale e si realizza attraverso la trasformazione delle basi di dati di diversa natura in grafi semantici, come il modello RDF, che possono poi essere arricchiti semanticamente allo scopo di facilitarne l'integrazione e migliorarne la fruizione finale da parte degli utenti. La trasformazione di basi di dati diverse in campo bibliografico avviene con RDA, ma il principio è lo stesso e lo scopo è sempre l'integrazione e l'arricchimento semantico del dato. Dunque, non può snaturare la tradizione catalografica ma semmai collocarsi in una linea di sviluppo della catalogazione come metodo di organizzazione della conoscenza. RDA è uno standard basato su FRBR e FRAD, creato come evoluzione delle regole di catalogazione AACR2 per metadattare le risorse bibliografiche digitali, in modo del tutto compatibile con i record catalografici precedentemente presenti negli OPAC. Non stiamo parlando di cose diverse, ma solo di un naturale processo di evoluzione.

Valdo Pasqui

Absolutamente no! Anzi è ormai giunto il momento di abbandonare completamente anche nel linguaggio corrente il concetto di record. Tale termine non ha più senso, in primo luogo tecnicamente, poiché i componenti (item, singoli metadati o attributi) che lo compongono sono dispersi nelle strutture dati sottostanti (tabelle di database relazionali, strutture dati di database non relazionali, indici ecc.) ed eventualmente vengono ricomposte sotto forma di record MARC\* solo quando viene chiesto tale tipo di visualizzazione o esportazione. In secondo luogo, non ha più senso anche da un punto di vista logico, perché, data una opera/oggetto/risorsa descritta da una molteplicità di metadati, posso scegliere quali visualizzare/esportare e in quale formato (per esempio, RDF piuttosto che record MARC o MARC in formato XML ecc.).

Tiziana Possemato

Non so immaginare una tradizione catalografica che prescindere dalle funzioni del catalogo: *trovare, identificare, selezionare, ottenere e navigare*. Se le fonti informative utili per un utente sono uscite fuori dall'ambito ristretto del catalogo, tentare di contenerne i limiti entro i confini tradizionali rischia di rivelarsi un pericoloso errore

di valutazione. L'*entity modeling* e l'attività di metadattazione a esso collegata prevedono necessariamente un arricchimento dei dati riferibili a ciò che sto descrivendo con elementi provenienti dall'esterno, al di fuori dei propri confini, ove questi dati si trovino. Il che non sminuisce affatto l'apporto o il ruolo del catalogatore, ma lo trasferisce su un piano diverso: la capacità di costruire, con dati propri o provenienti dall'esterno, una descrizione ricca ed esaustiva, utile per rispondere ai requisiti richiesti a un catalogo ma nello stesso tempo riutilizzabile da molti. Per fare questo le regole di catalogazione, gli standard, i linguaggi di metadattazione si evolvono e si arricchiscono, ma non si snaturano: le competenze dei catalogatori si modificano, come la storia di questa disciplina ha sempre richiesto che avvenisse, ma senza perdere nulla in termini di specializzazione e autorevolezza. I cataloghi di buona qualità o di pessima qualità sono sempre esistiti, e un cambio di ottica e di pratica nella loro costruzione, come l'incontro con il web sta richiedendo, non migliora né peggiora le cose. Le ripropone su più larga scala, richiedendo un'esposizione che quasi costringe a innalzare il livello di strutturazione e qualità dell'informazione e, dunque, delle competenze.

Roberto Raieli

Il prodotto della descrizione "catalografica" passa dalla scheda a sé stante, "monolitica", al record delle entità "in relazione", consentendo un approccio più ampio e globale alle risorse, inquadrando come insiemi di dati collegati, da trattare, diffondere, condividere. La prospettiva tradizionale del record bibliografico è sostituita da quella del "set di dati", composto dal collegamento di dati in sé autonomi. Il set di dati non necessariamente di natura bibliografica, riutilizzabile anche in ambito bibliografico, appare come un'opportunità di completezza, nonché di maggiore qualificazione delle informazioni raccolte e aggregate nella descrizione bibliografica. Tale informazione isolata, però, come già detto, può appartenere a tutti i contesti possibili, e può essere perso il contesto originario di appartenenza di una risorsa. La condivisione, quindi, resta un valore elevato, per le tante ragioni prima esposte, ma essa – come la collegata cultura della "convergenza" – non deve andare a danno della riconoscibilità della risorsa stessa che è condivisa insieme ai "propri" dati. Se il paradigma dei LOD fornisce la base alle comunità bibliografiche, archivistiche e museali, insieme a molte altre comunità più o meno strettamente culturali, per realizzare la condivisione dei patrimoni e della cultura, ognuna delle comunità coinvolte deve svolgere il proprio ruolo, valorizzare la propria autorità, per raggiungere una convergenza che non sia omologazione, appiattimento o dispersione. Ogni contesto deve essere individuabile

e affidabile, precisamente definito, e in grado di rapportare e contestualizzare al proprio interno e al proprio esterno i diversi dati informativi che lo costituiscono. Anche se la rinuncia alle chiusure della specializzazione, alla “firma” sui dati, è un elemento di sincera volontà di condividere tra i diversi organismi, non si deve rinunciare all’identità, alla presenza del “carattere” con cui contraddistinguere una risorsa.

#### *Maurizio Vivarelli*

Alla luce anche di alcuni elementi presenti nelle considerazioni proposte in precedenza, la mia impressione di sintesi è che, nel presente e nel futuro, continueranno a essere presenti record sia bibliografici sia non bibliografici, e più in generale documentari ed extradocumentari. In questo senso l’apertura della prospettiva bibliografica, con i suoi principi e i suoi metodi, credo che rappresenti un elemento di arricchimento. Il problema che si apre è allora quello di governare e interpretare questa ricchezza di contenuti informativi eterogenei e diversi, alcuni rappresentati secondo modelli bibliografici, nel senso in cui in questa sede stiamo parlando, e altri secondo modelli non bibliografici.

*Si potrebbe pensare (o temere) che se i metadati con le loro connessioni costituiscono una struttura reticolare, non più rigidamente gerarchizzata, ove ogni nodo della rete può costituire punto d’accesso, allora tutto è un grande frullato in cui l’autorevolezza si perde?*

#### *Stefano Bargioni e Alberto Gambardella*

Il pericolo del frullato esiste laddove non c’è una sufficiente conoscenza e consapevolezza nella creazione dei legami; tale competenza, che va sviluppata, può rappresentare uno dei nuovi percorsi della professione. Se per esempio si tratta di costruire un legame tra una persona e un suo profilo biografico, è possibile scegliere di creare un collegamento con la voce presente nel “Dizionario biografico degli italiani online” oppure con il profilo biografico compilato da una casa editrice, o altro. Questa decisione, che dipende dal contesto d’uso, crea la differenza. La rete, come ben sappiamo, offre un range di informazioni qualitativamente ampio. La valutazione delle fonti e la loro scelta appropriata, costituisce sicuramente una delle nuove competenze lavorative. Inoltre, anche la presenza di linee guida interne all’agenzia, che RDA chiama *application profiles*, gioca un ruolo di equilibrio all’interno di questo sistema.

#### *Giovanni Bergamin*

È sicuramente un aspetto che va affrontato. I metadati pubblicati come linked data offrono la possibilità di in-

dicare la fonte e il contesto di qualsiasi dichiarazione (o tripla). In altre parole, è possibile offrire elementi che permettono a chi accede ai dati di verificarne l’autorevolezza. Per il riuso delle triple tra domini differenti questa pratica è assolutamente necessaria anche se ancora non largamente diffusa.

#### *Calro Bianchini*

La mancanza di gerarchia è un principio di democrazia, ma allo stesso tempo non significa mancanza di controllo. Chiunque può produrre i propri dati ed esporli sul web, esattamente come oggi chiunque può creare un proprio sito web e popolarlo di qualsiasi contenuto. Il web dei documenti pone problemi di autorevolezza, che non sono risolvibili tramite automatismi, ma sulla base di scelte consapevoli. Allo stesso modo, nel web semantico l’autorevolezza dei dati si basa sulla scelta, consapevole, di fonti di dati autorevoli da parte dell’istituzione che decide di presentare i dati in una forma arricchita. Anche se è vero che la struttura del web semantico è reticolare, per il principio che ogni “cosa” (per citare un termine usato da Tim Berners-Lee, ma anche da IFLA LRM!) deve essere identificata sul web, è possibile stabilire con certezza la fonte di ogni singolo dato. Dal punto di vista dell’autorevolezza, è possibile essere molto più precisi nella presentazione consentita dai metadati che in una normale registrazione bibliografica o di autorità. Per esempio, in alcune voci di autorità di SBN vengono presentati i dati e, separatamente, nel campo *Fonti* vengono indicate le fonti (per esempio, <https://opac.sbn.it/nome/RAVV042068>). In questo caso, non è possibile stabilire da quale delle fonti citate è stata presa la data di nascita, o quella di morte, o la forma variante del nome ecc. Nella metadateazione, invece, per ogni singolo dato viene specificata la fonte precisa (o più fonti) di quel dato: una o più fonti per la data di nascita, per la data di morte, per il luogo di nascita, per il luogo di morte ecc. Inoltre, con i metadati è possibile rappresentare anche eventuali differenze tra le fonti (per esempio, due fonti che indicano una data differente, magari perché sono indicate due date diverse o perché una fonte ha la precisione al giorno e un’altra fonte ha una precisione all’anno soltanto). In una registrazione tradizionale, invece, il catalogatore fa inevitabilmente una scelta, una sintesi, che non necessariamente è corretta, che può diventare facilmente obsoleta e che, soprattutto, non informa in modo sufficiente il lettore del problema. Un esempio: si confronti il trattamento della data di nascita e del paese di Ludwig van Beethoven in <https://opac.sbn.it/nome/CFIV005324> e in <https://www.wikidata.org/wiki/Q255>.

#### *Pierluigi Felicciati*

La moltiplicazione dei punti di accesso, più o meno

generalisti, costituisce una tendenza del tutto indipendente dalla catalogazione basata sulla metadattazione, come ho accennato di sopra. L'impatto virtuoso della cattura ed esposizione di linked open data sulla qualità dei dati in rete è soprattutto indiretto, credo. Rispetto all'accesso diretto ai dati, non è certo pensabile che gli OPAC o le interfacce di accesso alle descrizioni si trasformino in access point SparQ! Il processo di evoluzione dei motori di ricerca commerciali e degli OPAC integrati verso l'inclusione intelligente delle fonti dati esposte come LOD è un processo inevitabile e in corso. In questo senso, è da sperare che gli investimenti dei big players, più attenti delle istituzioni culturali, di solito, alla facilità d'uso delle interfacce velocizzino tale processo. L'obiettivo non è selezionare e leggere dei LOD se non come preziosa fonte di supporto al controllo d'autorità oppure per facilitare lo scambio dei dati, ma accedere con più efficacia al *knowledge graph*, alla conoscenza registrata, valorizzata semanticamente dagli agenti intelligenti e sempre più liberata da restrizioni d'accesso, smuovendo il protezionismo alle volte reazionario degli editori.

#### *Claudio Forziati*

La possibilità di avere molteplici punti di accesso non ha di per sé lo scopo, né necessariamente la conseguenza, di ribaltare l'organizzazione gerarchica delle informazioni. Ogni risorsa, di qualsiasi dominio, in una corretta metadattazione, porta con sé uno strato articolato di informazioni strutturate che non si perdono solo perché possiamo utilizzare strumenti che ci consentono di evidenziare le relazioni di specifiche porzioni di esse in uno scenario differente da quello nativo. Per quanto i metadati possano descrivere una risorsa creata per uno scopo individuabile, essi hanno la funzione di rendere esplorabili le informazioni di contesto che enfatizzano le relazioni intercorrenti tra tutti gli "attori" e i concetti che quella risorsa rappresenta, senza escludere la possibilità, per queste informazioni, di ricombinarsi attraverso similarità, identità e differenze in domini molto distanti. Non percepisco perciò un rischio concreto di dispersione o perdita di autorevolezza; credo invece che meritino particolare attenzione oggi gli strumenti sviluppati da e con le biblioteche per rappresentare "graficamente" le serializzazioni dei grafi RDF, nei quali ho l'impressione manchi ancora l'orientamento all'utilizzatore umano e alle sue necessità. Grazie a una usabilità migliorata delle interfacce oggi vediamo esposti efficacemente i livelli sovraordinati, mentre percepisco ancora come tortuoso il percorso che l'utente è costretto a fare per arrivare dall'opera al possesso concreto dell'item. Mi sembra ancora realistico quello che afferma il *Manifesto per il Web Semantico*, curato dal gruppo

di studio AIB CILW, quando dice: "dobbiamo ancora sviluppare strumenti amichevoli e usabili che permettano agli utenti di utilizzare tali dati, in modo comparabile ai browser che hanno contribuito al successo del World Wide Web".

#### *Maurizio Lana*

All'interno di una struttura reticolare l'autorevolezza deve essere cercata e riconosciuta. Cioè applichiamo ai dati il medesimo approccio che abbiamo appreso e adottiamo già nel mondo fisico nei confronti delle risorse informative: ricerca e valutazione. Si cercano le risorse, si analizzano, si valutano, si trattengono quelle valide e su di esse si costruisce lo studio. Ovviamente possiamo giudicare che per una serie di ragioni una struttura rigidamente gerarchizzata sia preferibile per gestire, organizzare, usare, le risorse informative. Il problema, se così si può dire, è che il mondo digitale come noi lo conosciamo è caratterizzato dall'organizzazione reticolare dei dati, quindi, in prospettiva, non si perde l'autorevolezza, ma essa si esprime in nuove forme. La crescita del numero delle pubblicazioni per unità di tempo, nell'universo della stampa, ha presentato anch'esso una sfida all'autorevolezza: c'è stato un tempo non remotissimo in cui ogni studioso di un campo disciplinare poteva conoscere tutti i suoi colleghi e avere conoscenza di tutte le loro pubblicazioni. Potremmo dire che c'era un'analogia tra la struttura gerarchizzata della catalogazione e l'autorevolezza degli studiosi nel campo disciplinare: "tutto è noto e sotto controllo". Con il crescere del numero degli studiosi e del numero delle loro pubblicazioni, anche se la struttura della catalogazione continua a essere rigidamente gerarchizzata, per lo studioso si presenta la necessità di individuare e valutare le risorse informative che la catalogazione ha ordinato e reso disponibili.

#### *Rossana Morriello*

I metadati servono proprio a tenere insieme e collegare i diversi nodi della rete. Siamo nell'epoca della complessità e dobbiamo affrontarla, non possiamo pensare di semplificarla. La definizione della Treccani ci spiega come la parola "complesso", nel suo significato etimologico di *complexus* (e contrario di "semplice"), si riferisca a qualcosa che risulta dall'unione di più parti o elementi, tessute o tenute insieme. Non vi è altro modo di affrontare la complessità documentaria e in particolare del mondo digitale se non creando strumenti che consentano di tessere le parti e tenerle unite. Questo lo si può fare per mezzo di una struttura reticolare, e non è più pensabile usare strutture gerarchiche poiché è evidente che esse non sono in grado di cogliere tale complessità né di rappresentare efficacemente le caratteristiche del

mondo digitale. Ma sarebbe sbagliato pensare che una struttura reticolare sia meno funzionale o meno autorevole. Come ho scritto sopra rispondendo a un'altra domanda, per le macchine il concetto di punto di accesso è relativo e la struttura gerarchica non è importante, ed era già così con gli OPAC, eppure i record catalografici dei cataloghi online rivestono la stessa autorevolezza e funzionalità delle schede cartacee. Non è la gerarchia a dare autorevolezza a un sistema di organizzazione della conoscenza e di certo non lo può essere nel multiverso documentario contemporaneo. L'autorevolezza si crea con i metadati di qualità.

*Valdo Pasqui*

L'autorevolezza dipende dalle fonti che generano i metadati e soprattutto da quelle che governano la definizione dei set di metadati (per esempio, DC, BIBFRAME, RDA) e i domini di valori controllati (authority file, schemi di classificazione) assicurando la coerenza dei rispettivi identificatori, la loro persistenza nel tempo e la loro evoluzione consistente.

*Tiziana Possemato*

L'immagine di un reticolo informativo, costituito da nodi e archi di collegamento, è sicuramente affascinante, e in concreto genera un grafo in cui ogni piccola componente (tripla) contribuisce ad arricchire l'informazione ma a volte anche a complicarne l'utilizzo: ove inizi una catena di dati utile a soddisfare il mio bisogno informativo e dove questa finisca, è certamente complicato da definire in un contesto in cui salta il concetto "rassicurante" del confine fisico, che possa controllare. E, infatti, le tematiche di gestione di queste strutture dati sono ampie e complesse, spaziando dai quesiti sulla copia o il collegamento a dati esterni nelle fasi di arricchimento, alle questioni sugli aggiornamenti, a quelle sulla permanenza degli identificatori. Il tema dell'affidabilità e autorevolezza dei dati è uno di quelli più attuali e discussi in questo ambito, ma è anche uno di quelli per i quali una soluzione condivisa esiste ed è sempre più largamente applicata: la trasformazione della tripla in quadrupla, dove il quarto elemento è costituito dalla *provenance*, cioè l'autore dell'asserzione. Non avrò più un'affermazione del tipo "questa risorsa è una monografia" ma "l'istituzione x dichiara che questa risorsa è una monografia": un'assunzione di responsabilità che serve a esigere il rigore dell'asserzione e la sua certificabilità. Anche il meccanismo del *cluster* come rappresentazione di un'entità nelle sue possibili sfaccettature espresse in varianti del nome, che è di per sé un meccanismo democratico e antigerarchico per eccellenza, in realtà deve confrontarsi, per esempio nella presentazione del dato, con la possibile scelta di una forma *preferita* (non

più *autorizzata*), che può cambiare a seconda di differenti contesti culturali, geografici, linguistici o di scopo (finalità di ricerca o di gestione del dato).

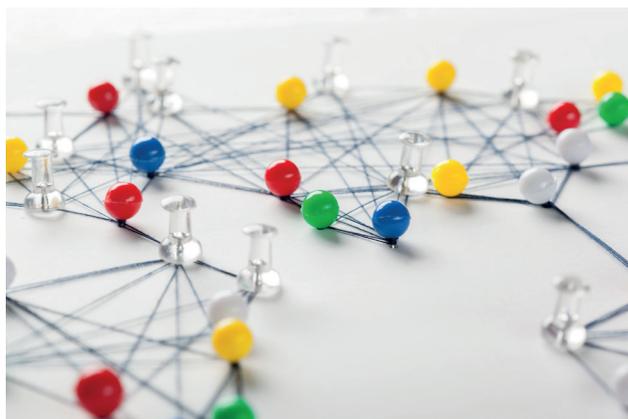
*Roberto Raieli*

Richiamando la risposta alla domanda precedente, non si deve perdere la "riconoscibilità" della risorsa stessa, per quanto i relativi dati possano essere separati e condivisi all'infinito. Ognuna delle comunità coinvolte, si ripete, deve valorizzare la propria autorità, ogni contesto deve essere individuabile e affidabile, precisamente definito dalla comunità che ne è responsabile, che deve rapportare e contestualizzare all'interno e all'esterno i diversi dati informativi che costituiscono una risorsa della conoscenza del proprio ambito. Il rischio è che gli utilizzatori dei contenuti della risorsa cui fanno riferimento i dati granularizzati possano iniziare a vedere i documenti "coerenti" originari, nati digitali o meno, quasi solo come "contenitori" di contenuti da disaggregare, per creare intorno a questi percorsi semantici spesso non "forti", per riaggregare ogni volta liberamente nuovi contenitori, senza un'opportuna considerazione dei vincoli semantici che reggono i contenuti originariamente creati fuori da tale logica di riutilizzo, creati in base ad altri schemi di strutturazione del "senso". L'attenzione verrebbe posta, in tal caso, solo sull'"informazione" intesa in generale e in assoluto, e non sulle "fonti" della conoscenza, sulla rete dei collegamenti e non sui contenuti collegati, sui dati e non sulla loro interpretazione, sulle descrizioni e non sul "descritto". Anche nei nuovi sistemi della rete e delle risorse digitali, dunque, si deve sviluppare la ricerca di soluzioni per garantire ai dati e ai relativi costrutti una validazione di autorità, di datazione, di integrità, nonché di coerenza semantica, secondo principi riferibili al mondo dei documenti "classici". La necessità è quella di contenere la "liquidità" dell'informazione, la decontestualizzazione e la reinterpretabilità perpetue, tramite sistemi dotati di quell'affidabilità caratteristica delle metodologie della tradizione bibliografica.

*Maurizio Vivarelli*

Il problema del modello di organizzazione della conoscenza è antico quanto la nostra tradizione culturale, e si correla all'antichissima questione delle relazioni tra "ordine" e "disordine"; credo che in questa nostra stagione, postmoderna e secondo alcuni anche postumana, sia indispensabile prendere atto, anzitutto, di questa complessità, entro la quale i dati non solo sono eterogenei e diversi, strutturati e non strutturati, normalizzati e non normalizzati, ma anche prodotti sia da agenti umani che da agenti artificiali. D'altra parte, il concetto di "autore" e di "autorevolezza" a esso strettamente col-

legato, ha iniziato la sua crisi ben prima della diffusione delle culture e delle tecnologie digitali. A mio parere, in questa fase di grandi e profonde trasformazioni, è dunque in primo luogo necessario porsi in una condizione di ascolto, e in sostanza di cercare di interpretare alcune delle caratteristiche principali della complessità contemporanea. In questo senso uno dei problemi che emergono è quello qui indicato, e che riguarda le relazioni tra *gerarchia* e *reticolarità*; la prima poggiava sul concetto e sul modello di “sistema”, che potremmo definire come un insieme di parti interagenti finalizzato a uno scopo; la seconda come caratterizzata da nodi e relazioni che cambiano in base al variare del punto di vista secondo il quale vengono trattate. Per questo motivo credo che sia indispensabile cercare di capire anzitutto che cosa le reti siano, e iniziare ad acquisire confidenza con i principi, i modelli, le tecniche delle quali è possibile avvalersi per studiare le reti: diciamo che anche la nostra comunità disciplinare dovrebbe iniziare ad acquisire maggiore confidenza con la network science. Questo è il motivo per cui da tempo sostengo che possa essere utile acquisire elementi di consapevolezza metodologica di questa possibile “scienza” delle reti, descritta ad esempio, a un livello di alta divulgazione, nelle opere del fisico statunitense Albert-László Barabási.



***L'utilizzo dei metadati per produrre sistemi di raccomandazione per la lettura è un rischio o un'opportunità? Aumenta la possibilità di disintermediazione e di distacco dal bibliotecario, oppure aiuta a rendersi autonomi nella ricerca, almeno in certi contesti?***

*Stefano Bargioni e Alberto Gambardella*

La maggiore esposizione dei dati culturali ai motori di ricerca comporta certamente una maggiore possibilità di profilazione dell'utente anche per quanto riguarda i suoi interessi culturali, e come tale può essere considerata un rischio per la libertà individuale. D'altra parte,

apre all'utente nuovi percorsi di ricerca (serendipità) nel solco della pratica della navigabilità così connotata al web semantico. La possibilità di registrare i dati in maniera più granulare costituisce uno dei cardini della metadattazione, ma ciò non implica che si vada necessariamente verso la disintermediazione e il distacco dal bibliotecario. Sicuramente cambia il ruolo del bibliotecario – e soprattutto del servizio di reference – nel senso che l'OPAC locale si allarga fino a poter offrire percorsi per accostare l'utente anche a risorse non possedute dalla biblioteca. L'abbondanza dell'offerta dei dati opportunamente strutturati, la costruzione di percorsi strutturati di ricerca, le potenzialità offerte dalla modularità e interscambiabilità dei dati costituisce piuttosto il nuovo campo di forza della biblioteca rispetto al “rumore” disordinato della rete.

*Giovanni Bergamin*

Le biblioteche non rientrano evidentemente nelle logiche oggi predominanti, ovvero chiedere dati in cambio di servizi (sfruttamento commerciale dei dati generati dagli utenti), ma stanno prendendo in considerazione ipotesi di trattamento di questo tipo di dati per migliorare i servizi. Più in generale si parla di *analytics* di una metodologia di analisi di grandi quantità di dati (*big data*) con le tecnologie dell'intelligenza artificiale con l'obiettivo di promuovere azioni, soprattutto nel campo del marketing, come ad esempio personalizzazione dell'interfaccia, suggerimenti ecc. Si parla soprattutto di azioni *data-driven* (o guidate dai dati). L'ultimo fascicolo – gennaio 2022 – di TILT (“Trends and issues in library technology”) della Information Technology Section dell'IFLA è dedicato proprio ai big data. In particolare, un articolo porta il titolo *You may like: how the National Library Board of Singapore uses machine learning to recommend books* e parla dell'uso di *Amazon personalize* nella loro biblioteca. *Amazon personalize* si definisce come “un servizio di machine learning completamente gestito che consente agli sviluppatori di offrire in modo semplice esperienze personalizzate agli utenti”. Per quanto riguarda i big data e in particolare il sistema delle raccomandazioni le biblioteche sono ancora agli inizi e dovranno sicuramente trovare la loro strada tra pericoli ben noti in letteratura quali il “capitalismo della sorveglianza” di Zuboff o la “bolla di filtraggio” di Pariser.

*Rossana Morriello*

È di certo un'opportunità, a condizione che venga governata dalle biblioteche e non lasciata agli operatori commerciali. In passato, il dibattito biblioteconomico è stato centrato per lungo tempo sul rischio di disintermediazione. Ma non ci si è resi conto di come, al contrario,

i grandi colossi commerciali che operano nel mondo editoriale e dell'informazione, nel frattempo, stessero diventando tali proprio rafforzando l'intermediazione. Si tratta, in questo caso, di un'intermediazione nascosta e non percepita dagli utenti, ma forte, molto più forte di quella che fanno le biblioteche. Non siamo affatto nell'epoca della disintermediazione ma siamo invece nell'epoca dell'iperintermediazione (o della rimediazione, se considerata dal punto di vista dei media, in linea con la definizione coniata da Jay D. Bolter e Richard Grusin). Google fa proprio questo, opera un'efficace mediazione tra le esigenze dell'utente e l'universo documentario di cui dispone in quanto indicizzato dal motore di ricerca. Con questo non intendo dire che le biblioteche debbano inseguire i modelli commerciali, ma che devono cominciare a utilizzare anche le tecnologie informatiche usate dagli altri operatori, come i sistemi di filtraggio e di raccomandazione, per riconquistare la loro funzione di mediazione. Tali sistemi e tecniche rappresentano una soluzione all'*information overload* e sono, dunque, utili per gli utenti.

#### *Valdo Pasqui*

Rappresenta un'opportunità e un valore aggiunto. Tuttavia, come in tutte le modalità social, richiede forme adeguate di moderazione e controllo per evitare degenerazioni e abusi. Questo prefigura anche un ruolo nuovo per alcuni bibliotecari.

#### *Tiziana Possemato*

Questa è una sfida che le nuove tecnologie del web si pongono ma è, a mio avviso, ancora presto per valutare: i progetti di adeguamento filosofico e tecnologico ai modelli e linguaggi del web, in ambito culturale, per ora sono concentrati soprattutto sulla trasformazione dei dati, con processi di conversione dei cataloghi dai formati tradizionali all'RDF. Ma tutta una serie di funzioni pure largamente previste dal web semantico, prima fra tutte quella dei processi inferenziali che potrebbero generare nuova conoscenza e che dovrebbero favorire l'autonomia della ricerca da parte degli utenti, è ancora da sperimentare in pratica. Per ora, anche i più importanti progetti di pubblicazione dei dati in LOD, prevedono una certa dose di arricchimento del dato di origine, ma con ancora bassa applicazione di algoritmi di logica inferenziale. Il che ha ovviamente una sua motivazione radicata nella giovane età di questi progetti, nella concentrazione ancora focalizzata sull'obiettivo di portare il catalogo nel web. La componente dei progetti relativa alla presentazione e fruizione del dato è ancora molto legata alla tradizione di pubblicazione degli OPAC, con i dati di origine arricchiti, sì, ma ancora poco collegati con altre fonti. Non so, dunque, dare una risposta defi-

nitiva su questo, perché l'interessante di questi dichiarati e in alcuni casi ancora solo potenziali incontri di mondi informativi diversi, ciascuno con la propria specificità e la propria ricchezza deve, secondo me, ancora arrivare.

#### *Roberto Raieli*

Pur avendo approfondito poco tale argomento, posso dire che sarà sempre necessario spiegare agli utenti le strategie di ricerca, la valutazione critica dei risultati, e i rischi dell'eccessiva disintermediazione, nonché definire il livello in cui i sistemi automatici restano limitati all'essere strumenti per una prima ricerca e scoperta delle risorse, e il livello in cui possono consentire di raggiungere il risultato finale – quali alcuni prototipi fondati sull'intelligenza artificiale. Le possibilità di essere ben "consigliati" ci sono, ma il rischio presente nelle elaborazioni della rilevanza algoritmica dei risultati delle ricerche è quello di lasciare troppo alla mediazione del software – scambiata spesso per "semplice" disintermediazione – e dimenticare l'utilità dei dati di qualità inseriti dai bibliotecari, nonché dei percorsi di ricerca da essi guidati. La "mediazione" di una raccomandazione robotica è ben più pericolosa della disintermediazione che si crea nelle ricerche autonome con uno strumento ben conosciuto, che non si "permette" di consigliare, ma si limita a elaborare e soddisfare una ricerca. È necessario, quindi, chiedersi fino a che punto l'estrema dinamicità degli strumenti debba essere libera di svilupparsi, e quanto debba essere controllata nell'ambito di un compromesso da raggiungere tra bibliotecari, utenti e produttori dei sistemi. Serve necessariamente svelare quanto all'ingombrante mediazione delle biblioteche si stia sostituendo una anche più ingombrante mediazione dei software, per di più nascosta e travestita da disintermediazione. Nell'insieme, la biblioteca, anche come semplice "indice", media la propria raccolta fisica unitamente a una raccolta virtuale finita e infinita, assumendo nell'uno e nell'altro caso una precisa responsabilità e identità culturale. La mediazione, ben altrimenti che scomparire in quella sorta di disintermediazione implicata dai nuovi strumenti di ricerca, è più che mai presente nelle azioni stesse che l'utente compie attraverso i mezzi indicati dalla biblioteca, siano i discovery tool o altri motori di ricerca. Ogni biblioteca, dunque, dovrà anzitutto preoccuparsi di stabilire e "insegnare" agli utenti le potenzialità e l'utilità dei nuovi strumenti in rapporto a quelli tradizionali, di spiegare loro le differenze che li fanno restare "distinti" per distinti scopi, così come di spiegare le strategie di ricerca, la valutazione critica dei risultati e i rischi dell'eccessiva disintermediazione, o della mediazione dei consigli delle macchine.

*Maurizio Vivarelli*

La risposta a questa domanda mi coinvolge in misura molto consistente, visto l'impegno dedicato negli ultimi mesi al progetto Reading(&)Machine, che si basa proprio sullo sviluppo e la prototipazione di un sistema di raccomandazione, che a sua volta, si basa appunto sia su metadati bibliografici che su dati non bibliografici, provenienti dalla piattaforma di social reading aNobii e anche da social network generalisti. Si tratta di una scelta che ha proprie caratteristiche peculiari, e che è strutturalmente diversa, ad esempio, da quella utilizzata nel progetto finlandese BookSampo, che prevede invece l'utilizzo esclusivo di metadati bibliografici, riconducibili alla struttura concettuale del modello FRBR. Il problema di fondo con il quale dobbiamo confrontarci è tuttavia costituito dalle diverse tipologie di lettura che vengono praticate nello spazio bibliografico della biblioteca. Una di queste è senz'altro quella che potremmo definire *lettura documentaria*, che ha per oggetto i metadati, e che dunque si situa esclusivamente all'interno dell'universo bibliografico. Ma, nello spazio bibliografico, le persone utilizzano e praticano molte altre forme di lettura, che in un articolo in corso di stampa su "AIB studi" ho provato sommariamente a mappare: la *lettura paratestuale*, *sincretica*, *letteraria*, e altre ancora. Insomma se vogliamo utilizzare *anche* i metadati bibliografici per arricchire e promuovere l'esperienza della lettura è indispensabile fuoriuscire dai confini dell'universo bibliografico, e avventurarsi (non solo ma anche) in quello che personalmente ho provato a chiamare Multiverso bibliografico, che relaziona i contenuti, cognitivi ed emotivi, che caratterizzano l'esperienza e le pratiche dei Lettori Empirici, visti nella realtà della concretezza della propria specifica esperienza di lettura. Si aprono, dunque, in questo modo nuovi modelli di mediazione, in questo senso ampliato, che la comunità bibliotecaria, a mio giudizio, dovrebbe inserire nella propria agenda, anche solo a partire dalla presa d'atto che questo paesaggio documentario e cognitivo è quello che può essere con facilità individuato all'interno delle biblioteche reali, e tra i lettori empirici che in esse si muovono.

**Quali sono i modelli formativi per il bibliotecario che si occupa di metadattazione; quali competenze?**

*Stefano Bargioni e Alberto Gambardella*

Premesso che i modelli formativi per la metadattazione hanno diversi punti di contatto con le competenze tradizionali, come la buona conoscenza dei formati quali MARC/BIBFRAME e standard come ISBD ecc., a queste vanno aggiunti:

- la pianificazione di linee guida per la registrazione dei dati in quanto non tutti i dati vanno registrati, la loro presenza o assenza va impostata in relazione al tipo di biblioteca e all'utenza;
- una conoscenza costante e aggiornata dei repertori, liste controllate, banche dati, siti di svariata natura disponibili in rete così da creare un sistema funzionale e condiviso di gerarchia delle fonti utili a una registrazione corretta e puntuale del dato;
- la costruzione di sistemi di registrazioni di dati con menù a tendina, possibilmente in riferimento a "oggetti reali" (GeoNames), sistemi di scrittura autosuggest che garantiscono l'uniformità del dato;
- la capacità di registrare i dati, laddove viene usato il linguaggio naturale, in maniera specifica, chiara e concisa;
- la capacità di creare legami secondo modelli gerarchici che seguono delle logiche interne stabilite;
- l'utilizzo di strumenti di lavoro duttili, ovvero in grado di essere modificati dall'operatore come, per esempio, accade in Wikidata dove le liste controllate possono essere modificate e corrette.

Tali competenze vanno sviluppate secondo logiche strettamente connesse all'identità della biblioteca (quali risorse possiede) e alla tipologia degli utenti che la frequentano.

*Giovanni Bergamin*

Un bibliotecario che si occupa di metadattazione (o di organizzazione dell'informazione) non deve essere un ingegnere informatico, ma sicuramente, oltre alle competenze di dominio, dovrebbe anche avere una buona familiarità con i vari strumenti software di trattamento e gestione dei dati; dovrebbe avere inoltre nozioni di base sugli strumenti di recupero dell'informazione e sui problemi di accesso nel lungo periodo alle risorse digitali; dovrebbe infine conoscere anche i meccanismi di base delle applicazioni statistiche.

*Carlo Bianchini*

In una biblioteca, i primi e principali metadati da produrre sono quelli necessari al funzionamento dell'ecosistema bibliografico. Perciò, le competenze devono necessariamente includere quelle catalografiche tradizionali, perché i metadati devono essere usati per produrre anche uno strumento "tradizionale". A queste se ne devono aggiungere altre, nuove, come richiesto inevitabilmente dalla quinta legge della biblioteconomia. Ciò che è cambiato è l'ecosistema intorno alla biblioteca e alle sue collezioni. La dieta informativa dei lettori si è modificata profondamente, nei contenuti e soprattutto nei canali; la biblioteca deve essere presente, visibile

e raggiungibile anche attraverso il web semantico. È per questo motivo che si passa dalla catalogazione alla metadattazione: per poter fare di più con gli stessi dati. Quindi, come sempre è successo nella storia delle biblioteche, le nuove competenze sono legate alla comprensione e all'uso dei nuovi strumenti che abbiamo a disposizione per fornire un servizio migliore ai nostri lettori.

#### *Maurizio Lana*

In termini concisi si potrebbe sostenere che il bibliotecario che si occupa di metadattazione deve essere *information literate*. Questo potrebbe apparire come una tautologia (ogni bibliotecario è per formazione e per definizione un soggetto *information literate*) ma non è così. Il bibliotecario è per formazione *information literate* nell'universo della biblioteca. Ma nel digitale l'universo delle risorse informative è (molto) più ampio dell'universo della biblioteca "tradizionale"; e questo universo si caratterizza per essere reticolare e non gerarchizzato. E con questi due elementi (e anche con altri) permette di fare i conti *l'information literacy*. L'ampiezza al di là delle possibilità di controllo da parte dell'individuo e l'assenza di gerarchie note, condivise creano incertezza, introducono limiti pratici all'esplorabilità dell'insieme delle risorse contenute nell'universo informativo. Come la *literacy*, l'alfabetizzazione, mette in condizione di lavorare con tanti "testi" differenti, così *l'information literacy* mette in condizione di lavorare con tanti tipi diversi di informazione. Dall'obiettivo di poter "conoscere tutto" che porta con sé (o che è espressione di) uno studioso-dio nel suo campo disciplinare, si passa all'obiettivo di "conoscere bene" (= in modo metodologicamente corretto) che porta con sé l'idea di una costruzione collaborativa del sapere a cui ogni studioso contribuisce per la sua parte. Come un edificio che si regge e dura perché tutti i fornitori hanno portato materiali di qualità per la costruzione.

#### *Rossana Morriello*

Un report del 2020 dell'Australian Library and Information Association intitolato *The Future of Library and Information Science in Australia*, specifica che un bibliotecario "deve essere un pensatore concettuale" (<https://read.alia.org.au/future-library-and-information-science-education-australia-discussion-paper>). La metadattazione sposta il livello dalle competenze puramente tecniche alla necessità di possedere un pensiero concettuale per affrontare e gestire i sistemi documentari complessi con modelli in grado di costruire la "struttura reticolare", l'ossatura del patrimonio culturale, di tessere le diverse parti del multiverso bibliografico, come fanno per esempio i linked open data. I bibliotecari non

devono creare software né saper creare un algoritmo perché sono gli ingegneri ed esperti informatici a farlo. Però i creatori di software e algoritmi si aspettano di avere nei bibliotecari degli interlocutori in grado di padroneggiare concettualmente l'organizzazione della conoscenza nel mondo analogico e digitale, di intervenire efficacemente nelle diverse fasi della gestione del dato bibliografico. Di conseguenza, la formazione richiesta non può che essere di alto livello. Nei paesi anglosassoni si definisce bibliotecario solo chi è in possesso di una laurea o di un master in Library and Information Science, mentre gli altri operatori sono tecnici di biblioteca o assistenti di biblioteca. La tradizione biblioteconomica italiana è diversa da quella anglosassone, ma non per questo presenta sfide minori, anzi la stratificazione culturale che caratterizza il nostro paese consente potenzialmente prospettive molto più ampie e allo stesso tempo più complesse. Per tessere insieme le diverse parti di un sistema complesso, per esempio attraverso la metadattazione, la conoscenza delle singole parti e delle singole tecniche non è sufficiente, ma serve una visione olistica capace di mettere in relazione le diverse parti. Dunque, al bibliotecario che si occupa di metadattazione serve una formazione che offra un'ampia preparazione culturale generale e conoscenze profonde del mondo documentario, analogico e digitale, e dei metodi per organizzarlo.

#### *Valdo Pasqui*

XML, RDA, RDF, concetti del web semantico, motori di ricerca, discovery tool, strumenti e approcci di AI.

#### *Tiziana Possemato*

Il cambiamento di approccio alla creazione e distribuzione delle informazioni nell'ambito della biblioteconomia sta generando un profondo cambiamento nel profilo professionale degli operatori del settore, con un avvicinamento sempre più prossimo a molte delle competenze tecnologiche richieste per operare nel web. Senza voler ripercorrere la storia della catalogazione e di come, nel tempo, abbia richiesto competenze sempre più specialistiche, tornerò a richiamare l'attenzione sui fattori che hanno generato una esplosione delle tipologie di metadati e una necessità di estendere la propria competenza oltre il dato strettamente catalografico. Questo fenomeno è in stretta dipendenza con la natura della biblioteca, il livello di servizio offerto, il tipo di utenza servita e la vocazione tecnologica espressa. È chiaro che in tantissime biblioteche la conoscenza della normativa catalografica nazionale (ammesso che abbia ancora senso parlare di una normativa catalografica nazionale), di un formato MARC e del software di catalogazione adottato, siano competenze già sufficien-

ti per portare avanti il proprio lavoro. Ma parliamo di contesti il cui orizzonte di riferimento è limitato alla biblioteca stessa e ai suoi servizi di base. Non appena si esca fuori da questi stretti limiti, ci si scontra con comunità e sollecitazioni professionali ben più ampie e complesse. Solo per citare due esempi: la University of Alberta ha presentato a gennaio 2022 un *Linked data implementation plan*, un piano strategico quinquennale ad altissimo contenuto tecnologico. Il piano, non a caso, è presentato dalle figure apicali degli uffici di Cataloguing Strategies e Metadata Strategies. Nel 2020 la Bibliothèque Nationale de France ha pubblicato una Digital Rodmaps che include anche tecniche di intelligenza artificiale come supporto ai processi di catalogazione.

*Roberto Raieli*

In generale, il bibliotecario che si occupa di metadattazione, come faceva il “vecchio” catalogatore, deve entrare in simbiosi con tutte le altre professionalità della biblioteca, e i suoi strumenti in interazione con gli altri strumenti bibliotecari. Data la gestione *in cloud* di molti software, per i problemi tecnici si può fare sempre riferimento al produttore, ma il personale con competenze informatiche è necessario per una serie di questioni collegate, come le connessioni, i servizi di proxy e di autenticazione, le pagine web che ospitano i servizi, la gestione dei dati e la relativa pubblicazione. Anche i bibliotecari che si occupano delle banche dati, degli ejournal e degli ebook devono interagire costantemente tra loro e con l'addetto alla metadattazione, riguardo il controllo della ricercabilità e accessibilità delle risorse, gli aggiornamenti dei metadati eccetera. Stessa cosa vale per i bibliotecari che si occupano dei servizi del catalogo e degli altri database dell'istituzione, via via verso i servizi di reference, quelli di valutazione e quelli di istruzione degli utenti, per non parlare della “promozione” dei servizi. Se l'assistenza dei produttori, dei fornitori e degli editori è in genere sempre presente e relativa a ognuno di questi aspetti, molte decisioni e attività sono di tipo “politico” e non possono essere supportate con efficacia da un soggetto esterno. Il fatto di creare un gruppo di lavoro esteso e dinamico, dunque, è sempre collegato alle aspettative di impatto che sono riposte nell'intero servizio della biblioteca. Il valore aggiunto da parte delle biblioteche non cambierà nella sostanza e nei principi, e sarà sempre la capacità di valutare, selezionare e mediare tra le risorse disponibili, sarà la competenza per conoscerle e descriverle in modo appropriato all'utenza di riferimento. È necessario, infine, che tutti i professionisti dell'informazione si preparino a fornire sempre il migliore e più adeguato aiuto alle persone, qualunque livello di competenza informativa esse abbiano, per insegnare come scegliere

e valutare criticamente le informazioni, le risorse, gli strumenti di ricerca e scoperta, anche in ambiti diversi da quelli delle biblioteche, anche e soprattutto nel web.

*Maurizio Vivarelli*

Credo che da un lato sia necessaria anzitutto una adeguata preparazione che introduca alla storia della catalogazione bibliografica, dal punto di vista dei principi, dei modelli, delle procedure. Contestualmente credo che sia altrettanto necessaria una apertura di questa linea diciamo così *verticale*, ad alcuni dei suoi contesti e delle sue implicazioni *orizzontali*, e in primo luogo quelle che riguardano l'uso che le persone fanno di questi modelli di rappresentazione, all'interno dei propri comportamenti informativi. Infine, vedo uno strato ancora più esterno che potrebbe riguardare l'acquisizione di competenze, epistemologiche e riflessive, dedicate proprio al concetto di *modello di rappresentazione*, in modo da poter valorizzare la comprensione delle relazioni tra modelli linguistici diversi e plurali. Penso insomma, anche se in questa sede non è possibile trattare compiutamente l'argomento, che dovrebbe essere costituito, a livello di formazione accademica, uno strato inter e transdisciplinare, in cui convergano principi e metodi della tradizione storico-filologica, delle scienze sociali, delle culture digitali. Secondo questa prospettiva stiamo cercando di muoverci con il Corso di laurea magistrale in Scienze del libro, del documento, del patrimonio culturale dell'Università di Torino, all'interno del quale è programmato il corso Cultura bibliografica in ambiente digitale.

***Il processo della metadattazione modifica le relazioni tra i vari agenti coinvolti nella filiera del libro, stimola e favorisce una maggiore connessione tra editori e biblioteche che potrebbero finalmente collaborare nell'interesse dell'utente; cosa ne pensate?***

*Stefano Bargioni e Alberto Gambardella*

La collaborazione indubbiamente può arricchire entrambi anche se le biblioteche, al contrario degli editori, sono organismi non commerciali. Gli editori, per esempio, spesso sono in contatto diretto con gli agenti e quindi sono in grado di fungere da prezioso serbatoio di dati, come data, luogo di nascita, campo di attività ecc., spesso altrove irreperibili. Gli editori, che pubblicano in rete tali dati (per esempio i profili biografici di Aracne), contribuiscono e sostengono il compito “identificativo” svolto tradizionalmente dalle biblioteche. Chi non ricorda come in passato nei record di autorità relativi a persone, venisse apposto il “sigillo di qualità” scrivendo, in un apposito campo, “telefonato all'autore”? Per altri

versi, la metadattazione e la spinta alla granularità che la sostiene crea un circuito capillare di diffusione della cultura che indubbiamente può favorire gli editori. Va ricordato come già ora alcuni degli standard utilizzati in biblioteca, come per esempio il MARC21 per i record bibliografici, prevedono, attraverso l'utilizzo del campo 856, la possibilità di creare un legame tra la risorsa catalogata e il profilo dell'editore che l'ha prodotta.

#### *Giovanni Bergamin*

Occorre dire che almeno in Italia la collaborazione nella creazione di metadati tra editori e biblioteche è rimasta sempre una possibilità progettuale fin dai tempi dell'avvio del Servizio bibliotecario nazionale. Non sono mai nate tuttavia iniziative di collaborazione consolidate. I nuovi strumenti che facilitano la collaborazione e l'interesse dell'utente (ricordato nella domanda) potrebbero essere argomenti decisivi per affrontare in maniera innovativa e collaborativa il processo di metadattazione.

#### *Calro Bianchini*

Il passaggio alla produzione di metadati invece che soltanto di strumenti tradizionali meno flessibili – come i cataloghi – è un fenomeno che riguarda tutti gli attori del web, non soltanto musei, archivi e biblioteche. In passato, la disponibilità dei siti web degli editori ha reso più semplice anche il lavoro di catalogazione e il lavoro di controllo di autorità, grazie alla pubblicazione di molte informazioni. È evidente che la disponibilità di metadati interoperabili renderà ancora più semplice la connessione e la collaborazione tra il mondo dell'editoria e quello delle biblioteche, con vantaggi reciproci. Fermo restando che ciascun attore continuerà a produrre – e poi a condividere – i dati di proprio specifico interesse, la condizione essenziale per avere vantaggi reciproci è quella dell'interoperabilità dei dati. Un aspetto fondamentale che risulta dalla connessione tra editori e biblioteche è che i dati descrittivi diventeranno sempre più facilmente disponibili anche per le biblioteche, che potranno invece dedicarsi al nuovo lavoro del bibliotecario che si occupa dei metadati, ovvero la gestione delle identità (*identity management*). Questo consiste proprio nella costruzione della rete di corrispondenze, di rinvii, di identità, che caratterizza il web semantico, proprio in quanto associa a una stringa un significato preciso, definito con il collegamento di una particolare entità con i dati su quella medesima entità espressi in uno o più insiemi di dati.

#### *Rossana Morriello*

Credo sia un processo indispensabile in questa fase, e uno stimolo sia per le biblioteche che per gli editori. L'esigenza di standardizzare i metadati tra biblioteche e

editori si è già manifestata in passato con formati di metadati come ONIX, ma limitatamente a certi aspetti delle transazioni. Oggi proprio la metadattazione consentirebbe di estendere la collaborazione su molti fronti con reciproco vantaggio, e a beneficio dell'utente finale. Le biblioteche e gli editori usano spesso informazioni di tipo diverso, per esempio per classificare i libri, ma molte informazioni sono già comuni, come il titolo di un libro, l'autore, l'ISBN, l'anno di pubblicazione, la descrizione fisica, il prezzo. Estendere la collaborazione a una più ampia serie di metadati costruiti e modellati in collaborazione, per le diverse tipologie documentarie, strutturati e collegati tra loro, consentirebbe di amplificare la presenza e l'impatto dell'insieme degli agenti coinvolti nella filiera del libro. Ciascuno degli agenti potrebbe poi usare i metadati d'interesse per le proprie specifiche applicazioni, laddove vi siano esigenze diverse.

#### *Valdo Pasqui*

Il bibliotecario in futuro si dovrebbe dedicare di più all'arricchimento semantico del catalogo, all'interconnessione tra le varie risorse per mezzo dei metadati e al rapporto collaborativo con gli utenti invece che alla catalogazione di base o primaria che può essere recuperata alla fonte.

#### *Tiziana Possemato*

La catena produttiva dell'informazione è diventata ben più complessa e articolata rispetto al passato, includendo una molteplicità di attori e di fonti di dati eterogenee e non necessariamente autorevoli: dalle biblioteche agli editori, dai siti web alle grandi comunità collaborative, come Wikidata. Il che, di per sé, non garantisce una maggiore capacità di collaborazione e di riutilizzo del dato. Certo è che la tendenza a utilizzare linguaggi comprensibili e condivisibili anche al di fuori delle biblioteche, cosa che il formato MARC non ha mai davvero favorito, creando in molti casi delle vere barriere comunicative, dovrebbe facilitare questo tipo di connessione e di cooperazione. La tendenza si era già manifestata nel più recente passato con l'adozione da parte delle biblioteche di tool di conversione dei dati, per esempio dall'ONIX – utilizzato da alcuni editori – al MARC. Ma le conversioni tra formati di dati diversi sono strategie di riutilizzo del dato costose, in termini di energie, e che richiedono continui adattamenti e revisioni. E d'altra parte, l'adozione di modelli di dati agnostici rispetto ai domini, come l'RDF indicato per la strutturazione dei linked data, che pure indica la strada di un superamento importante delle specifiche competenze di dominio con i relativi linguaggi, non è ancora così largamente diffusa se non in contesti per ora altamente specializzati. In questo momento storico bisogna guardare alla

*tendenza* alla cooperazione che spinge a trovare modi di dialogo con istituzioni diverse, più che all'idea di un obiettivo raggiunto. La pratica di dataset open, che fino a qualche anno fa sembrava inaccettabile non solo a enti che sui dati costruiscono il proprio mercato ma addirittura alle biblioteche, sta diventando una modalità corrente di condivisione e creando nuove dinamiche di relazione basate sul riutilizzo delle informazioni.

*Roberto Raieli*

Partendo dall'esempio limitato della collaborazione tra editori e biblioteche all'interno dei nuovi strumenti di ricerca – per esempio i discovery tool – la questione è già complessa da dirimere. Ci si aspetta, intanto, che i diversi editori siano sempre più propensi a fornire ai produttori dei discovery tool la licenza per indicizzare i contenuti che essi pubblicano, da includere nell'indice unico insieme ai dati dei contenuti locali della biblioteca, recuperati tramite *harvesting* dei metadati dei repository istituzionali e dei record dei cataloghi. Nella pratica la collaborazione con le biblioteche non è diretta nemmeno su questo fronte, dato che l'elenco dei database in cui un sistema effettua la ricerca è sempre compilato di *default* dai produttori in base alle licenze editoriali che gli conviene acquisire, e in base alle “alleanze” con gli editori stessi. In ogni caso, la selezione operata dai produttori dei sistemi di ricerca è sempre estremamente limitata – nel bene e nel male – rispetto all'ambito illimitato in cui cercano i motori generalisti. I principali editori di database, di ebook e di ejournal, comunque, si stanno preoccupando di favorire l'interoperabilità, rendendo i propri metadati aperti a tutti i discovery tool, consentendo a sistemi di ogni “marca” di utilizzarli per le ricerche complessive. La raccolta dei metadati da parte dei produttori dei sistemi, però, è soggetta a differenti tipi di accordi sottoscritti con diversi editori e aggregatori, per cui i dati sono prodotti in forma differente, con un grado di accuratezza molto variabile e, per giunta, non tutti i *content providers* sono raggiunti e collaborano all'indice. In ogni caso non si possono lamentare mancanze relative ai contenuti prodotti dai principali editori “occidentali”. I primi a tenere all'inclusione nell'indice dei diversi strumenti di ricerca sono gli editori stessi, che vedono incrementare il regime delle vendite in proporzione alla diffusione dell'informazione sulla propria produzione. Più difficile è la copertura della produzione non in lingua inglese, o di paesi poco sviluppati, e di tutti quegli editori che non sono ben noti nella scena internazionale.

*Maurizio Vivarelli*

La relazione tra mondi dell'editoria e delle biblioteche può essere certamente migliorata. Sarebbe certamente

auspicabile, ad esempio, un maggior allineamento dei modelli di descrizione e indicizzazione dei libri, fisici e digitali, utilizzati all'interno di questi due mondi, che garantirebbe una valutazione comparativa dei dataset molto utile per migliorare, sul versante delle biblioteche, la qualità delle procedure di gestione e sviluppo delle collezioni; e, sul versante editoriale, la comunicazione dei contenuti alle diverse comunità di *stakeholders*. In questo caso specifico si profilerebbero, dunque, concreti benefici in primo luogo per le comunità tecnico-professionali, sui due versanti editoriale e bibliotecario; e in secondo luogo per quelle dei lettori, che potrebbero disporre tendenzialmente di collezioni di migliore qualità.

***Secondo voi quali sono i progetti in corso in Italia e in varie parti del mondo per sviluppare una tecnologia sempre più funzionale a descrivere le risorse e ad aiutare gli utenti a trovare, identificare, selezionare e ottenere le informazioni che desiderano?***

*Stefano Bargioni e Alberto Gambardella*

COBIS, Alfabetica, Museo Galileo di Firenze, Share Catalogue sono alcuni esempi italiani dove l'attenzione alla metadattazione e il riuso di dati di Wikidata e altre fonti sta innovando la loro offerta. Attualmente in Italia ci sono 22 biblioteche o sistemi che hanno una proprietà in Wikidata, da SBN fino a piccole biblioteche accademiche. Anche altrove si assiste a realizzazioni basate su BIBFRAME, o basate sul modello IFLA LRM come AlKindi Ideo Library (<https://alkindi.ideo-cairo.org>) specializzata in studi islamici. Un ambito simile ad Alfabetica e dove i linked open data esprimono tutto il loro potenziale è certamente <https://data.bnf.fr>. In Data BNF, la strutturazione delle risorse ottenuta tramite la metadattazione ne permette la navigabilità, offrendo



all'utente informazioni ricche, puntuali e amichevoli. ICCD ha realizzato il Catalogo generale dei beni culturali (<http://www.iccd.beniculturali.it>) con il dichiarato intento di promuoverne l'apertura e al tempo stesso il riuso e l'interoperabilità. Questi obiettivi possono e debbono essere perseguiti anche dalle biblioteche, proprio tramite la metadattazione.

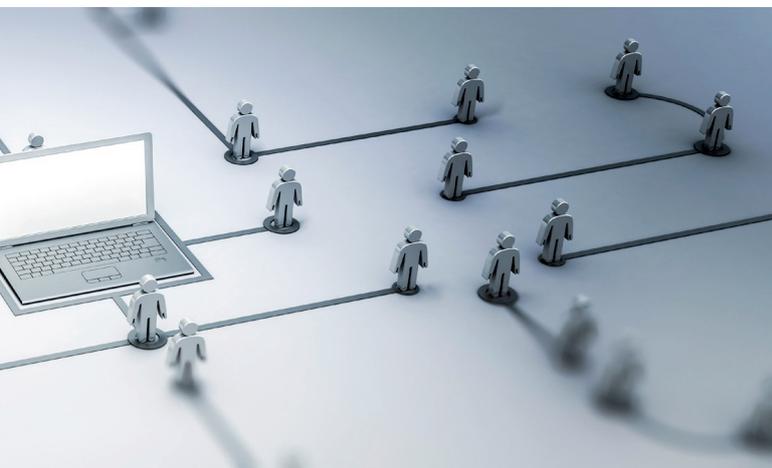
#### *Giovanni Bergamin*

Vorrei citare tre iniziative significative nel campo della descrizione delle risorse bibliografiche. Le prime due sono state presentate al Convegno internazionale Il controllo bibliografico nell'ecosistema digitale (Firenze, 8-12 febbraio 2021) e riguardano casi d'uso – a regime, non a titolo sperimentale – delle tecnologie dell'apprendimento automatico (o machine learning) per il controllo bibliografico. La prima iniziativa è stata presentata da Elisabeth Mödden della Deutsche Bibliothek che ha parlato del trattamento delle pubblicazioni che nascono digitali e che sono oggetto di deposito legale: ebook, articoli di riviste accademiche e tesi di dottorato. Le tecnologie dell'intelligenza artificiale vengono in questo caso usate per la produzione della Bibliografia nazionale tedesca e in particolare per la classificazione e la soggettazione. La seconda iniziativa è stata presentata da Osmo Suominen e riguarda l'esperienza della Biblioteca nazionale finlandese. L'obiettivo anche in questo caso è quello di classificare e soggettare pubblicazioni che nascono digitali e che sono oggetto di deposito legale. Quello che caratterizza il progetto finlandese (ANNIF) è – in estrema sintesi – il fatto che le pubblicazioni da indicizzare automaticamente vengono trattate in successione con differenti applicazioni open source esistenti nel campo del machine learning e che il risultato viene messo a confronto in fase di controllo di qualità. Occorre sottolineare che in queste due esperienze le tecnologie non sostituiscono la professio-

nalità del bibliotecario. Nel resoconto dei due progetti l'intelligenza artificiale può dare risultati soddisfacenti solo se accompagnata da bibliotecari esperti in soggettazione e classificazione che in questo contesto hanno sostanzialmente due compiti: 1) classificare e soggettare “manualmente” alcuni insiemi per “insegnare” alla macchina; 2) lavorare al controllo di qualità ed essere coinvolti nello sviluppo di nuove applicazioni. La terza iniziativa riguarda invece il consolidarsi dell'uso di uno strumento come Wikibase per la gestione dei dati bibliografici. Segnalo in particolare la BnF con NOEMI, ovvero il nuovo ambiente di metadattazione conforme a IFLA LRM, e – ancora – la Deutsche Bibliothek con il progetto legato al Wiki Library Manifesto di dare una “seconda casa” su Wikibase al suo sistema integrato di gestione del controllo di autorità (GND). È da notare che mentre NOEMI è un ambiente di metadattazione riservato ai bibliotecari della BnF, il progetto per il GND è invece finalizzato anche a prospettive di raccolta di contributi esterni (*crowdsourcing*).

#### *Roberto Delle Donne*

Mi limito a riferire del riutilizzo di metadati bibliografici in linked open data (LOD) nell'ambito delle edizioni di fonti medievali, realizzate nel quadro della collaborazione tra l'Università Federico II di Napoli, l'Università della Basilicata, Scaipuntoit, @Cult, Wikidata e Factgrid. Tra queste segnalo l'edizione di un registro di età aragonese reperibile all'indirizzo [www.detail.unina.it](http://www.detail.unina.it). Significativo è stato il confronto con Tiziana Possemato, avviato nel 2013 per la realizzazione del catalogo interuniversitario SHARE-CAT, un'iniziativa ancora oggi fortemente innovativa, non solo nel panorama europeo. Per chiarire il contesto più generale in cui il riutilizzo di metadati bibliografici in LOD è avvenuto, ricordo che tra gli storici è da tempo presente la consapevolezza che le reti telematiche hanno trasformato profondamente le metodologie e le pratiche della ricerca storica, le forme di comunicazione e le modalità di diffusione dei suoi risultati, provocando considerevoli mutamenti anche nei processi di ricerca, selezione, studio, interpretazione e critica delle fonti, quali si erano andati consolidando nel corso degli ultimi due secoli. Agli storici è del resto noto che la trasposizione in formato digitale di una fonte analogica o l'esame di fonti native digitali richiedano peculiari strumenti ecdotici che tengano conto dell'eventuale perdita dei contenuti informativi veicolati dall'originario supporto analogico oppure del loro possibile arricchimento determinato dal loro inserimento in una rete di relazioni ipertestuali con altre fonti, con trascrizioni, edizioni critiche e altri strumenti informativi che, inevitabilmente, ne condizionano la comprensione e l'interpretazione. Nell'ultimo decen-



nio, con il moltiplicarsi dei produttori di contenuti in grado di inserire facilmente in internet informazione strutturata, è infatti cresciuta l'esigenza di descrivere e organizzare tale informazione, per garantirne la reperibilità ed evitare che si perda nel rumore di fondo della rete. In tale prospettiva sono stati determinanti gli orientamenti del web semantico o *web di dati*, secondo cui il lavoro di organizzazione e di gestione dell'informazione deve essere in gran parte automatico e basato su descrizioni fortemente standardizzate e formalizzate, elaborate da specialisti, come bibliotecari e archivisti. I sistemi di classificazione dell'informazione alla base del web semantico sono ontologie formali, schemi di ordinamento dei documenti gerarchico-enumerativi oppure analitico-sintetici elaborati da esperti del settore, espressi in maniera uniforme e rigorosa e associati all'informazione primaria attraverso l'uso di linguaggi e formalismi, a loro volta rigidamente strutturali e ben definiti, comprensibili dalle macchine, offrendo nuove potenzialità alla indicizzazione di dati rilevanti per gli studi di tipo storico-filologico, oltre che per la razionalizzazione delle informazioni di tipo biblioteconomico. In questo contesto, al web dei documenti basato sul linguaggio HTML si aggiunge un insieme strutturato di informazioni collegate tra loro: i cosiddetti linked data, i dati connessi, basati sulla sintassi Resource Description Framework (RDF), lo strumento base proposto dal W3Consortium per la codifica, lo scambio, il riutilizzo e l'interoperabilità semantica. I LOD costituiscono quindi una tecnologia e un insieme di buone pratiche per pubblicare dati sul web in una modalità leggibile, interpretabile e omogeneizzabile da agenti non umani, come ad esempio i motori di ricerca. Dal 2015 sono stati avviati diversi progetti di ricerca volti a delineare modelli e pratiche ecdotiche di documenti medievali basate sui LOD. Si tratta di attività indirizzate allo studio e alla realizzazione di ontologie legate al "dominio" storico, che mirano alla pubblicazione digitale di documenti storici secondo la codifica semantica dei testi in XML-TEI (Text Encoding Initiative) e secondo la sintassi RDF, caratterizzata dalla sequenza "soggetto-predicato-oggetto" (cosiddette triple). In questa prospettiva, oltre alle collaborazioni con FactGrid, con WikiData e con il consorzio Data for History, per collocare il progetto entro una comunità scientifica più ampia di storici interessati alle potenzialità di ontologie e triple per lo sviluppo di nuovi strumenti di ricerca, sono stati riutilizzati anche i metadati creati in LOD in contesti bibliografici. Del resto, proprio questo orientamento mi ha indotto anche a elaborare e a esporre in XML-RDF, in Zenodo, con licenza CC0 1.0 Universal, i metadati bibliografici di tutti i contributi depositati in Reti Medievali Open Archive (<https://zenodo.org/badge/DOI/10.5281/zenodo.2628263.svg>). Le

tecnologie sviluppate per descrivere le risorse bibliografiche nel corso della realizzazione del progetto interbibliotecario SHARE si sono quindi rivelate utili anche per l'identificazione e il riutilizzo delle informazioni in un diverso contesto.

#### *Pierluigi Feliciati*

Di progetti nazionali e internazionali su fonti dati strettamente bibliografiche diranno senza dubbio i colleghi più specialisti di me. In Italia costituisce un'eccellenza per la comunità culturale il gruppo Wikidata MAB, che a neanche due anni dal suo lancio ha raccolto intorno a sé una comunità variegata e motivata e sta attuando obiettivi di analisi, visualizzazione, scoperta, correzione e arricchimento dei dati, soprattutto bibliografici. Questa azione volontaristica, "dal basso", pur se svolta per iniziativa di esperti, si inquadra nel contesto partecipativo di cui si è detto, gettando peraltro ponti tra comunità disciplinari e favorendo la ricomposizione semantica dei dati relativi all'agire umano. Da segnalare sempre nel nostro paese è il progetto ArCO, Architettura della Conoscenza, sviluppato dall'Istituto centrale del catalogo del MiC insieme al CNR per costruire il knowledge graph del patrimonio culturale italiano, esprimendo in RDF, sulla base delle relative ontologie, sette vocabolari che descrivono il dominio dei beni culturali e i dati estratti dal Catalogo Generale dei Beni Culturali dell'ICCD-MiC. L'esposizione dei LOD può costituire un punto di partenza importante per ricomporre le basi conoscitive suddivise in ambiti istituzionali e disciplinari. Nel settore archivistico sono stati esposti in forma di LOD diverse basi di dati, per iniziativa dell'Istituto Centrale degli Archivi e dell'Archivio Centrale dello Stato. La vera criticità è che le ontologie adottate sono "locali", ovvero esprimono le proprietà dello specifico dominio conoscitivo in forma di base di dati, e sono, dunque, interoperabili molto limitatamente. Si è infatti in attesa di un'ontologia riconosciuta come standard dalla comunità archivistica internazionale, come la RiC-O dell'International Council of Archives, attualmente in versione di bozza 0.2 e sperimentata sinora soprattutto in Francia, presso gli Archives Nationales. Un progetto di respiro davvero internazionale è costituito da Interpares Trust AI, avviato nel 2021 per iniziativa della School of Information dell'Università della British Columbia (Vancouver, Canada), che lavorerà per 5 anni sulla sperimentazione delle soluzioni di intelligenza artificiale per rispondere efficacemente alle funzioni archivistiche, dalla creazione dei *record* all'accesso da parte degli utenti.

#### *Claudio Forziati*

Gli esempi sono tanti anche se ci limitiamo solo a quei progetti, di ampiezza molto variabile, che hanno adottato i linked data per la rappresentazione e la condivi-

sione delle informazioni; una varietà che a mio avviso testimonia grande vivacità nei temi e negli obiettivi. Tra le iniziative di ampio respiro è impossibile non segnalare Linked Data for Production (LD4P), un progetto finanziato dalla Andrew W. Mellon Foundation, che vede coinvolte sei istituzioni statunitensi e ha l'obiettivo di proporre un quadro di riferimento per la creazione e il riutilizzo dei metadati bibliografici, applicando una effettiva transizione da formati a base MARC ai LOD. Verso il consolidamento di questa prospettiva si muovono i cataloghi innovativi della SHARE Family, tra i quali SHARE VDE, Kubikat-LOD, Parsifal e SHARE Catalogue. Va inoltre citato CoBIS, frutto di uno sforzo consortile di biblioteche speciali piemontesi. Esistono poi iniziative molto specifiche che affrontano la questione del 'pregiudizio' negli strumenti di soggettazione generalisti. Un caso emblematico è Homosaurus, vocabolario di termini LGBTQ+ in linked data, che, in combinazione con vocabolari più ampi come LCSH, consente una efficace descrizione delle risorse sul tema, ampliandone la possibilità di reperimento presso le istituzioni culturali. Infine, vale la pena di segnalare strumenti tecnologicamente più tradizionali, nordamericani e australiani, dedicati alla sottorappresentazione delle culture e del patrimonio dei popoli nativi (tema al quale è dedicato il numero speciale di IFLA Journal dell'Ottobre 2021), tra i quali Māori Subject Headings Thesaurus (Nga Upoko Tukutuku) e Indigenous subject headings modification project (ISHM).

#### Rossana Morriello

La metadattazione è una fase imprescindibile per qualsiasi processo di gestione dei dati digitali. Le banche dati editoriali, e di ogni genere, usano la metadattazione per organizzare i contenuti. Gli OPAC e i portali digitali bibliotecari usano i metadati per permettere agli utenti di trovare le risorse. E dalla metadattazione partono i progetti più avanzati, per esempio per l'applicazione dell'intelligenza artificiale. In ambito bibliotecario ci sono molti progetti in corso nel mondo e qualcuno anche in Italia. Durante la *International Conference Bibliographic Control in the Digital Ecosystem*, organizzata dall'Università di Firenze con altri partner a febbraio 2021, ne sono stati presentati diversi. Un progetto molto interessante di applicazione dell'intelligenza artificiale alle collezioni delle biblioteche, *Reading (&) Machine*, è in corso a Torino ed è stato presentato da Maurizio Vivarelli e Marco Mellia al Convegno delle Stelline *La biblioteca piattaforma della conoscenza* nel 2021. La Library of Congress lavora su questi temi da molti anni, così come la British Library che, per esempio, collabora insieme ad altre istituzioni al progetto *Living with Machines*. Infine, un'applicazione molto avanzata di realtà

aumentata, che ci mostra le frontiere raggiungibili per le biblioteche, è quella prototipata dalla Mandal Public Library e dalla Science Library dell'Università di Oslo, e presentata in un video su YouTube ([https://www.youtube.com/watch?v=jSfdG\\_45iqA](https://www.youtube.com/watch?v=jSfdG_45iqA)), di cui consiglio vivamente la visione.

#### Valdo Pasqui

In Italia non sono a conoscenza di progetti che vadano in questa direzione e mi sembra che ormai, sia nell'ambito delle biblioteche aderenti a SBN che nelle altre, ci sia un notevole appiattimento sui servizi e le funzionalità offerti dai pochi grandi fornitori di ILS e discovery tool. Qualche sforzo nella direzione social è stato fatto anni fa da Sosebi. Un altro fornitore che ha proposto soluzioni innovative è @Cult. In Italia vi è ancora una cultura molto legata al MARC\* e allo sviluppo di aggregatori che operano interagendo con gli OPAC tradizionali e che nel migliore dei casi forniscono funzionalità di restituzione dei risultati ordinati per rilevanza, faccette, area personale per salvare query e risultati ma quasi mai la "FRBRizzazione" e la ricerca/navigazione LOD (anche il recente Alfabetica di ICCU rientra in questa accezione). All'estero è sicuramente rilevante il lavoro portato avanti negli ultimi anni dalla BnF.

#### Tiziana Possemato

Diversi sono i progetti che si pongono questo obiettivo e lavorano in questo senso. Primo fra tutti, citerei Wikidata, che ha aggregato una comunità cross-domain estesa in tutto il mondo, che si occupa di creare informazioni aperte e tool tecnologici collaborativi. Non posso non ricordare qui Share-VDE (Share Virtual Discovery Environment), un'iniziativa che riunisce, arricchisce e collega i cataloghi bibliografici e di authority di una ampia comunità di biblioteche in un ambiente condiviso basato su linked data. Share-VDE ha progressivamente ampliato la sua portata abbracciando una comunità di istituzioni autorevoli in diversi ambiti e costituendo la Share Family. L'iniziativa è promossa da Casalini Libri, agenzia bibliografica internazionale e membro del Programma di Catalogazione Cooperativa; @Cult, fornitore di ILS e soluzioni di web semantico, con input e partecipazione attiva di un gruppo internazionale di biblioteche, comprese biblioteche nazionali e di ricerca, e influenzata dalla visione del progetto LD4P. Dopo due fasi di ricerca e sviluppo avviate nel 2016 e con la collaborazione della Library of Congress, Share-VDE e la Share Family stanno ora passando a una fase di produzione collegando i cataloghi di biblioteche negli Stati Uniti, in Canada e in Europa, compreso il progetto italiano Share Catalogue che è stato davvero pioniere in questo senso, in Italia e fuori. L'iniziativa di collaborazione è aperta a

qualsiasi biblioteca ed è guidata dalla comunità bibliotecaria. Far parte della Share Family significa supportare e facilitare le istituzioni nel delicato passaggio da un ambiente di catalogazione tradizionale a modelli innovativi basati sulle nuove tecnologie del web, offrendo prima di tutto ai partecipanti uno spazio di confronto professionale e di condivisione di esperienze e di competenze in diverse aree della conoscenza, facilitando questa transizione altrimenti così complessa.

*Roberto Raieli*

Non sono molto aggiornato in proposito, ma a buon senso posso, intanto, citare i progetti storici. In Italia non si può non citare SHARE Catalogue, parte di un articolato progetto denominato SHARE (Scholarly Heritage and Access to Research), orientato alla cooperazione territoriale e alla condivisione dei servizi bibliotecari e documentari. In SHARE Catalogue sono convertiti in linked data, secondo il modello RDF, i dati dei cataloghi bibliografici dei sistemi bibliotecari in convenzione. L'utente potrà, poi, cercare agevolmente informazione attraverso un'unica interfaccia di navigazione, organizzata secondo il modello FRBR. All'estero, è storico anzitutto il progetto della Bibliothèque nationale de France, che applica gli standard del web semantico per la pubblicazione dei cataloghi della Biblioteca e i dati della biblioteca digitale Gallica. Segue il progetto della British Library, che pubblica la BNB (British National

Bibliography) in linked data. Ancora, da evidenziare, il linked data service della Deutsche Nationalbibliothek (DNB) che ha avviato la conversione e pubblicazione dei propri record d'autorità e dei record bibliografici, e il progetto della Biblioteca Nacional de España per la pubblicazione dei propri record in linked data. Da ricordare, poi, il Linked Data for Libraries Project, creato congiuntamente dalla Cornell University Library, dallo Harvard Library Innovation Lab e dalle Stanford University Libraries con il sostegno della Fondazione Andrew W. Mellon. Infine, un progetto storico che ha fatto molta scuola è Linked Jazz, che si basa sull'idea che le persone appassionate di jazz possano arricchire di contenuti l'archivio, contribuendo a esplicitare la natura delle relazioni che collegano le entità in esso descritte.

*Maurizio Vivarelli*

Un progetto molto interessante all'interno del quale sono presenti elementi significativi della tradizione documentaria, della data science, della linguistica computazionale è Living with machines (<https://www.turing.ac.uk/research/research-projects/living-machines>), promosso dall'Istituto Alan Turing e dalla British Library, che ha per oggetto lo studio della Rivoluzione industriale in Gran Bretagna tra fine del Settecento e inizio dell'Ottocento, che forse si situa sul confine esterno della metadattazione intesa nella sua specifica dimensione concettuale, metodologica, procedurale e tecnologica.

## ABSTRACT

*Metadating, in the field of librarianship, is the dimension assumed by cataloging in the digital age; like cataloging, it refers to the process of recording metadata, that is, data functional to the identification and discovery of a bibliographic resource. It has been a reality for at least fifteen years in the international library context, especially in Northern Europe, the United States, Canada and Australia. With metadata, all relevant data continues to be recorded in an open environment – preferably according to cataloging rules suitable for metadata. What changes is the presentation, i.e., the way the created data is returned to the reader. Italian librarians and professors of librarianship answer 10 questions, offering, for the first time in an Italian journal, an overview of the topic of great interest in the international librarianship scene.*