



Università
Ca' Foscari
Venezia

Corso di Laurea magistrale
(ordinamento ex D.M. 270/2004)
in Storia e gestione del patrimonio archivistico e bibliografico

Tesi di Laurea

Dati bibliografici e linked (open) data
Prospettive e criticità

Relatore

Ch. Prof. Paul Gabriele Weston

Correlatore

Ch. Prof. Riccardo Ridi

Laureanda

Camilla Fusetti

Matricola 845606

Anno Accademico

2015 / 2016

Indice

Indice delle immagini	1
Introduzione	2
I dati bibliografici e la loro evoluzione	6
1.1. Dalla Great Tradition ai Principi Internazionali di Catalogazione del 2016.....	9
1.2. I formati bibliografici e gli standard	24
1.3. Gli strumenti di accesso ai dati bibliografici.....	29
1.4. Il presente e il futuro	31
Linked (open) data	33
2.1. Linked data e web semantico	33
2.1.1. Dal web dei documenti al web dei dati.....	35
2.2. Uno sguardo alle tecnologie del web semantico	38
2.2.1. XML.....	38
2.2.2. RDF.....	40
2.2.3. URI.....	44
2.2.4. Ontologie e vocabolari.....	45
2.3. Dai linked data ai linked open data	51
I linked open data e l'universo bibliografico	53
3.1. Il Library Linked Data Incubator Group	55
3.2. Ontologie e vocabolari per rappresentare i dati bibliografici.....	61
3.2.1. La scelta delle ontologie e i profili di applicazione	64
3.3. Library linked open data cookbook.....	68
3.4. Alcune criticità	74
Casi di applicazione dei linked open data	78
4.1. Europea.....	78
4.2. Digital Public Library fo America	88
4.3. Online Computer Library Center	93
Conclusioni	100
Bibliografia	105
Sitografia	116
Ringraziamenti	124

Indice delle immagini

Figura 1: Anteprima di un oggetto digitale nel portale di Europeana	79
Figura 2: Classi principali di EDM	82
Figura 3: Art Faces	84
Figura 4: Visualizzazione di un record nel portale di Europeana.....	85
Figura 5: Linked open data in un record del portale di Europeana	86
Figura 6: Relazioni di similarità tra record di Europeana	87
Figura 7: DPLA Metadata Model.....	90
Figura 8: Color Browse	92
Figura 9: Metadata Explorer.....	93
Figura 10: Aggiornamenti del vocabolario VIAF RDF	95
Figura 11: WorldCat Entities.....	96
Figura 12: WorldCat Works	97
Figura 13: Multilinguismo per i record di WorldCat.	98
Figura 14: Markup di entità di WorldCat	98

Introduzione

Il presente lavoro ha come oggetto lo studio delle relazioni tra i dati bibliografici, creati e gestiti dalle biblioteche, e i linked open data (LOD), una nuova tecnologia in grado di originare dati aperti semanticamente interoperabili nel web.

Oggi sono molti i progetti di ambito bibliotecario che hanno sperimentato questo approccio. La scelta di adottare i linked open data sottende spesso varie motivazioni. In alcuni casi si vuole aumentare l'interoperabilità tra tipologie differenti di dati, in altri affermarsi come fonti autorevoli in un contesto più ampio, in altri ancora allargare il pubblico degli utenti e migliorare l'esperienza di accesso alle informazioni. Tra i progetti più importanti ricordiamo quelli della Deutsche Nationalbibliothek¹ e delle biblioteche nazionali di Svezia², Ungheria e Spagna³, il cui scopo è quello di rendere i dati catalografici di tipo locale o regionale visibili a livello mondiale, aprendo i database al più ampio contesto del web. La British Library ha applicato queste tecnologie alla British National Bibliography⁴, in risposta alle politiche di apertura dei dati promosse dal Governo inglese, ma anche per svincolarsi dall'uso di formati di ambito prettamente bibliotecario. La Bibliothèque Nationale de France⁵ ha reso le risorse accessibili in modo che gli utenti le possano citare nei siti, nei blog o nelle email; inoltre è stato predisposto *data.bnf.fr*, un portale che potenzia la funzione di scoperta grazie alla ricca rete di relazioni tra dati. Anche diversi strumenti semantici di comune utilizzo nella pratica catalografica sono attualmente disponibili come linked open data. Tra i principali vi sono LCSH (Library of Congress Subject Headings)⁶, VIAF (Virtual International Authority File)⁷, la DDC (Dewey Decimal Classification)⁸ e il Thesaurus del Nuovo Soggettario⁹.

¹ «Deutsche Nationalbibliothek - Linked Data Service», http://www.dnb.de/EN/Service/DigitaleDienste/LinkedData/linkedata_node.html.

² «LIBRIS», <http://libris.kb.se/>.

³ «datos.bne.es», <http://datos.bne.es/inicio.html>.

⁴ «Welcome to bnb.data.bl.uk | The British National Bibliography», <http://bnb.data.bl.uk/>.

⁵ «index (data.bnf.fr)», <http://data.bnf.fr/>.

⁶ «Home - LC Linked Data Service (Library of Congress)», <http://id.loc.gov/>.

⁷ «VIAF», <https://viaf.org/>.

⁸ «Dewey Web Services | OCLC Developer Network», <https://www.oclc.org/developer/develop/web-services/dewey-web-services.en.html>.

⁹ «Nuovo Soggettario - Thesaurus», <http://thes.bncf.firenze.sbn.it/ricerca.php>.

Oltre al Thesaurus, in Italia ci sono svariate sperimentazioni inerenti la conversione di dati in LOD. Lo Share Catalogue¹⁰ è un portale per navigare cataloghi bibliografici, che vede riunite le Università di Napoli, Lecce, Benevento, Potenza e Salerno. A breve sarà pubblicata anche la BNI (Bibliografia Nazionale Italiana)¹¹. Un gruppo di lavoro dell'ICCU (Istituto Centrale per il Catalogo Unico) sta portando avanti un'attività di ricerca per la conversione dei record di SBN in linked open data¹², principalmente per consentire lo scambio di informazioni tra istituti culturali di ambiti differenti. Parallelamente, un *team* ancora informale, che vede coinvolti bibliotecari della Biblioteca Nazionale Centrale di Firenze e collaboratori di Wikipedia, sta tentando di inserire queste stesse descrizioni in Wikibase, una piattaforma dell'enciclopedia online a contenuto aperto, in modo che i dati siano riutilizzabili a partire da un'unica base¹³. Altre istituzioni italiane di ambito culturale hanno iniziato a convertire i propri dati e a pubblicarli con queste nuove modalità. Ad esempio, l'Istituto per i beni artistici culturali e naturali di Ravenna¹⁴, la Fondazione Zeri¹⁵ e il Museo Galileo¹⁶.

Il dibattito catalografico oggi ruota attorno alla possibilità di utilizzare i linked open data per soddisfare le funzioni e gli obiettivi stabiliti dai *Principi Internazionali di Catalogazione*. Ci si chiede in sostanza se i linked open data garantiscano una maggiore interoperabilità e apertura, anche per ridurre il costo economico della catalogazione, e se possano migliorare l'esplorazione delle informazioni rese disponibili dalle biblioteche. Inoltre l'IFLA FRBR Review Group ha presentato un nuovo modello concettuale dell'universo bibliografico, FRBR-LRM (FRBR – Library Reference Model)¹⁷, che ha lo scopo di essere maggiormente compatibile con le tecniche del web semantico e con la

¹⁰ «Share Catalogue», <http://catalogo.share-cat.unina.it/sharecat/clusters>.

¹¹ Luca Bellingeri e Maria Chiara Giunti, «BNI Open and Collaborative: How and Why», *JLIS.it*, n. 1 (2017): 67-76, doi:10.4403/jlis.it-12167.

¹² «ICCU | Istituto Centrale per il Catalogo Unico - Gruppo di lavoro Linked Open Data SBN», http://www.iccu.sbn.it/opencms/opencms/it/main/attivita/gruppilav_commissioni/pagina_0005.html.

¹³ Cristian Bacchi. «Marc2Wikibase: ipotesi di caricamento dei dati bibliografici in Wikibase».

Presentazione tenuta a SBN: un'eredità per il futuro, Ravenna, 16 settembre 2016.

http://www.bibliotecheromagna.it/admin/PagPar.php?op=fg&id_pag_par=1683&fld=file.

¹⁴ «Linked open data — E-R Istituto per i beni artistici culturali e naturali», <http://ibc.regione.emilia-romagna.it/servizi-online/lod>.

¹⁵ «Zeri&LODE», <http://data.fondazionezeri.unibo.it/>.

¹⁶ «Banca dati cumulativa - Museo Galileo»,

<http://www.museogalileo.it/esplora/biblioteche/biblioteca/bancadaticumulativa.html>.

¹⁷ Pat Riva, «On the New Conceptual Model of the Bibliographic Universe: the FRBR Library Reference Model», *AIB studi*, n. 2 (2016): 265-75, doi:10.2426/aibstudi-11480.

strutturazione dei dati in LOD. Il modello è ancora in fase di revisione internazionale; per questo non è stato oggetto della trattazione. È comunque importante tenere presente la direzione verso cui sta andando la disciplina catalografica. Queste realtà e la riflessione che esse sollevano sono alla base del presente studio.

L'obiettivo della tesi è duplice: da una parte fornire un inquadramento teorico, sia dei dati bibliografici che dei linked open data, e dall'altra cercare di valutare l'applicazione pratica di queste tecnologie e i risultati che essa può avere.

In particolare, sono stati analizzati i principali documenti di rilievo internazionale inerenti questi settori. Per la catalogazione ci si è basati sulle varie versioni dei Principi Internazionali, sui modelli concettuali e sugli standard per i dati. Per i linked open data, invece, si è fatto riferimento alla documentazione emanata dal W3C (World Wide Web Consortium), l'ente che ha lo scopo di promuovere standard e buone pratiche per lo sviluppo del web. Riguardo la parte pratica, si è partiti dalle linee guida fornite dal Library Linked Data Incubator Group, e dall'analisi dei portali dei progetti scelti come casi di studio.

Il primo capitolo cerca di analizzare l'evoluzione dei dati bibliografici e della disciplina che ne stabilisce obiettivi e funzioni, ovvero la catalogazione. Questo per comprendere quali siano gli ultimi sviluppi e quali i principali limiti delle strutture catalografiche attuali. Nuovi standard come BIBFRAME (Bibliographic Framework Initiative) e RDA (Resource Description and Access) sono stati pensati proprio per l'ambiente web e in particolare per essere compatibili con le tecnologie dei linked open data.

Viene poi affrontato l'ambito del web semantico perché i LOD nascono come soluzione pratica alla sua realizzazione. L'idea di poter estendere il web dei documenti verso il web dei dati comporta l'utilizzo di diverse tecnologie. In questo capitolo se ne prendono in considerazione le principali, come XML (eXtensible Markup Language), URI (Uniforme Resource Identifier) e RDF (Resource Description Framework), senza pretesa di esaustività, poiché sono linguaggi e standard in continua evoluzione. Vengono inoltre analizzati i Principi dei linked data proposti da Tim Berners-Lee e il passaggio che ha portato all'idea di apertura dei dati, e quindi ai linked *open* data.

Gli ultimi due capitoli cercano di valutare le applicazioni pratiche di queste tecniche ai dati bibliografici. Quali sono i vantaggi dell'utilizzo dei linked open data? A questa domanda si è tentato di rispondere analizzando le raccomandazioni presenti nel *Report* del Library Linked Data Incubator Group, il primo gruppo ufficiale di lavoro che, con la partecipazione di biblioteconomi ed esperti di web semantico, ha analizzato queste tecnologie proprio in rapporto ai dati bibliografici creati dalle biblioteche. L'approccio dei linked open data prevede la possibilità di riutilizzare ontologie e vocabolari preesistenti, ma per questo bisogna capire quali siano i più adatti. Si è cercato quindi di individuare le strategie di scelta per una istituzione e le modalità di predisposizione dei profili di applicazione. Questi ultimi sono i documenti in cui si dichiara quali schemi di dati o modelli sono stati prescelti per rappresentare il proprio dominio conoscitivo.

La conversione dei dati e la loro pubblicazione segue di solito un processo ben definito. Nel paragrafo intitolato "Library linked open data cookbook" vengono presentate le fasi principali sulla base delle raccomandazioni del W3C. La presa in considerazione dei passaggi concreti che consentono di realizzare linked open data mette in luce alcune criticità, che si è cercato di valutare.

Infine sono stati analizzati tre progetti di ambito culturale che hanno deciso di sperimentare queste tecnologie, in relazione ai dati bibliografici: Europeana, la Digital Public Library of America e alcuni progetti di OCLC (Online Computer Library Center). Per ognuno di questi sono stati esaminati il modello di dati, i profili di applicazione predisposti, e la scelta delle licenze per i metadati. È stata considerata anche l'interfaccia utente, in relazione alle funzioni consentite dai dati. Questo per capire come i linked open data possano valorizzare le informazioni bibliografiche migliorando l'esperienza degli utenti.

Nelle conclusioni sono stati esposti i risultati di questo lavoro. I linked open data possono valorizzare i dati bibliografici e soprattutto sono in linea con gli obiettivi stabiliti dai *Principi Internazionali di Catalogazione*. La loro applicazione è però ancora in fase sperimentale, molte problematiche emergono man mano che i progetti procedono. Inoltre la loro adozione comporta un investimento a lungo termine. Serve personale specializzato per seguire tutte le fasi di pubblicazione e l'utilizzo di strumenti e linguaggi nuovi da parte di chi gestisce i dati delle biblioteche.

CAPITOLO PRIMO

I dati bibliografici e la loro evoluzione

I dati bibliografici sono informazioni che consentono l'accesso alle risorse messe a disposizione dalle biblioteche. L'insieme dei dati porta ad avere delle descrizioni, strutturate sulla base di specifiche normative, che riproducono simbolicamente le risorse e le rendono ricercabili da parte degli utenti attraverso i cataloghi¹⁸. I dati bibliografici vengono ricondotti a due principali tipologie:

- descrittivi, rappresentano gli oggetti e ne permettono l'identificazione;
- di autorità, vengono ricavati dalla risorsa o dal suo contenuto e costituiscono i primari elementi¹⁹ di accesso alle informazioni.

Nel primo caso si parla di *descrizione bibliografica*, nel secondo di *accessi formali* (per autori o titoli) e di *accessi semantici* (per soggetti e classi)²⁰. La catalogazione è l'ambito entro cui collocare le funzioni, gli obiettivi e le modalità di strutturazione dei dati bibliografici. Quando si parla di catalogazione ci si riferisce però ad almeno tre elementi:

- un nucleo teorico, costituito dai principi, dai modelli concettuali e dalle norme;
- i dati stessi, intesi come prodotto di una pratica;
- i cataloghi, ovvero gli strumenti di mediazione informativa tra gli utenti e le raccolte bibliografiche.

Ebbene, i dati bibliografici si collocano al centro di questa triade: da una parte essi infatti vengono creati sulla base di elementi normativi e prescrittivi, dall'altra necessitano dei cataloghi come strumenti di interfaccia alle informazioni. Le norme e gli standard cercano di concretizzare le funzioni e gli obiettivi previsti dai principi di catalogazione e dai modelli, che forniscono indicazioni sull'organizzazione dei dati. Gli elementi normativi

¹⁸ Andrea Marchitelli, *Orientarsi tra le informazioni in biblioteca: cataloghi, banche dati, motori di ricerca* (Milano: Editrice Bibliografica, 2015), 10.

¹⁹ I dati di autorità sono detti tali poiché vengono sottoposti a controlli volti a definire le forme preferite e non preferite di un nome proprio o di un termine. In questo modo viene garantita la funzione di raggruppamento e di organizzazione all'interno di un catalogo.

²⁰ Marchitelli, *Orientarsi tra le informazioni in biblioteca*, 14.

sono importanti perché creano un linguaggio comune che facilita la consultazione e lo scambio di informazioni tra i cataloghi delle biblioteche. Il catalogo è invece l'elemento che lega intimamente i dati bibliografici da una parte alle tecnologie e dall'altra agli utenti.

Le innovazioni tecnologiche hanno delle ricadute sulle modalità di pubblicazione, diffusione e fruizione dei documenti, rendendo necessari nuovi strumenti informativi. Si è passati dalla tavoletta di argilla ai rotoli di papiro, dalla pergamena alla carta, per arrivare oggi all'infrastruttura digitale.

Analizzare l'evoluzione dei dati bibliografici significa dunque mettere in luce i cambiamenti che hanno caratterizzato i tre elementi costitutivi della catalogazione: la teoria catalografica, i dati, e il catalogo stesso. Il senso di questi elementi è dato dalla loro relazione e dalle ricadute che ognuna di queste entità ha sulle altre.

Il rinnovamento costante della catalogazione, che emerge fin dalla nascita della *Great Tradition*²¹, può essere ravvisato nel fatto che “la catalogazione considera la realtà nel suo divenire e come tale si rinnova costantemente in rapporto con la cultura, della quale la tecnologia costituisce parte integrante”²². La teoria e la pratica catalografica vivono nella costante tensione che si crea tra la necessità di adeguamento ai nuovi bisogni degli utenti e la stabilità delle norme che serve a garantire la coerenza del catalogo²³. Questi strumenti sono strutture volte alla conservazione, alla diffusione e allo scambio dei dati bibliografici; la coerenza, intesa come omogeneità interna del catalogo, è un elemento fondamentale in quanto garantisce uniformità nella presentazione e nella fruizione dei dati e garantisce la possibilità di condivisione e riutilizzo. Allo stesso tempo è inevitabile che i cataloghi, per essere attuali, debbano “recepire i mutamenti della forma assunta dai documenti, del linguaggio, della cultura”²⁴.

L'avvento del web ha riportato la catalogazione al centro della riflessione biblioteconomica²⁵. Il dibattito ruota attorno al ruolo del catalogo nello scenario reso

²¹ Michael Gorman, «Changes in Cataloguing Codes: Rules for Entry and Heading», *Library Trends* 25, n. 3 (1977): 587–601.

²² Carlo Revelli e Giulia Visintin, *Il catalogo* (Milano: Editrice Bibliografica, 2008), 484.

²³ *Ibid.*, 485.

²⁴ *Ibid.*, 54.

²⁵ Paul Gabriele Weston, «La gestione elettronica delle biblioteche», in *Biblioteconomia: principi e questioni*, a c. di Giovanni Solimine e Paul Gabriele Weston, (Roma: Carocci, 2007), 235-6.

disponibile da internet e alle caratteristiche che devono avere i dati bibliografici per essere efficacemente utilizzati dagli utenti. Il web ha sostanzialmente cambiato tutto, come afferma Karen Coyle durante un intervento tenuto al convegno della Library and Information Technology Association²⁶. Viviamo in un mondo altamente computerizzato e interattivo, moltissime attività oggi vengono svolte nella rete e l'aspettativa delle persone è di poter intervenire, interagire ed esprimere la propria opinione nel web²⁷. Questa infrastruttura ha creato una realtà pluralistica, dove blogger indipendenti²⁸ sono diventati commentatori seguiti da molte persone, e organizzazioni come WikiLeaks²⁹ hanno influenzato il corso della politica e della diplomazia internazionale. Molte risorse sono già digitali o stanno per essere digitalizzate, il che non comporta ovviamente la perdita di interesse verso i documenti cartacei. Il paragone fatto da Karen Coyle con il teatro rende bene la situazione: 150 anni fa le persone che volevano vedere uno spettacolo potevano vederlo solo dal vivo, recandosi fisicamente nei luoghi di rappresentazione. Oggi il teatro non è scomparso e chi apprezza la peculiarità delle prestazioni *live* può ancora vivere questo tipo di esperienza, che però non è più l'unica³⁰. Lo si può ascoltare da podcast, lo si può guardare dalla televisione o tramite video in streaming o si possono avere esperienze di realtà virtuale. Inoltre la maggior parte delle risorse oggi è facilmente disponibile e recuperabile; tramite i motori di ricerca o i database si arriva all'oggetto ricercato e se questo è online si accede direttamente al contenuto³¹. Ciò non toglie che servano gli strumenti per usufruire di questo tipo di risorse e che questi strumenti non siano necessariamente facili da usare per tutti. L'accesso al web e ai documenti in esso contenuti richiede ad esempio un dispositivo e un software, che comunque devono essere aggiornati e non è detto funzionino sempre bene.

Le modalità comunicative e le esigenze degli utenti sono dunque completamente cambiate. La comunicazione è sempre più interattiva, remota e veloce. Le persone

²⁶ Karen Coyle «On the Web, of the Web», 2011, in <http://www.kcoyle.net/presentations/lita2011.html>.

²⁷ Ibid.

²⁸ Come esempio, molto vicino anche al mondo delle biblioteche e al loro ruolo nel contesto del web, si ricorda Aaron Swartz, per un avere un'introduzione sul suo percorso e il suo ruolo di attivista si può fare riferimento al documentario diretto da Brian Knappenberger, *The Internet's Own Boy: The Story of Aaron Swartz*, video YouTube, 1:44:59, caricato da Documentary Storm, 30 giugno 2014, <https://www.youtube.com/watch?v=M85UvH0TRPc>.

²⁹ «WikiLeaks», <https://wikileaks.org/>.

³⁰ Coyle, «On the Web, of the Web».

³¹ Ibid.

possono interagire in ogni momento le une con le altre o con le informazioni stesse, aggiungere o modificare i contenuti del web, tramite le chat, i tweet, i forum e altri ambienti sociali del web, creando una quantità enorme di informazioni “informali”³². Il concetto stesso di *accesso* viene inteso dagli utenti come possibilità di avere direttamente la copia dell’oggetto, per poter quindi usufruire del contenuto³³.

Il nostro mondo oggi è digitale e interattivo e i documenti sono facilmente manipolabili e granulari³⁴. La comunicazione avviene per lo più a distanza e in modo veloce, supportata spesso dall’utilizzo di media come i video e le immagini, a discapito del testo. Come si sta muovendo il sistema catalografico a fronte di tutto ciò?

Le risposte possono essere individuate a partire da una panoramica sul contesto catalografico. L’analisi prende avvio dai *Principi Internazionali di Catalogazione*, che individuano le funzioni e gli obiettivi della normativa catalografica e dei cataloghi. A questi si affiancano i modelli concettuali, che offrono una modalità di organizzazione dell’universo bibliografico.

1.1. Dalla Great Tradition ai Principi Internazionali di Catalogazione del 2016

Nel 1961 si tiene a Parigi un convegno internazionale, l’ICCP (*International Conference on Cataloguing Principles*), che vede riuniti i principali esperti di catalogazione e che porta alla pubblicazione dello *Statement of Principles Adopted by The International Conference on Cataloguing Principles*³⁵, informalmente conosciuti come i Principi di

³² Karen Coyle, «Linked Data: An Evolution», *JLIS.it* 4, n. 1 (2013): 53, doi.org/10.4403/jlis.it-5443.

³³ Lynn Silipigni Connaway e Timothy J. Dickey, «The Digital Information Seeker: Report of Findings from Selected OCLC, RIN and JISC User Behaviour Projects», in *The Library in the Life of the User: Engaging with People Where They Live and Learn*, a c. di Lynn Silipigni Connaway (Dublin: OCLC Research, 2015), 102–9; Karen Calhoun et al., *Online Catalogs: What Users and Librarians Want: An OCLC Report* (Dublin: OCLC, 2009).

³⁴ Il concetto di granularità può essere spiegato come la possibilità per i documenti digitali di articolarsi in parti sempre più piccole e autonome o anche come la scomposizione di un’unità in elementi più piccoli e indipendenti. Riccardo Ridi, *La biblioteca come ipertesto: verso l’integrazione dei servizi e dei documenti*, (Milano: Editrice Bibliografica, 2007), 28-9.

³⁵ IFLA Cataloguing Section e IFLA Meetings of Experts on an International Cataloguing Code, «Statement of International Cataloguing Principles (ICP)», (2016), http://www.ifla.org/files/assets/cataloguing/icp/icp_2016-en.pdf.

Parigi. L'incontro diventa una pietra miliare: per la prima volta il confronto relativo alla catalogazione avviene in termini di cooperazione internazionale.

I Principi di Parigi consolidano la riflessione della *Great Tradition*, ovvero la catalogazione di ambito angloamericano, la cui nascita viene convenzionalmente fatta coincidere con la pubblicazione delle *91 Regole* di Antonio Panizzi, nel 1841³⁶.

È importante tracciare a grandi linee i principali passaggi che hanno portato ai Principi di Parigi, partendo proprio da Panizzi.

Il catalogo a stampa, redatto sotto la sua guida, presenta un'introduzione metodologica nella quale si prescrive l'utilizzo del frontespizio come fonte della pubblicazione da cui trarre le informazioni, in modo da garantire uniformità alla catalogazione e di conseguenza facilitare agli utenti la comprensione delle informazioni. Inoltre la costruzione del catalogo, in questo caso a stampa, viene portato avanti da un gruppo di lavoro e non più da una sola persona³⁷. Si comincia dunque a riflettere sull'opportunità di uniformare la presentazione dei dati, e sui benefici agli utenti, in quanto fruitori finali delle informazioni veicolate dai cataloghi.

È con Charles Coffin Jewett che si presenta invece l'idea della standardizzazione, ovvero la necessità di utilizzare una veste grafica standard che prescriva l'ordine delle informazioni all'interno di una scheda catalografica, in modo da poter scambiare anche con altre biblioteche le notizie catalografiche³⁸. L'apporto di Jewett è "il punto di partenza di tutti i progetti di catalogazione partecipata e derivata, che hanno [...] reso particolarmente efficiente il sistema bibliotecario nordamericano"³⁹.

Qualche decennio dopo, nel 1876, viene pubblicata l'opera *Rules for a Dictionary Catalog* di Charles Ammi Cutter, un'opera ancora oggi molto importante, sui cui concetti principali vale la pena soffermarsi. Il principio più significativo viene enunciato nella premessa alla quarta edizione, in cui si legge "the convenience of the public is always to

³⁶ Mauro Guerrini e Carlo Bianchini, «Catalogazione», in *Biblioteconomia: guida classificata*, (Milano: Editrice Bibliografica, 2007), 372–90.

³⁷ Renzo Esposito, «Regole per la compilazione del catalogo» (Atti del Convegno di studi su Antonio Panizzi: Roma, 21-22 aprile 1980, Galatina: Editrice salentina, 1980), 225–41.

³⁸ Paul Gabriele Weston, *Il catalogo elettronico: dalla biblioteca cartacea alla biblioteca digitale*, (Roma: Carocci, 2002).

³⁹ *Ibid.*, 76.

be set before the ease of the cataloger”⁴⁰. La riflessione sulla possibilità di avere un impianto teorico complessivo che guidi lo sviluppo di norme, deve quindi essere sempre indirizzato al soddisfacimento dell’interesse degli utenti, piuttosto che a quello dei catalogatori. Con Cutter nasce l’idea dell’importanza di avere dei criteri fondanti applicabili a contesti differenti. Inoltre vengono poste le basi per comprendere la differenza tra i principi, intesi come valori “alti”, e le norme, ovvero prescrizioni pratiche sulla creazione delle notizie bibliografiche⁴¹. Le norme devono essere guidate da solide basi concettuali, che Cutter espone negli obiettivi del catalogo⁴²:

- consentire a una persona di trovare un libro di cui si conosca:
 - a) l’autore;
 - b) il titolo;
 - c) il soggetto;
- mostrare che cosa la biblioteca possiede:
 - a) di un determinato autore;
 - b) su un determinato soggetto;
 - c) in un determinato genere letterario
- facilitare la scelta di un libro:
 - a) attraverso la sua edizione (in termini bibliografici)
 - b) attraverso la sua caratterizzazione (in senso letterario o topico)

Vengono dunque chiaramente enunciati gli scopi a cui deve tendere la costruzione di un catalogo. La riflessione di Cutter è importante perché le funzioni di molti cataloghi ancora oggi non riescono a soddisfare gli obiettivi da lui fissati, come ad esempio la possibilità di cercare le notizie a partire dal genere letterario. La situazione attuale non ha realizzato i propositi di Cutter poiché invece di avere una solida base teorica e poche norme prescrittive si ha a che fare con codici sempre più ampliati in termini di contenuti e prescrizioni⁴³. Dopo un periodo di fervore che vede coinvolta nella prima metà del

⁴⁰ Charles Ammi Cutter, *Rules for a Dictionary Catalog* (Washington: Government Printing Office, 1904), 6, <https://archive.org/stream/rulesforadictio06cuttgoog#page/n10/mode/1up>.

⁴¹ Elaine Svenonius, Council on Library Resources, e Conference on the Conceptual Foundations of Descriptive Cataloging, a c. di, *The Conceptual Foundations of Descriptive Cataloging* (San Diego: Academic Press, 1989).

⁴² Cutter, *Rules for a Dictionary Catalog*.

⁴³ Le Regole Italiane di Catalogazione per Autori (RICA) pubblicate nel 1979, erano contenute in un volume di 290 pagine, le Regole Italiane di Catalogazione (REICAT) del 2009 sono passate a 632 pagine.

Novecento la catalogazione americana⁴⁴, si arriva a Seymour Lubetzky. Nel 1953, Lubetzky scrive *Cataloguing Rules and Principles*⁴⁵, portando “il rasoio di Occam ad un effetto devastante. La sua parafrasi [...] fu “is this rule necessary?”. Si tratta di una domanda apparentemente semplice che però spazzò via le leggi del 1949”⁴⁶. La volontà è quella di eliminare le contraddizioni tra le norme presenti nell’*ALA Catalogue Rules*⁴⁷, che si erano accumulate negli anni senza coerenza interna. La pubblicazione del 1960 del *Code of Cataloguing Rules* porta Lubetzky a stabilire due obiettivi principali al catalogo:

- permettere la localizzazione di una particolare edizione di un autore;
- creare delle relazioni e raggruppare le edizioni di una determinata opera o le opere di uno specifico autore.

Rispetto a Cutter va segnalato un cambiamento destinato a segnare tutto lo sviluppo della teoria catalografica, ovvero la sostituzione della parola “libro” con “edizione” ed “opera”. Il termine “opera” e la possibilità, ad esempio, di raggruppare sotto il titolo dell’opera le varie traduzioni, erano già stati proposti da Panizzi nelle sue *91 Regole*⁴⁸. Lubetzky tuttavia tratta in modo più consapevole il concetto di entità letteraria, che viene distinta dalle sue possibili e diverse materializzazioni fisiche, rappresentate dalle varie edizioni. Il pensiero di Lubetzky è alla base dei Principi di Parigi.

La conferenza del 1961 ha come obiettivo quello di “facilitare la circolazione internazionale della conoscenza realizzando la più ampia uniformità nei cataloghi delle biblioteche”⁴⁹, poiché le differenti modalità di trattazione dei dati bibliografici ostacolano lo scambio delle notizie.

Resource Description and Access (RDA), nell’edizione rivista del 2015 e nella versione a stampa presenta 1052 pagine.

⁴⁴ I cui principali attori sono la Library of Congress e l’American Library Association (ALA) che pongono al centro del dibattito le regole di tipo descrittivo e le regole per la creazione dei punti di accesso. Guerrini e Bianchini, «Catalogazione».

⁴⁵ Michael Gorman, «Seymour Lubetzky uomo di principi: discorso tenuto al Seymour Lubetzky 100. Birthday Symposium, 18 aprile 1998, University of California, Los Angeles», *Biblioteche oggi* 18, n. 6 (2000): 6–11.

⁴⁶ Ibid.: 8.

⁴⁷ Guerrini e Bianchini, «Catalogazione».

⁴⁸ Antonio Panizzi, *Catalogue of Printed Books in the British Museum: Volume I*. (London: Order of the Trustees, 1841), <https://catalog.hathitrust.org/Record/001761602>.

⁴⁹ Weston, *Il catalogo elettronico*, 84.

Nei Principi di Parigi viene stabilito che “un catalogo deve essere uno strumento efficace per accertare se”⁵⁰:

- la biblioteca contiene un libro particolare identificato:
 - a) per mezzo del suo autore e del titolo, oppure,
 - b) se l’autore non è nominato sul libro, per mezzo del titolo soltanto, oppure,
 - c) se autore e titolo sono inadatti o insufficienti all’identificazione, un conveniente sostituto del titolo;
- quali opere di un particolare autore e
 - a) quali edizioni di una particolare opera esistono nella biblioteca.

È importante puntualizzare che in quegli anni il catalogo è composto dalle schede cartacee: per questo ci si sofferma molto sulla forma e la scelta delle intestazioni, elementi fondamentali dei cataloghi per autori e titoli.

Dopo i Principi di Parigi il dibattito si concentra sugli aspetti normativi della catalogazione; vengono infatti pubblicati in questo periodo diversi codici catalografici di stampo nazionale. Bisogna attendere gli anni Novanta del secolo scorso perché ci si focalizzi nuovamente su aspetti di tipo teorico. Nel 1990 si tiene il Seminario di Stoccolma dedicato agli standard per i record bibliografici⁵¹. Le principali questioni al centro della discussione sono⁵²:

- il crescente costo della catalogazione e la conseguente ricerca verso una semplificazione dei contenuti bibliografici dei record;
- la diminuzione del costo della catalogazione attraverso una maggiore condivisione dei record, a livello sia nazionale che internazionale, per ridurre le duplicazioni;
- il forte incremento dei materiali pubblicati in formati diversi, e il maggiore interesse verso il controllo bibliografico universale⁵³;

⁵⁰ International Conference on Cataloguing Principles, «Statement of Principles Adopted by The International Conference on Cataloguing Principles», (1961), 1, http://www.ifla.org/files/assets/cataloguing/IMEICC/IMEICC1/statement_principles_paris_1961.pdf.

⁵¹ Olivia M. A. Madison, «The Origins of the IFLA Study on Functional Requirements for Bibliographic Records», *Cataloging & Classification Quarterly* 39, n. 3–4 (2005): 15–37, doi:10.1300/J104v39n03_02..

⁵² Ibid.

⁵³ Le attività centrali del controllo bibliografico sono l’analisi bibliografica e la catalogazione, con questo termine si intende l’insieme delle attività di acquisizione, descrizione e scambio sia delle risorse

- la consapevolezza dei benefici che sarebbero derivati dall'adattamento delle pratiche e dei codici catalografici al nuovo contesto elettronico;
- l'importanza di esaminare il ruolo degli universi bibliografici attraverso il punto di vista e le necessità degli utenti.

La domanda centrale del Seminario “che sintetizzava tutte queste questioni era: la catalogazione può essere semplificata e, allo stesso tempo, adattata ai contesti elettronici e alle necessità degli utenti?”⁵⁴ Viene così evidenziata la necessità di trovare un accordo internazionale per stabilire i requisiti minimi funzionali dei record bibliografici, in rapporto ai bisogni degli utenti e all'aumento della condivisione internazionale dei dati. Negli anni immediatamente successivi viene creato un Working Group che porta nel 1998 alla pubblicazione dei *Functional Requirements for Bibliographic Records*⁵⁵. Il modello FRBR cerca di rispondere alle mutate esigenze informative degli utenti, dovute all'avvento dell'elettronica; stabilisce nuove funzioni e scopi del catalogo e apre una nuova visione dell'universo bibliografico. Vengono delineate quattro funzioni principali⁵⁶, relative alle operazioni che gli utenti possono effettuare utilizzando i dati bibliografici per:

- trovare risorse che corrispondono alla ricerca;
- identificare un'entità⁵⁷;
- selezionare un'entità adeguata ai bisogni degli utenti;
- acquisire o ottenere l'accesso alle entità descritte.

documentarie che delle relative registrazioni catalografiche. L'idea del controllo bibliografico universale si fa risalire all'incirca all'inizio del XVI secolo, anche Jewett aveva avanzato l'idea di poter censire “quanto prodotto a livello internazionale dalle principali biblioteche degli Stati Uniti”, nel 1969 l'International Meeting of Cataloguing Experts (IMCE) tenuto a Copenaghen “getta le basi per lo scambio internazionale di informazioni per mezzo di descrizioni bibliografiche standard”. Giuliano Genetasio, «Analisi e controllo bibliografico», in *Biblioteconomia: guida classificata*, a c. di Mauro Guerrini, (Milano: Editrice Bibliografica, 2007), 355–61.

⁵⁴ Galeffi e Sardo, *FRBR* (Roma: Associazione italiana biblioteche, 2013), 10.

⁵⁵ IFLA Study Group on the Functional Requirements for Bibliographic Records e International Federation of Library Associations and Institutions, a c. di, *Functional requirements for bibliographic records: final report* (München: K.G. Saur, 1998).

⁵⁶ Ibid.

⁵⁷ Le entità rappresentano gli elementi di interesse per gli utenti che utilizzano i record bibliografici.

A queste funzioni Elaine Svenonius ha aggiunto quella della navigazione⁵⁸, ovvero la possibilità di esplorare l'universo bibliografico muovendosi da un'entità all'altra, attraverso le relazioni che vengono stabilite nelle descrizioni bibliografiche.

Rispetto ai Principi di Parigi, FRBR presenta un totale cambio di approccio. Le funzioni del catalogo vengono adesso intese come una serie di azioni compiute dagli utenti e tenute distinte da termini come *libro*, *autore*, *edizione* ed *opera*. Si parla infatti di *entità* per svincolare le modalità di ricerca dal legame con uno specifico supporto o con una specifica tipologia di oggetti bibliografici. La scelta è dovuta all'aumento esponenziale della varietà che caratterizza i documenti, non più meramente identificabili con l'oggetto *libro* e le sue peculiarità. Come già aveva teorizzato Cutter e come intorno agli anni Sessanta aveva analizzato Akos Domanovszky, che si era occupato dettagliatamente degli oggetti della catalogazione e ne aveva evidenziato aspetti intellettuali e materiali⁵⁹. Non vanno dimenticate nemmeno le riflessioni di Alfredo Serrai, sulle relazioni tra opera, documento e livello testuale, che probabilmente hanno fornito argomentazioni e spunti al nuovo modello concettuale⁶⁰. In sostanza “FRBR parte dall'oggetto d'interesse catalografico, e cerca di definire le caratteristiche per cui esso viene comunemente individuato sulla base delle entità che lo rappresentano”⁶¹. FRBR non rappresenta però dei principi, il suo scopo è quello di essere un modello concettuale. Barbara Tillett spiega molto bene il perché sia importante avere dei modelli concettuali:

Very simplistically, the bibliographic universe includes all of the things that libraries and archives collect; things like books, serials, maps, manuscripts, musical scores, soundrecordings, motion pictures, photographs, slides, etc. Currently the developers of the principles and rules are using the term resources to indicate these things. The bibliographic universe also includes the people and corporate bodies that play some role with respect to these things. This universe is subjective depending on the viewer. It may be a different set of things to a library or an archive, or to a music library or law library. Developing tools to help find things in this universe is a complex task that is greatly helped by having a plan. The plan gives guidance to the creators of the collections and to the creators of the descriptions of the things in those collections and to the designers of tools to help users find what they need and ideally to help them in a way that seemseasy to the user. This is the great work of libraries

⁵⁸ Galeffi e Sardo, *FRBR*.

⁵⁹ Akos Domanovszky, *Functions and objects of author and title cataloguing* (Budapest: Akademiai Kiado, 1974).

⁶⁰ Galeffi e Sardo, *FRBR*, 9.

⁶¹ Mauro Guerrini, «Le funzioni del catalogo dall'ICCP a FRBR», 2000, <http://www.aib.it/aib/sezioni/toscana/conf/frbr/guerrini.htm>.

and other information providers. A plan may be guided by models, be based on principles, follow certain rules and standards, that in turn follow policies, procedures, and agreements of those institutions that share in the work of providing a bibliographic universe and making it accessible⁶².

L'universo bibliografico include tutte le cose che le biblioteche, ma anche altre istituzioni culturali, collezionano. Le tipologie di materiali sono moltissime: libri, manoscritti, riviste, fotografie, musica registrata. Per indicare questi oggetti, chi sviluppa principi e norme catalografiche attualmente utilizza il termine *risorsa*, che si riferisce in modo generico a tutti questi oggetti. La presenza di modelli concettuali, basati su principi condivisi, facilita la costruzione di strumenti che semplificano l'accesso a tutte queste risorse. I principi guidano lo sviluppo di standard e norme. L'idea del modello concettuale nasce perché

solo un modello concettuale permette di avere una visione unificante dell'universo bibliografico che può essere letto, interpretato e, quindi, descritto in modi e con strumenti differenti, a seconda delle finalità si vogliono perseguire⁶³.

FRBR si sviluppa a partire dal modello entità-relazioni (e/r), un approccio molto conosciuto nella costruzione di database relazionali. Questo metodo si differenzia dagli altri, in quanto parte da uno schema concettuale astratto dell'universo preso in considerazione, che viene poi caratterizzato sulla base delle entità e delle relazioni che intercorrono tra di esse⁶⁴. Il modello comprende:

- le entità: tutti gli oggetti di *livello alto* sui quali si vogliono raccogliere informazioni. Il modello non prende in considerazione tutti i singoli *dati* ma parte dalle *cose* che quei dati descrivono. Ogni entità del modello serve dunque a raggruppare tutti i dati che presentano caratteristiche comuni;
- le proprietà: ogni entità definita nel modello è associata ad una serie di attributi, o caratteristiche che servono a descrivere e rendere ricercabile l'entità stessa. Gli

⁶² Barbara Tillett, *The bibliographic universe and the new IFLA cataloging principles: lectio magistralis in library science = L'universo bibliografico e i nuovi principi di catalogazione dell'IFLA: lectio magistralis di biblioteconomia* (Fiesole: Casalini libri, 2008), 6.

⁶³ Galeffi e Sardo, *FRBR*, 12; Tillett, *The bibliographic universe and the new IFLA cataloging principles: lectio magistralis in library science = L'universo bibliografico e i nuovi principi di catalogazione dell'IFLA: lectio magistralis di biblioteconomia*.

⁶⁴ Madison, «The Origins of the IFLA Study on Functional Requirements for Bibliographic Records».

attributi sono le informazioni che rappresentano le entità e che vengono generalmente cercate dagli utenti per trovare le risorse;

- le relazioni: rappresentano le connessioni tra le entità e consentono agli utenti di navigare all'interno dell'universo bibliografico. Generalmente gli utenti utilizzano gli attributi per trovare le entità desiderate. Le relazioni, invece, che vengono visualizzate in un record bibliografico, aiutano gli utenti ad esplorare le connessioni tra le entità trovate e le altre entità ad essa relazionate.

Partendo dalle funzioni *trovare*, *identificare*, *selezionare* e *ottenere* il team di lavoro ha identificato le entità principali sulle quali si vogliono raccogliere informazioni, suddividendole in tre gruppi:

- Gruppo 1: i prodotti della creazione intellettuale e artistica, costituiti da *opera*, *espressione*, *manifestazione* ed *esemplare*;
- Gruppo 2: *persone* ed *enti*, ovvero coloro che hanno una responsabilità nei confronti del contenuto intellettuale o artistico, della produzione fisica, della disseminazione o della custodia delle entità del Gruppo 1;
- Gruppo 3: *concetto* (una nozione astratta), *oggetto* (una cosa materiale), *evento* (un'azione o un avvenimento), *luogo* (un posto fisico), rappresentano le entità che possono fungere da soggetto per le opere. I soggetti inoltre possono essere rappresentati anche da tutte le entità del Gruppo 1 e 2.

Per fare un esempio pratico riguardo le entità del Gruppo 1:

- opera**¹ *Freedom* di Jonathan Franzen
 - espressione**¹ il testo originale in lingua inglese
 - manifestazione**¹ il libro pubblicato nel 2010 da Farrar, Straus & Giroux
 - esemplare**¹ la copia autografata dall'autore
 - espressione**² il testo in lingua italiana tradotto da Silvia Pareschi
 - manifestazione**² il libro pubblicato nel 2011 da Einaudi
 - esemplare**² la copia posseduta da una biblioteca, contrassegnata dal rispettivo numero di inventario

Ai fini di questa trattazione non è necessario entrare nel dettaglio del modello; è importante però segnalare come FRBR cerchi di chiarire la polisemia del termine *libro*. Come afferma Patrick Le Boeuf, quando diciamo *libro* forse abbiamo in mente uno specifico oggetto fisico, che FRBR chiama *esemplare*; oppure pensiamo ad una determinata pubblicazione di cui abbiamo l'ISBN e FRBR la chiama *manifestazione*; o ancora cerchiamo un libro in una determinata traduzione o un film doppiato nella lingua che ci interessa: FRBR chiama questo *espressione*. Quando invece ci chiediamo chi sia l'autore di un certo libro, forse ci stiamo riferendo ad un livello ancora più astratto, al contenuto concettuale che è alla base di tutte le versioni linguistiche, alla storia di fondo che rende tale quel libro. FRBR chiama questo livello *opera*⁶⁵.

Dai Principi di Parigi viene dunque ripresa la distinzione tra *opera* ed *edizione* ed estesa attraverso l'individuazione delle entità del Gruppo 1.

FRBR imprime una forte svolta al sistema catalografico. Dopo la pubblicazione del *Final Report* l'universo bibliografico viene analizzato alla luce di questo nuovo modello concettuale: i dati bibliografici presentano una struttura pre-FRBR, le funzioni dei cataloghi non sembrano in linea con le nuove esigenze degli utenti, inoltre si apre la possibilità di dialogare con altre istituzioni culturali, come gli archivi e i musei⁶⁶. Si sente quindi l'esigenza di una revisione dei Principi di Parigi che risalgono al 1961 e dopo trent'anni sembrano ormai obsoleti. La funzione dei principi, all'interno del sistema catalografico, viene chiarita da Elaine Svenonius, che precisa

principles, on the other hand, are directives for the design of the bibliographic language used to create such a system. This language normally takes the form of a code of rules. However, principles themselves are not rules but rather guidelines for the design of a set of rules⁶⁷.

I principi stabiliscono delle direttive utili alla creazione di un linguaggio che consenta la costruzione di sistemi bibliografici. Questo linguaggio prende normalmente la forma dei codici catalografici. Essi sono le linee guida, gli indirizzi, utilizzati per creare delle norme. FRBR invece è un modello concettuale che cerca di offrire una visione sull'universo

⁶⁵ Patrick Le Boeuf, «Brave New FRBR World», (2003), 1-2, http://www.ifla.org/files/assets/cataloguing/IMEICC/IMEICC1/papers_leboeuf.pdf.

⁶⁶ Carlo Bianchini e Mauro Guerrini, «Quis custodiet ipsos custodes?: osservazioni sulle relazioni fra FRBR, ICP, ISBD e RDA», *Il bibliotecario* 3, n. 1 (2009): 59–85.

⁶⁷ Elaine Svenonius, *The Intellectual Foundation of Information Organization* (Cambridge: MIT Press, 2009), 67-8.

bibliografico, sull'organizzazione delle entità che lo popolano, sulle informazioni utili ad identificare le entità e infine sulle relazioni che intercorrono in questo universo⁶⁸. Questa disintizione, tra i modelli concettuali e i principi, accresce l'esigenza di aggiornare i Principi di Parigi.

A partire dal 2003 prende così avvio un processo di revisione internazionale, promosso dall'IFLA⁶⁹, con lo scopo di redigere nuovi principi di catalogazione⁷⁰. Nel 2009 si arriva alla pubblicazione degli *Statement of International Cataloguing Principles (ICP)*. Alla base di questo documento vi erano tre obiettivi⁷¹:

- adattare i principi internazionali di catalogazione all'ambiente catalogafico contemporaneo;
- armonizzare i codici di catalogazione nazionali;
- suggerire delle regole guida per un codice di catalogazione internazionale.

Gli ICP vengono pensati per il catalogo elettronico e la rete; si propongono di essere validi per ogni risorsa bibliografica e per tutte le tipologie di biblioteca e istituzioni culturali; si fa inoltre riferimento al modello concettuale FRBR⁷².

I principi generali, ora sono⁷³:

- interesse dell'utente: tutte le decisioni catalogafiche devono essere adottate tenendo presente l'utente;
- uso comune: il lessico deve essere conforme a quello utilizzato dalla maggior parte degli utenti;

⁶⁸ Tillett, *The bibliographic universe and the new IFLA cataloging principles: lectio magistralis in library science = L'universo bibliografico e i nuovi principi di catalogazione dell'IFLA: lectio magistralis di biblioteconomia*.

⁶⁹ International Federation of Library Associations and Institutions, la federazione mondiale che riunisce associazioni e istituzioni bibliotecarie. Promuove la ricerca, lo sviluppo e la cooperazione internazionale, riguardo tutti i settori dell'ambito bibliotecario.

⁷⁰ Bianchini e Guerrini, «Quis custodiet ipsos custodes?: osservazioni sulle relazioni fra FRBR, ICP, ISBD e RDA».

⁷¹ Giuliano Genetasio, «I Principi Internazionali di Catalogazione (ICP) e il loro futuro», *JLIS.it* 3, n. 1 (2012): 1-18, doi:10.4403/jlis.it-4936.

⁷² Ibid.

⁷³ IFLA, Barbara B. Tillett, e Ana Lupe Cristán, a c. di, *IFLA Cataloguing Principles: the Statement of International Cataloguing Principles (ICP) and its glossary: in 20 languages* (München: K.G. Saur, 2009).

- presentazione: le descrizioni e le forme controllate dei nomi delle entità per i punti di accesso devono basarsi sul modo in cui ciascuna entità descrive sé stessa;
- accuratezza: l'entità descritta deve essere rappresentata fedelmente;
- sufficienza e necessità: nelle descrizioni e nelle forme controllate per i punti di accesso devono essere presenti solo gli elementi necessari per soddisfare l'utente e identificare in modo univoco un'entità;
- significatività: i dati devono essere bibliograficamente significativi;
- economia: bisogna preferire le modalità che contribuiscono all'economia complessiva (costo minore o approccio più semplice);
- coerenza e standardizzazione: le descrizioni e la creazione dei punti di accesso devono essere il più possibile normalizzate per una maggiore coerenza e una più facile condivisione dei dati;
- integrazione: le descrizioni devono basarsi il più possibile su regole comuni.

Gli obiettivi del catalogo, sulla base di FRBR, diventano:

- trovare risorse bibliografiche a partire da attributi e relazioni;
- identificare una risorsa bibliografica, avere quindi conferma che ciò che si trova corrisponda a ciò che si cerca;
- selezionare una risorsa appropriata alle necessità degli utenti;
- acquisire o ottenere accesso ad un esemplare descritto;
- navigare in un catalogo e al di fuori di esso.

Si inizia dunque a ragionare in termini di funzioni, cioè di operazioni che l'utente può fare quando cerca informazioni a partire dai dati bibliografici. Il suo interesse viene anteposto a tutte le scelte catalografiche, come già sosteneva Cutter. Non solo egli ora può usufruire della *navigazione* a partire dalle informazioni bibliografiche, ma si parla anche della possibilità che il catalogo instauri relazioni con l'ambiente web esterno.

Nel 2009 gli ICP gettano le basi affinché il catalogo elettronico si apra all'ambiente del web, con la conseguenza di modifiche normative e tecnologiche⁷⁴.

⁷⁴ Marchitelli, *Orientarsi tra le informazioni in biblioteca*, 24.

Le potenzialità offerte dall'infrastruttura della rete vengono ancora più percepite nell'ultima revisione degli ICP, pubblicata nel 2016. Rispetto al 2009 ai Principi generali vengono aggiunti quattro importanti concetti⁷⁵:

- interoperabilità: i dati bibliografici e di autorità devono poter essere condivisi e riutilizzati all'interno della comunità bibliotecaria e al di fuori di essa;
- apertura: le restrizioni sui dati, se presenti, devono essere minime e chiaramente indicate, per favorire la trasparenza e la conformità ai principi dell'*Open Access*;
- accessibilità: l'accesso ai dati bibliografici e di autorità, nonché le funzionalità di ricerca degli strumenti adottati, devono rispettare gli standard internazionali sull'accessibilità;
- razionalità: le regole dei codici catalografici devono essere sostenibili e non arbitrarie.

Questi elementi rappresentano la grande novità degli ICP 2016.

Il concetto di interoperabilità è un elemento fondamentale del paradigma attuale legato all'ambito informatico. La sua importanza è dovuta alla possibilità di rendere i dati delle biblioteche interscambiabili e riutilizzabili, sia all'interno dei propri sistemi che con sistemi esterni⁷⁶. Ci sono diverse tipologie di interoperabilità, le principali sono⁷⁷:

- interoperabilità tecnica: "si basa sullo sviluppo e l'applicazione di standard per la comunicazione, il trasferimento, l'archiviazione e la rappresentazione dei dati, come ad esempio i protocolli, i formati e i linguaggi di marcatura"⁷⁸;
- interoperabilità semantica: prevede lo sviluppo di tecnologie che stabiliscano corrispondenze fra termini appartenenti ad ambiti diversi. Questa necessità è dovuta al fatto che spesso vengono utilizzati termini diversi per descrivere concetti simili (ad esempio *author*, *creator*, *composer*), oppure termini identici

⁷⁵ IFLA Cataloguing Section e IFLA Meetings of Experts on an International Cataloguing Code, «Statement of International Cataloguing Principles (ICP)».

⁷⁶ Paul Gabriele Weston, *Gli strumenti della cooperazione in rete. Dal catalogo elettronico ai sistemi della ricerca interdisciplinare* (Napoli: ClioPress, 2003).

⁷⁷ Paul Miller, «Interoperability: What Is It and Why Should I Want It?», *Ariadne*, n. 24 (2000), <http://www.ariadne.ac.uk/issue24/interoperability>.

⁷⁸ Weston, *Gli strumenti della cooperazione in rete. Dal catalogo elettronico ai sistemi della ricerca interdisciplinare*, 11. Un esempio di protocollo è lo Z39.50 che serve a ricercare e recuperare attraverso una rete informatica informazioni in uno o più database, lo standard è coordinato dalla Library of Congress. «Z39.50 Maintenance Agency Page», <https://www.loc.gov/z3950/agency/>.

per indicare concetti differenti o analoghi, creando confusione per chi riutilizza quei dati;

- interoperabilità politica/umana: richiede la scelta consapevole da parte delle istituzioni di rendere accessibili e disponibili i dati delle proprie risorse con nuove procedure gestionali e personale qualificato, sensibilizzando nel contempo la propria utenza ad usufruire di una rete conoscitiva più ampia e complessa⁷⁹;
- interoperabilità internazionale: riguarda la possibilità di condividere i dati a livello internazionale, cercando soluzioni che oltrepassino le barriere linguistiche e culturali⁸⁰;
- interoperabilità multidisciplinare: presuppone l'avvio di iniziative congiunte da parte di biblioteche, archivi e musei⁸¹.

L'inserimento dell'interoperabilità tra gli ICP 2016 sottointende, da parte della comunità catalografica, la consapevolezza del valore dei dati creati e gestiti dalle biblioteche. Da un punto di vista tecnologico, attualmente il principale protocollo utilizzato per lo scambio dei dati è lo Z39.50, sviluppato per le biblioteche e limitato principalmente a tale ambito. Questo ostacola la possibilità di poter condividere e riutilizzare i dati appartenenti ad altre comunità⁸². La garanzia dell'interoperabilità semantica è data, nel contesto catalografico, dalla struttura dei dati e dai formati che vengono scelti, come vedremo meglio nel paragrafo successivo.

Il riferimento al concetto di apertura dei dati è un altro elemento connesso al cambiamento del paradigma informativo. Nel testo dei Principi 2016 si fa riferimento *all'IFLA Statement on Open Access*, un documento nel quale l'IFLA chiarisce il suo impegno nel rispettare i principi della libertà di accesso alle informazioni, fondamentale per il benessere della società⁸³. Il movimento *Open Access* rientra nei vari movimenti *open* e “si concretizza nella disponibilità dei contenuti digitali online” rivolgendosi in particolare “agli articoli di riviste e alle ricerche svolte in campo universitario, consentendone la

⁷⁹ Weston, *Gli strumenti della cooperazione in rete. Dal catalogo elettronico ai sistemi della ricerca interdisciplinare*, 11-2.

⁸⁰ Ibid.

⁸¹ Ibid.

⁸² Fabio Di Giammarco, «Le biblioteche nella rete dei linked data», *Digitalia* 1, (2012): 138–41, <http://digitalia.sbn.it/article/view/544/383>.

⁸³ «IFLA Statement on Open Access», (2011), <http://www.ifla.org/files/assets/hq/news/documents/ifla-statement-on-open-access.pdf>.

pubblicazione e l'accesso/utilizzo gratuito”⁸⁴. Negli ICP 2016 il concetto di *openness* viene dunque limitato allo specifico movimento dell'*Open Access*, scelta che in qualche modo circoscrive il significato di apertura dei dati all'accessibilità dei contenuti delle pubblicazioni digitali o digitalizzate, in particolare di ambito scientifico e universitario. Probabilmente perché le descrizioni bibliografiche stesse possono essere considerate alla stregua di documenti scientifici.

Infine si parla di *accessibilità* in riferimento all'*IFLA Code of Ethics for Librarians and other Informations Workers*, nella quale si specifica che i bibliotecari e gli addetti nel campo dell'informazione devono assicurarsi che i siti delle biblioteche o di altre istituzioni culturali siano conformi agli standard internazionali sull'accessibilità, e che l'accesso alle informazioni non venga ostacolato da barriere⁸⁵. L'accessibilità in questo contesto riguarda la capacità dei sistemi informatici di offrire informazioni e contenuti consultabili da tutti, anche a coloro che a causa di disabilità possono avere esigenze particolari riguardo la fruizione dei dati⁸⁶.

L'accessibilità dei siti e dei portali delle biblioteche è garantita dalla presentazione e dalle modalità di fruizione dei contenuti, dalle interfacce che vengono create, dalle tecnologie che ne sono alla base e dalle soluzioni adottate. L'apertura dei dati può riguardare la scelta delle licenze ma anche dei formati standard, nel senso che il formato che viene scelto stabilisce i gradi di condivisione dati⁸⁷. La loro struttura inoltre incide sull'interoperabilità a livello di condivisi internazionale, con altri ambiti, e da un punto di vista tecnologico e semantico.

È attraverso la scelta dei formati, ovvero i metadati che possono essere elaborati e trattati dai software, che si può incidere profondamente sull'interoperabilità e sulle possibilità di condivisione e riutilizzo dei dati. Ma tale scelta offre anche la possibilità di presentare

⁸⁴ Antonella De Robbio e Silvia Giacomazzi, «Dati aperti con LODe», *Bibliotime* XIV, n. 2 (2011), <http://www.aib.it/aib/sezioni/emr/bibtime/num-xiv-2/derobbio.htm>.

⁸⁵ «IFLA Code of Ethics for Librarians and other Information Workers (full version)», 2012, <http://www.ifla.org/publications/node/11092>.

⁸⁶ La normativa italiana riguardo l'accessibilità si occupa di indicare gli aspetti tecnici, infrastrutturali e amministrativi, attraverso l'emanazione di linee guida, leggi e circolari. In particolare si fa riferimento all'Allegato A del Decreto Ministeriale dell'8 luglio 2005, che a sua volta si ispira alle raccomandazioni del World Wide Web Consortium «Accessibilità | AgID», <http://www.agid.gov.it/agenda-digitale/pubblica-amministrazione/accessibilita>; «Linee guida per l'accessibilità dei contenuti Web (WCAG) 2.0», <https://www.w3.org/Translations/WCAG20-it/>.

⁸⁷ De Robbio e Giacomazzi, «Dati aperti con LODe».

all'utente un universo conoscitivo fatto di relazioni e connessioni, sia tra i dati interni di un sistema che tra dati di sistemi esterni. Nel paragrafo successivo verranno dunque presi in esame i più importanti formati e standard bibliografici, per capire in che direzione si sta andando

1.2. I formati bibliografici e gli standard

Dall'analisi dell'evoluzione dei principi e dei modelli catalogafici si è visto come l'informatica e l'infrastruttura della rete abbiano profondamente inciso sulle linee guida destinate alla creazione, gestione e diffusione dei dati bibliografici.

L'avvento dell'informatica ha avuto una forte ripercussione anche sulle modalità di strutturazione e scambio dei dati catalogafici, rendendo necessari i *formati bibliografici*, sistemi automatici che aiutano a gestire il contenuto dei dati⁸⁸. I formati bibliografici fanno parte dei linguaggi di marcatura e in particolare di quelli dichiarativi e descrittivi. Essi forniscono all'elaboratore delle indicazioni sulle caratteristiche formali o di presentazione dei documenti. Questi linguaggi sono formati da una serie di marcatori, detti tag, che hanno la funzione di contrassegnare ed esplicitare la funzione logico-strutturale dei blocchi informativi ai quali si riferiscono⁸⁹.

La famiglia di formati standard più conosciuti ed utilizzati dalla comunità bibliotecaria è il MARC (Machine-Readable Cataloging), il cui progetto pilota risale al 1968⁹⁰. Il MARC rappresenta una tappa fondamentale nella storia della catalogazione. Non solo infatti rappresentò una tecnologia all'avanguardia per quei tempi, ma riuscì anche a diffondere lo sviluppo dell'automazione all'intera comunità bibliotecaria. Lo scopo principale era quello di trovare un mezzo che permettesse di passare dall'utilizzo delle descrizioni bibliografiche cartacee ad un sistema automatizzato, garantendo allo stesso tempo descrizioni analitiche e comprensibili anche al di là delle differenze linguistiche. Il MARC serve a scambiare i dati tra biblioteche o reti di biblioteche. Esso si basa su un

⁸⁸ Stefano Gambari e Matteo Barucci, «Formati di registrazione per record bibliografici leggibili dalla macchina (MACHINE-Readable Cataloging, MARC)», in *Biblioteconomia: guida classificata* (Milano: Editrice Bibliografica, 2007), 416–20.

⁸⁹ Andrea Marchitelli e Giovanna Fragimelica, *OPAC* (Roma: Associazione italiana biblioteche, 2012).

⁹⁰ *The MARC Pilot Project: Final Report on a Project Sponsored by the Council on Library Resources, Inc* (Washington: Government Printing Office, 1968).

sistema di numeri corrispondenti ai vari blocchi informativi della notizia bibliografica, a loro volta suddivisi in sottocampi contrassegnati da lettere, che servono a specificare il contenuto delle informazioni contenute nei campi. La struttura del MARC consente lo scambio dei dati, mentre il contenuto delle etichette (*tag*) varia a seconda del formato di riferimento. Ad esempio il MARC21 e l'UNIMARC utilizzano tag differenti per esprimere lo stesso contenuto. Infine i valori che vengono immessi nei vari campi dipendono dalle norme catalografiche di riferimento.

Nel 2002 Roy Tennant intitola un suo celebre articolo *MARC must die*⁹¹, nell'intervento il MARC viene etichettato come anacronistico per la rigidità e il basso livello di granularità. L'articolo si chiude con questa esortazione

Libraries exist to serve the present and future needs of a community of users. To do this well, they need to use the very best that technology has to offer. With the advent of the web, XML, portable computing, and other technological advances, libraries can become flexible, responsive organizations that serve their users in exciting new ways. Or not. If libraries cling to outdated standards, they will find it increasingly difficult to serve their clients as they expect and deserve⁹².

Il web e i suoi nuovi linguaggi vengono presi in considerazione anche dalla comunità bibliotecaria, si comincia infatti a valutare l'uso di XML (eXtensible Markup Language), sviluppato nel 1998 dal W3C (World Wide Web Consortium)⁹³, come meta-linguaggio descrittivo per la codifica di dati afferenti a documenti sul web⁹⁴. Il MARCXML⁹⁵ è stato sviluppato per rispondere alla necessità di codificare i record MARC in XML, rendendoli comprensibili al mondo del web. Inoltre sono stati creati appositi software che consentono di convertire in modo automatico i record esistenti in MARXML⁹⁶. Questo formato presenta però delle criticità, che Karen Coyle spiega efficacemente

MARCXML was developed as a pure serialization of the MARC format. "Serialization" means that the data encoding of MARC was translated directly to XML without any related transformation of the data itself. Although this produced a record that could be managed with XML-aware software, it did

⁹¹ Roy Tennant, «MARC Must Die», *Library Journal*, (2002), <http://lj.libraryjournal.com/2002/10/ljarchives/marc-must-die/>.

⁹² Ibid.

⁹³ «World Wide Web Consortium (W3C)», <https://www.w3.org/>.

⁹⁴ Oreste Signore, «Strutturare la conoscenza: XML, RDF e Semantic Web», (2003), <http://www.w3c.it/papers/ck2003.pdf>.

⁹⁵ «MARC in XML» <https://www.loc.gov/marc/marcxml.html>.

⁹⁶ Agnese Galeffi, «Standard di catalogazione», in *Biblioteche e biblioteconomia: principi e questioni*, a c. di Giovanni Solimine e Paul Gabriele Weston, (Roma: Carocci editore, 2015), 255–80.

nothing to improve the kind of data that could be conveyed in library bibliographic records. It also did nothing to address some of the limitations of the MARC record. The MARCXML standard is kept one-to-one with the original MARC record, with the single exception that field and record sizes are not enforced. [...] But the limitation on the number of subfields to a field remains, even though there are fields that have no open subfields available for expansion. Other inconveniences also remain, such as the non-repeatability of the MARC fixed field information, which then forces some repeatable elements like languages and dates to be coded in more than one field to accommodate repeatability. MARCXML was never allowed to develop as its own technology, and therefore did not present a change. Library data in XML, rather than in MARCXML, could have represented a real change in capabilities. It might also have provided a better transition to new technology than we now have, because we could have resolved some of the more awkward elements of MARC over a decade or more, with a gradual update to the library systems that use this data. Today we either have to carry those practices on to our future data, or we need to make a great leap forward and break with our past⁹⁷.

In sostanza la serializzazione del MARC in XML ripropone comunque tutte le limitazioni del MARC. Restano alcuni inconvenienti, dati dalla non ripetibilità di alcuni campi o dal numero limitato dei sottocampi. In questo senso la Coyle sostiene la necessità di passare alle nuove infrastrutture che il web semantico offre, con la tecnologia che ne sta alla base: RDF (Resource Description Framework), lo strumento proposto dal W3C per la codifica, lo scambio e il riuso dei metadati strutturati⁹⁸.

Una tappa decisiva in questa direzione è rappresentata dal rapporto *On the Record*⁹⁹, pubblicato nel 2008 ad opera di un gruppo di studio commissionato dalla Library of Congress per valutare il futuro dei record bibliografici e i possibili sviluppi del MARC. Le critiche che vengono mosse riguardano il fatto che il formato MARC è utilizzato principalmente dalla comunità bibliotecaria e questo impedisce lo scambio e il riutilizzo dei dati con altre istituzioni culturali, inoltre sono necessari metadati che accolgano le descrizioni dei professionisti ma anche le annotazioni e i commenti degli utenti. Si delinea poi l'importanza di utilizzare degli identificatori neutri per le intestazioni e i termini, poiché il linguaggio naturale non consente di stipulare equivalenze tra dati

⁹⁷ Karen Coyle, *FRBR, Before and After: A Look at Our Bibliographic Models* (Chicago: ALA Editions, 2016), 53.

⁹⁸ «RDF - Semantic Web Standards», <https://www.w3.org/RDF/>.

⁹⁹ Library of Congress Working Group on the Future of Bibliographic Control, «On the Record» (Washington: Library of Congress, 2008), <https://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf>.

linguisticamente diversi, e in questo senso il Rapporto fa riferimento a VIAF (Virtual International Authority File) come buon esempio¹⁰⁰.

On the Record porta nel 2011 all'iniziativa denominata *Bibliographic Framework for the Digital Age*, sempre promossa dalla Library of Congress, un progetto focalizzato sull'ambiente web e sulle tecnologie offerte dai linked data, che pone RDF come modello di dati principale¹⁰¹. Viene sviluppato così *Bibliographic Framework Initiative*, conosciuto come BIBFRAME, ispirato a FRBR, che si propone di sostituire il MARC e di consentire la piena integrazione dei dati bibliografici nel web¹⁰². BIBFRAME è uno standard di dati che presenta un vocabolario compatibile con il web aperto e i suoi linguaggi, e che cerca in qualche modo di riportare le pratiche catalogafiche del MARC in RDF¹⁰³.

Oltre ai formati di dati come il MARC e BIBFRAME, tra gli standard va annoverato anche ISBD (International Standard Bibliographic Description), la cui prima versione viene pubblicata nel 1974¹⁰⁴. L'obiettivo di ISBD è quello di fornire indicazioni sulla presentazione e sul contenuto dei *dati descrittivi* in modo da favorire lo scambio internazionale delle registrazioni e aiutare la loro comprensione anche al di là delle barriere linguistiche. Questo standard indica quali sono i dati necessari per rappresentare le risorse e le fonti da cui trarre le informazioni. Prevede inoltre che gli elementi vengano suddivisi in aree e assegna un valore alla sequenza dei dati attraverso l'utilizzo di una punteggiatura convenzionale. ISBD è sottoposto ad una continua revisione; nel 2008 la volontà di mantenere questo standard aggiornato si è concretizzata nella creazione dell'ISBD Linked Data Study Group, che ha lo scopo di promuovere l'interoperabilità e il riuso dei dati bibliografici nell'ambiente del web semantico e che ha portato alla pubblicazione degli elementi ISBD, espressi come classi e proprietà RDF¹⁰⁵.

¹⁰⁰ Antonella Trombone, «Il progetto BIBFRAME della Library of Congress: come stanno cambiando i modelli strutturali e comunicatIbid. dei dati bibliografici», *AIB studi* 55, n. 2 (2015), <http://aibstudi.aib.it/article/view/11100>.

¹⁰¹ Marchitelli e Fragimelica, *OPAC*.

¹⁰² Galeffi, «Standard di catalogazione».

¹⁰³ Coyle, *FRBR, Before and After*.

¹⁰⁴ Galeffi, «Standard di catalogazione».

¹⁰⁵ «IFLA -- ISBD Linked Data Study Group», <http://www.ifla.org/node/1795>.

Un ulteriore elemento di cambiamento è rappresentato da RDA (Resource Description and Access)¹⁰⁶, uno standard di contenuto nato per l'ambiente web, che presenta relazioni qualificate, come richiesto da RDF, e una maggiore granularità dei dati. RDA è stato pensato per essere utilizzato attraverso diversi modelli tecnologici ma in particolare esso è compatibile con il mondo del web semantico, spostando l'attenzione dal *record bibliografico* ai singoli *dati* che lo compongono¹⁰⁷. I record bibliografici attuali contengono moltissime informazioni semantiche sui dati che li compongono. I campi e i sottocampi del MARC, così come la punteggiatura ISBD, servono proprio ad attribuire significato ai dati. Quando però un record MARC viene trasferito nel web tradizionale esso perde molto del suo significato perché la semantica di questo formato non è comprensibile ai motori di ricerca, che trattano le informazioni come *blocchi di testo*. Inoltre affinché sia possibile riutilizzare le informazioni è necessario che esse siano qualificate come singoli dati significativi¹⁰⁸. RDA, da questo punto di vista, si propone come standard internazionale per la metadattazione delle risorse, muovendosi nella direzione del web semantico e dei linked data, favorendo l'utilizzo di vocabolari espressi in RDF¹⁰⁹.

¹⁰⁶ Joint Steering Committee for Development of RDA, «RDA: Resource Description and Access», <http://www.rdatoolkit.org/>.

¹⁰⁷ Carlo Bianchini, «Dai cataloghi alla navigazione semantica», in *Linked data per biblioteche, archivi e musei: perché l'informazione sia del web e non solo nel web* (Milano: Editrice bibliografica, 2015), 173–204.

¹⁰⁸ Mauro Guerrini e Tiziana Possemato, *Linked data per biblioteche, archivi e musei: perché l'informazione sia del web e non solo nel web*, (Milano: Editrice bibliografica, 2015).

¹⁰⁹ *Ibid.*, 132.

1.3. Gli strumenti di accesso ai dati bibliografici

I dati bibliografici vengono creati a partire dai codici di catalogazione e dagli standard, che a loro volta cercano di assorbire le indicazioni degli ICP e dei modelli concettuali come FRBR. Essi inoltre sono veicolati dai formati che ne consentono la fruizione e lo scambio. Il principale accesso ai dati creati dalle biblioteche oggi è fornito dagli OPAC (Online Public Access Catalogue) e dai Discovery Tool.

Gli OPAC sono cataloghi elettronici che gli utenti possono liberamente consultare senza doversi recare in biblioteca¹¹⁰. Consentono di fare ricerca sulle raccolte bibliotecarie, prenotare un libro, richiedere informazioni e inserire commenti o valutazioni, attraverso l'utilizzo di una connessione e di un computer. Alcuni OPAC fanno parte di ILS (Integrated Library System) o LMS (Library Management System), software che gestiscono in maniera integrata i processi gestionali che avvengono in biblioteca, come acquisizioni, catalogazione, circolazione. Altri OPAC invece sono software a sé stanti che, attraverso procedure di esportazione e conversione dei dati bibliografici, si occupano principalmente di esporre i dati agli utenti¹¹¹.

L'architettura degli OPAC si basa principalmente su tre elementi¹¹²:

- l'interfaccia: predisposta per consentire agli utenti di fare delle ricerche;
- il database: che contiene le notizie bibliografiche, generalmente di tipo relazionale basato su tabelle, che suddividono le varie tipologie di dati, riuniti attraverso linguaggi come SQL (Structure Query Language);
- il modulo di ricerca: utilizzato per recuperare i record sia a livello gestionale che a livello utente.

Dai primi cataloghi elettronici, che riproducevano l'organizzazione e le funzioni di ricerca dei tradizionali cataloghi cartacei a schede¹¹³, gli OPAC si sono evoluti sempre più verso il mondo del web, cominciando a presentare interfacce di ricerca più semplici e *friendly*.

¹¹⁰ Marchitelli e Fragimelica, *OPAC*.

¹¹¹ Ibid.

¹¹² Ibid.

¹¹³ Carlo Bianchini, «Universo bibliografico, descrizione e accesso», in *Biblioteche e biblioteconomia: principi e questioni*, a c. di Giovanni Solimine e Paul Gabriele Weston, (Roma: Carocci, 2015), 229–54.

Si cominciano a sviluppare OPAC *arricchiti e sociali*¹¹⁴, che favoriscono l'interazione tra gli utenti e il catalogo, attraverso l'aggiunta di copertine o abstract, l'utilizzo di icone, la possibilità di leggere o inserire recensioni sulle risorse o anche di interagire con altri lettori all'interno di blog. Un'altra caratteristica di questi OPAC sono le interfacce *google-like*, che riguarda non tanto le modalità di interrogazione dei dati e le risposte fornite, quanto invece la presentazione di un'unica stringa di ricerca in cui inserire le parole, proprio come avviene nei motori di ricerca¹¹⁵.

Nascono poi i cataloghi di nuova generazione, che sono in grado di fornire l'accesso ad un set di dati molto ampio, non solo relativi alla raccolta fisica posseduta dalla biblioteca, ma anche a banche dati, periodici elettronici o piattaforme con contenuti digitali¹¹⁶. Da questa funzione di ricerca allargata prendono il nome i *Discovery Tool*, strumenti con i quali viene fatta una ricerca generica, successivamente raffinata attraverso l'uso di faccette. Essi permettono la *navigazione*, prevista dagli ICP, che consente agli utenti di muoversi non solo all'interno del catalogo ma anche verso ambienti esterni.

I cataloghi oggi sono dunque più collaborativi, amichevoli e sociali, ma non ancora pienamente integrati nella rete. Sono *sul web* ma non fanno parte *del web*¹¹⁷. A proposito di questo Guerrini e Possemato affermano

Constatiamo che i dati prodotti dalle biblioteche – i cataloghi –, la cui redazione ha richiesto elaborazione di normative, competenze professionali e finanziamenti, non sono sul web, ma isolati dal web. I cataloghi non sono, infatti, integrati nel web, non sono interrogabili, pur essendo il web il luogo in cui la maggior parte degli utenti lavora, gioca, opera e crea altra informazione. La questione, dunque, è: Come modificare i cataloghi e i dati perché siano del web e non solo sul web?. È proprio la filosofia che sottende la tecnologia dei linked data che può offrire un interessante punto di partenza per il raggiungimento di questo obiettivo strategico, pena la morte dei cataloghi, abbandonati dagli utenti a favore di altri strumenti di reperimento dell'informazione, come i motori di ricerca¹¹⁸.

Il web dunque, che si configura oggi come il luogo privilegiato di accesso all'informazione, porta a ripensare la struttura e le funzionalità degli strumenti di mediazione informativa, rappresentati dai cataloghi. Soddisfare le Funzioni e gli Obiettivi

¹¹⁴ Carlo Bianchini, «Universo bibliografico, descrizione e accesso».

¹¹⁵ Ibid.

¹¹⁶ Marchitelli e Fragimelica, *OPAC*.

¹¹⁷ Coyle, «Linked Data».

¹¹⁸ Mauro Guerrini e Tiziana Possemato, «Linked Data: A New Alphabet for the Semantic Web», *JLIS.it* 4, n. 1 (2013): 67, doi 10.4403/jlis.it-6305.

individuare dagli ICP, e potenziare di conseguenza le connessioni tra il contesto delle biblioteche e ambiti esterni, può migliorare la funzione di navigazione. La scoperta di nuova conoscenza può essere generata proprio da un più ampio *network* informativo, presentato da nuove tipologie di cataloghi.

1.4. Il presente e il futuro¹¹⁹

In questo capitolo si è cercato di tracciare brevemente l'evoluzione dei dati bibliografici, partendo dai principi, attraverso i formati e gli standard, per arrivare infine al panorama attuale dei cataloghi.

Oggi i *Principi Internazionali di Catalogazione* utilizzano termini come *apertura*, *interoperabilità*, *accessibilità* e *navigazione*, tutte funzioni garantite dall'applicazione di determinate tecnologie informatiche, che non riguardano i contenuti dei dati ma il mezzo che viene utilizzato affinché essi acquisiscano queste proprietà.

BIBFRAME, che oggi vorrebbe sostituirsi al MARC, è uno standard di dati essenzialmente basato su RDF e rivolto ai linked data come nuova modalità di strutturazione e condivisione dei dati.

Gli elementi ISBD sono stati pubblicati su Open Metadata Registry, e le classi e le proprietà convertite in RDF¹²⁰.

Nel 2011 è stata pubblicata la versione ufficiale del modello concettuale FRBR, espresso come vocabolario RDF¹²¹.

Con FRBR e RDA si sancisce il passaggio dal *record* al *dato*, nella direzione segnata dal web semantico che prevede che le informazioni vengano scomposte in singoli dati identificati da URI (Uniform Resource Identifier), cioè identificatori univoci e persistenti¹²². Il tema degli URI era già emerso nel Rapporto *On the Record* dove si prescriveva non solo la necessità di rendere i dati univocamente identificabili nel contesto

¹¹⁹ Coyle, *FRBR, Before and After*, 53.

¹²⁰ «IFLA ISBD Vocabularies», <http://iflstandards.info/ns/isbd/>.

¹²¹ Coyle, *FRBR, Before and After*.

¹²² Antonella Iacono, «Dal record al dato. Linked data e ricerca dell'informazione nell'OPAC.», *JLIS.it* 5, n. 1 (2013): 77–102, doi 10.4403/jlis.it-9095.

del web, ma si poneva l'obiettivo di renderli compatibili con gli standard e le tecnologie del web¹²³.

Il mondo del web ed elementi come XML, RDF, URI, linked data e web semantico sono dunque chiaramente entrati nel linguaggio catalografico. Ma cosa sono i linked data e quali tecnologie li rappresentano? Cercare di delineare un quadro introduttivo di questo ambito è necessario per capire come i linked data possono essere applicati ai dati bibliografici.

¹²³ Guerrini e Possemato, «Linked Data».

CAPITOLO SECONDO

Linked (open) data

Il termine linked data si riferisce ad un insieme di buone pratiche per la strutturazione e la pubblicazione di dati nel web¹²⁴. I linked data si propongono di rendere i dati maggiormente interoperabili e di consentire la scoperta di nuova conoscenza a partire dalle connessioni che si possono creare tra i dati di differenti comunità¹²⁵. Esso rappresentano una possibilità per sprigionare le potenzialità dei dati facendoli emergere nel web, rendendoli condivisibili e accessibili attraverso formati standard. L'idea fu di Tim Berners-Lee, che presentò questa soluzione per realizzare il web semantico¹²⁶.

2.1. Linked data e web semantico

Con web semantico si indica un'estensione dell'attuale web nella quale alle informazioni viene assegnato un significato ben definito, capace di migliorare l'interazione tra computer e persone¹²⁷. Secondo Tim Berners-Lee gli sviluppi del web semantico offrono nuove funzionalità che permettono alle macchine di processare e "comprendere" i dati in modo più efficiente¹²⁸. Affinché ciò si realizzi, i dati sul web devono essere inseriti o convertiti in una forma che le macchine possano elaborare, in modo da generare nuove connessioni tra le informazioni, nuove applicazioni e sistemi di ricerca più avanzati¹²⁹.

La sfida di questa visione del web è fornire un linguaggio che esprima il significato dei dati e che consenta a qualunque sistema di rappresentazione della conoscenza di poter

¹²⁴ Seth van Hooland e Ruben Verborgh, *Linked Data for Libraries, Archives and Museums: How to Clean, Link and Publish Your Metadata* (Chicago: Neal-Schuman, 2014).

¹²⁵ Coyle, *FRBR, Before and After*.

¹²⁶ Tim Berners-Lee, «Linked Data - Design Issues», 2006, <https://www.w3.org/DesignIssues/LinkedData.html>.

¹²⁷ Tim Berners-Lee, James A. Hendler, e Ora Lassila, «The Semantic Web. A new form of web content that is meaningful to computers will unleash a revolution of new possibilities», *Scientific American magazine*, (2001), https://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf.

¹²⁸ Ibid.

¹²⁹ Tim Berners-Lee e Mark Fischetti, *L'architettura del nuovo Web: dall'inventore della rete il progetto di una comunicazione democratica, interattiva e intercreativa* (Milano: Feltrinelli, 2001).

diventare parte del web¹³⁰. Le varie comunità che popolano il web hanno infatti creato diverse descrizioni e definizioni per rappresentare i propri documenti e i propri dati, utilizzando linguaggi e formati spesso incompatibili tra di loro o che non ne esplicitano il significato in modo che questo possa essere processato dalla maggior parte dei programmi scritti per il web. Se una persona utilizza un comune motore di ricerca per cercare un libro posseduto dalla Biblioteca di Chioggia, inserendo nella stringa di ricerca ad esempio “*Promessi sposi – Chioggia – Biblioteca*” non vedrà tra i risultati le notizie contenute nel catalogo della Biblioteca, perché i dati non sono strutturati con una semantica comprensibile ai linguaggi del web. Se invece si inserisce nella stringa di ricerca di Google un titolo qualsiasi, come *Il vagabondo delle stelle*, compare sulla destra dello schermo un box, nel quale vengono presentate informazioni sull’*opera* letteraria, come il nome dell’*autore*, l’anno della prima *edizione*, il *genere*, un *abstract*, altre *opere correlate* etc. Questo riquadro sfrutta il Google Knowledge Graph, una funzione di ricerca che utilizza proprio la semantica di dati processabili¹³¹. La strutturazione e la presentazione dei dati che vengono visualizzati nel box di Google è molto simile al modello FRBR: le informazioni riguardano l’*opera* concettuale, vengono distinte *espressioni* e *manifestazioni*, e sono riportate anche delle *relazioni tra opere*. È evidente l’influenza dei modelli nati all’interno del contesto biblioteconomico. Nonostante questo al momento, i dati utilizzati provengono da Wikipedia o da librerie digitali e non dalle biblioteche. Fornire un quadro teorico sul web semantico significa in qualche modo capire come le biblioteche potrebbero valorizzare i propri dati facendoli emergere nel più ampio contesto del web, potenziandone così anche la ricercabilità e avvicinandosi in questo modo al numero sempre maggiore di utenti remoti.

Il web semantico mira dunque ad esplicitare il significato dei contenuti attraverso l’aggiunta di annotazioni semantiche, cioè informazioni che descrivono il significato dei contenuti¹³². La sua realizzazione richiede che i documenti pubblicati nel web vengano associati a metadati adatti all’elaborazione automatica: questa operazione rende possibili

¹³⁰ Berners-Lee, Hendler, e Lassila, «The Semantic Web. A new form of web content that is meaningful to computers will unleash a revolution of new possibilities».

¹³¹ «Algoritmi – Dentro la Ricerca – Google»,
<https://www.google.it/intl/it/insidesearch/howsearchworks/algorithms.html>.

¹³² Tom Heath e Christian Bizer, «Linked Data: Evolving the Web into a Global Data Space», *Synthesis Lectures on the Semantic Web: Theory and Technology* 1, n. 1 (2011): 1–136, doi:10.2200/S00334ED1V01Y201102WBE001.

ricerche più complesse, con un *recall* più preciso. Permette inoltre la costruzione di reti relazionali non solo tra i documenti, ma anche tra i dati contenuti in questi stessi documenti, secondo logiche più elaborate dei link ipertestuali¹³³.

Il web semantico vede il passaggio da un ambiente popolato da documenti e da link tra documenti ad un ambiente in cui sono i singoli dati ad essere relazionati gli uni agli altri.

2.1.1. Dal web dei documenti al web dei dati

Il web di oggi è definito *web di documenti*, tenuti insieme da una rete di nodi collegati tra loro tramite link. Il web semantico presuppone un'evoluzione verso un ambiente in cui i dati diventano il centro dei nodi e i link relazioni tra entità¹³⁴.

Nel 2008 Christian Bizer, Thom Heat e Tim Berners-Lee presentano alla 17. Conferenza internazionale sul World Wide Web un prospetto che chiarisce la differenza tra il web dei documenti e il web dei dati.

Il web dei documenti ha come componenti principali¹³⁵:

- HTML (HyperText Markup Language)¹³⁶: un linguaggio di marcatura usato per creare pagine web, che utilizza i tag per fornire informazioni sulla visualizzazione dei documenti. La sintassi di questo linguaggio di pubblico dominio è stabilita dal W3C¹³⁷;
- link non tipizzati, cioè non qualificati, tra documenti: collegamenti ipertestuali che demandano alle persone il compito di comprendere il significato della relazione;
- Web-API (Application Programming Interface): interfacce grafiche che consentono l'accesso a delle componenti interne di un software, in modo che i

¹³³ Tom Heath e Christian Bizer, «Linked Data: Evolving the Web into a Global Data Space»

¹³⁴ Berners-Lee, Hendler, e Lassila, «The Semantic Web. A new form of web content that is meaningful to computers will unleash a revolution of new possibilities».

¹³⁵ Christian Bizer et al., «Linked Data on the Web», in *Proceedings of the 17th international conference on World Wide Web, Beijing China, 2008* (New York: ACM Press, 2008), 1265, doi:10.1145/1367497.1367760.

¹³⁶ «W3C HTML», <https://www.w3.org/html/>.

¹³⁷ Il W3C è nato nel 1994 grazie a Tim Berners-Lee, questa comunità internazionale ha lo scopo di sviluppare standard applicabili al Web, per consentire il pieno sviluppo di tutte le sue potenzialità. «About W3C» <https://www.w3.org/Consortium/>.

programmatori possano sviluppare delle funzionalità aggiuntive o creare delle integrazioni con altre applicazioni¹³⁸. Ad esempio Google Maps, Facebook e Twitter mettono a disposizione delle API per consentire agli sviluppatori di integrarle nei loro programmi e personalizzarle;

- database: insiemi di dati che vengono memorizzati in un elaboratore elettronico¹³⁹.

Il web dei documenti è un ambiente popolato da documenti strutturati in pagine HTML o altri linguaggi di marcatura. Gli oggetti visualizzati sono creati per essere interpretati dagli esseri umani e non dalle macchine, pertanto ostacolano le applicazioni software nell'estrazione di singoli dati strutturati¹⁴⁰.

Il web dei documenti risulta dunque formato da oggetti strutturati in modo che le macchine comprendano come essi devono essere presentati e visualizzati, senza però dichiarare il contenuto degli elementi che essi racchiudono, e delle relazioni che instaurano con altri oggetti. Per superare questo limite nasce l'idea di atomizzare le informazioni aggiungendo significato semantico ai singoli dati. Ciò permette alle macchine di processare il loro significato¹⁴¹.

Nonostante i vantaggi indiscutibili, ai dati non sono stati applicati gli stessi principi che hanno permesso al web dei documenti di fiorire. Infatti i dati pubblicati nel web sono stati resi disponibili in formati quali CSV o XML, o con linguaggi di marcatura come l'HTML, sacrificando gran parte della loro struttura e della loro semantica. In questo senso, L'HTML non è sufficientemente espressivo per consentire di individuare le entità descritte in un documento, e per collegare con link tipizzati le entità correlate¹⁴².

Il web dei dati è invece caratterizzato dalle seguenti componenti¹⁴³:

- le cose del mondo reale (*things*): ad esempio persone, concetti, oggetti, luoghi;
- *typed links*, link qualificati che rendono esplicito il tipo di relazione tra dati.
- i database, rappresentati come contenitori che conservano le informazioni.

¹³⁸ «W3C WebApps Working Group», <https://www.w3.org/2008/webapps/>.

¹³⁹ «Base di dati», *Wikipedia*, https://it.wikipedia.org/w/index.php?title=Base_di_dati&oldid=85497469.

¹⁴⁰ Guerrini e Possemato, *Linked data per biblioteche, archivi e musei*.

¹⁴¹ Ibid.

¹⁴² Christian Bizer, Tom Heath, e Tim Berners-Lee, «Linked Data - The Story So Far», *International Journal on Web semantico and Information Systems* 5, n. 3 (2009): 1–22, doi:10.4018/jswis.2009081901.

¹⁴³ Bizer et al., «Linked Data on the Web».

Nel web dei dati i collegamenti devono avere un preciso significato ed essere strutturati per essere utilizzati dalle macchine. *Semantico* dunque assume il significato di comprensibile dalla macchina¹⁴⁴.

L'evoluzione del web, teorizzata da Berners-Lee¹⁴⁵, presenta un web di dati composto da molti livelli, o meglio di metadati che esprimano il contenuto dei documenti presenti in rete. Essi stabiliscono delle interazioni tra dati a partire da una loro strutturazione formale e promuovono la condivisione tra comunità che decidono di aprire e connettere al network i propri dati, linked data¹⁴⁶.

Alla base del web semantico vi è l'utilizzo combinato di una serie di tecnologie, tanto che si parla di *Web semantic stack*, letteralmente pila del web semantico o *Web semantic layer cake*¹⁴⁷, torta a strati del web semantico.

I linguaggi più importanti, in relazione alla strutturazione dei linked data, sono XML, RDF, URI e le ontologie. Lo scopo è quello di rappresentare semanticamente un dominio conoscitivo, utilizzando una rappresentazione formale di questa astrazione che risulti comprensibile sia all'uomo che alla macchina¹⁴⁸. Nel paragrafo successivo verranno presentati, senza pretesa di esaustività, quelli che vengono considerati i pilastri del web semantico.

¹⁴⁴ Guerrini e Possemato, *Linked data per biblioteche, archivi e musei*.

¹⁴⁵ Berners-Lee, Hendler, e Lassila, «The Semantic Web. A new form of web content that is meaningful to computers will unleash a revolution of new possibilities».

¹⁴⁶ Berners-Lee, «Linked Data - Design Issues».

¹⁴⁷ Eric Gordon, «Semantic Web Specification at W3C - Slide list», 2004, <https://www.w3.org/2004/Talks/0412-RDF-functions/Overview.html>.

¹⁴⁸ Dean Allemang e James A. Hendler, *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*, 2nd ed (Waltham: Morgan Kaufmann, 2011).

2.2. Uno sguardo alle tecnologie del web semantico

Il web semantico non è un linguaggio o uno standard. Esso prevede la compresenza di diverse tecnologie. Di seguito si prenderanno in considerazione le più importanti.

2.2.1. XML

XML è un meta-linguaggio di marcatura che fornisce regole sintattiche per modellare la struttura di documenti e dati. Progettato per consentire una maggiore interoperabilità tra applicazioni e per essere auto-descrittivo, in quanto gli elementi di marcatura risultano comprensibili sia alla macchina che all'uomo¹⁴⁹. Questo linguaggio rappresenta in qualche modo un'evoluzione dell'HTML:

I limiti di HTML emergono quando si hanno di fronte applicazioni che presentano requisiti più stringenti, determinati dalla complessità delle strutture dati da gestire, dalla necessità di utilizzare gli stessi dati in maniera diversa e di gestire in modo adeguato dati con un ciclo di vita significativo. Extensible Markup Language (XML) è nato per far fronte alle nuove applicazioni Web, in cui i dati costituivano un elemento essenziale (data-centric Web applications). XML è stato quindi il primo passo per assegnare una semantica ai tag e supportare le transazioni sul Web, permettendo lo scambio di informazioni tra database diversi. Ulteriori e significativi vantaggi sono costituiti dalla possibilità di avere viste diverse degli stessi dati, e la possibilità di personalizzare le informazioni mediante opportuni agenti. L'adozione di XML agevola la gestione di collezioni di documenti, e costituisce un supporto fondamentale per la pubblicazione di informazioni a livello internazionale, con il non piccolo vantaggio di essere indipendente dalla piattaforma e dal linguaggio¹⁵⁰.

L'XML non solo garantisce un web più strutturato rispetto all'HTML ma consente alle comunità di utilizzatori di poter definire i propri elementi di *markup*, da cui l'aggettivo di *estensibile* nel nome. Questo meta-linguaggio, rispetto all'HTML, si focalizza sulla struttura del testo da rappresentare e sul contenuto dei suoi elementi, mentre l'HTML si occupa principalmente della formattazione e della visualizzazione.

Un documento XML si compone, essenzialmente, di tre elementi principali:

1. la DTD (Document Type Definition) o l'XML Schema (opzionale), un documento che permette di definire *ex novo* un insieme di tag che rispettino le esigenze di codifica di un documento o di un insieme

¹⁴⁹ «XML Essentials - W3C», <https://www.w3.org/standards/xml/core>.

¹⁵⁰ Oreste Signore, «Il ruolo centrale di XML nell'evoluzione del Web», (2001), 3, <http://www.w3c.it/papers/SignorePerXMLDays2001.pdf>.

di documenti specifici. In altre parole si realizza un modello dati che ricalca la struttura logica e le componenti semantiche dell'informazione trattata;

2. l'istanza di documento, ossia uno o più documenti testuali in cui risiede il contenuto informativo, e che, eventualmente, rispetta/no il modello dati implementato;

3. il foglio di stile XSL (XML Stylesheet Language) per la rappresentazione del documento e la sua visualizzazione¹⁵¹.

Poiché in un documento XML ognuno può creare i propri elementi ed attributi, si è reso necessario un meccanismo che identifichi quali di questi appartengono ad uno stesso schema. Ad esempio due documenti possono utilizzare l'elemento *titolo*, ma uno in riferimento al titolo di un libro e l'altro alle qualifiche di tipo professionale, come dottore o ingegnere. La DTD o l'XML Schema non forniscono però un meccanismo che consenta di riutilizzare gli stessi elementi in contesti diversi. A partire da questa necessità è stato sviluppato l'XML Namespaces, che consente di poter utilizzare nomi già definiti in altri schemi, evitando la presenza di conflitti e ambiguità¹⁵². I *namespace* sono insieme di nomi, identificati in modo univoco da un URI, che hanno lo scopo di qualificare gli elementi e gli attributi di uno stesso schema. Se in un documento si decide di riutilizzare l'elemento Dublin Core¹⁵³ *Creator:http://purl.org/dc/term*, quando un elemento sarà qualificato con *creator* esso dovrà essere interpretato secondo la definizione data nello schema di Dublin Core¹⁵⁴.

Da diversi anni XML viene utilizzato anche nell'ambito bibliotecario, per rappresentare documenti elettronici o metadati e per la condivisione di dati tra applicazione diverse¹⁵⁵. Riguardo i dati bibliografici XML è stato applicato come linguaggio di sintassi per rappresentare dati strutturati e "assicurare la interoperabilità tra applicazioni di rete distribuite per mezzo di formati di interscambio dei dati"¹⁵⁶. Nel 2002 la Library of Congress ha utilizzato uno schema XML per il formato MARC21, il MARCXML e reso

¹⁵¹ Ilaria Barbanti, «La codifica informatica dei dati di archivio: EAD ed EAC», *Scrinia*, n. 2–3 (2005): 3, http://media.regesta.com/dm_0/regesta/regestacms/generic/documentazione/Barbanti-EADEAC2005.pdf.

¹⁵² Valdo Pasqui, «Xml: una introduzione. Nozioni di base e possibili aree di sviluppo nel contesto bibliotecario», *Biblioteche oggi* 19, n. 6 (2001): 18–24.

¹⁵³ «Dublin Core Metadata Element Set, Version 1.1», <http://dublincore.org/documents/dces/>.

¹⁵⁴ Hooland e Verborgh, *Linked Data for Libraries, Archives and Museums*.

¹⁵⁵ Pasqui, «Xml: una introduzione. Nozioni di base e possibili aree di sviluppo nel contesto bibliotecario».

¹⁵⁶ *Ibid.*, 22.

disponibili una serie di strumenti per convertire i record MARC in MARCXML e viceversa¹⁵⁷. Anche per l'UNIMARC sono stati predisposti degli schemi XML, nel 2004 la società BookMARC, nell'ambito dell'attività dell'IFLA, ha prodotto il *Manuale UNIMARC/B* che comprende regole strutturali e di contenuto oltre che diverse note esplicative¹⁵⁸. TurboMARC¹⁵⁹ è un altro esempio di strumento che consente la conversione di record MARC in XML e che presenta un formato più facilmente processabile da XSLT (eXtensible Stylesheet Language), che si occupa della presentazione del contenuto dei record.

La semantica dei dati viene dunque esplicitata attraverso l'uso degli elementi XML. Questo linguaggio però richiede comunque che chi utilizza gli elementi conosca lo schema e la documentazione che ne illustra il significato. Avvalersi dei *namespace* garantisce la possibilità di condividere uno stesso elemento tra diverse comunità ma è una pratica limitata ad un numero esiguo di elementi¹⁶⁰. Tali inconvenienti hanno in qualche modo indirizzato la nascita di ulteriori linguaggi diretti a standardizzare sempre di più il processo di strutturazione e condivisione dei dati nel web.

2.2.2. RDF

L'XML fornisce una sintassi di basso livello per rappresentare i dati e consentire una migliore transazione sul web¹⁶¹, cioè consente di poterli rappresentare attraverso una struttura ad albero e di aggiungere tag di natura semantica, ma non fornisce nessuna indicazione sul come esprimere la semantica dei tag. RDF sopperisce a questa necessità, fornendo un modello condiviso che rappresenta in qualche modo la grammatica della lingua del web semantico¹⁶². RDF si presenta come un modello astratto, proposto dal

¹⁵⁷ Maria Cristina Bassi, «Formati MARC-XML e servizi derivati: resoconto sulla sessione al congresso IFLA 2005», *Bibliotime* VIII, n. 3 (2005), <http://www.aib.it/aib/sezioni/emr/bibttime/num-viii-3/bassi.htm>.

¹⁵⁸ Maria Cristina Bassi, «Formati MARC-XML e servizi derivati: resoconto sulla sessione al congresso IFLA 2005»

¹⁵⁹ «Turbomarc, faster XML for MARC records | Index Data», <http://www.indexdata.com/news/2014/06/turbomarc-faster-xml-marc-records>.

¹⁶⁰ Hooland e Verborgh, *Linked Data for Libraries, Archives and Museums*.

¹⁶¹ Signore, «Il ruolo centrale di XML nell'evoluzione del Web», (2001), <http://www.w3c.it/papers/SignorePerXMLDays2001.pdf>.

¹⁶² Thomas Baker, «Designing Data For the Open World of the Web», *JLIS.it* 4, n. 1 (2013): 63-6, doi 10.4403/jlis.it-6308.

W3C, che non esplicita il significato dei metadati, ma fornisce un modello per poterlo esprimere, configurandosi come “una lingua franca, capace di mediare e far dialogare differenti modelli di dati utilizzati nel web”¹⁶³. Lo sviluppo di RDF è stato motivato dalla possibilità di implementare i seguenti usi¹⁶⁴:

- fornire informazioni sulle risorse web e sui sistemi che le usano;
- consentire lo sviluppo di applicazioni che necessitano di dati aperti piuttosto che di modelli che vincolano le informazioni;
- dare alle macchine informazioni processabili, come che il web ha fatto per gli ipertesti.;
- permettere alle applicazioni di combinare dati che provengono da diverse applicazioni per creare nuove informazioni. Il web semantico si propone di rendere i dati processabili anche in contesti differenti rispetto a quelli di origine;
- consentire ad agenti software¹⁶⁵ di applicare processi automatici nell’elaborazione dell’informazione. Il web sta cominciando ad avere informazioni “comprensibili” non solo dagli esseri umani ma anche dalle macchine, RDF consente di avere un linguaggio comune per questi processi di comprensione semantica dell’informazione.

RDF è stato dunque progettato per rappresentare le informazioni nel modo più flessibile possibile. Questo modello può essere utilizzato anche in applicazioni isolate ma il valore principale è la possibilità di condividere informazioni, rendendole accessibili alle diverse applicazioni esistenti nel web.

L’RDF si basa su due componenti:

- RDF Model and Syntax che espone la struttura fondamentale del modello RDF e la sua codifica in XML;

¹⁶³ Guerrini e Possemato, *Linked data per biblioteche, archivi e musei*, 47.

¹⁶⁴ «RDF - Semantic Web Standards».

¹⁶⁵ “Gli agenti software sono programmi che dovrebbero sfruttare la conoscenza contenuta nei metadati per fornire servizi complessi agli utenti del Semantic Web. Per far ciò, gli agenti devono essere dotati di un buon livello di autonomia (non possono rivolgersi continuamente all’utente per decidere cosa fare) e di capacità di ragionare in modo complesso sull’informazione a cui hanno accesso. Questo significa al minimo essere in grado di rappresentare gli obiettivi di un certo utente, di mettere in atto una sequenza di azioni che possa soddisfare tali obiettivi, ed eventualmente di cooperare con altri agenti per ottenere tale risultato”. Paolo Bouquet e Roberta Ferrario, «Il Semantic Web», *Networks. Rivista di filosofia dell’intelligenza artificiale e scienze cognitive*, n. 2 (2003): iv.

- RDF Schema che presenta la sintassi per definire schemi e vocabolari per i metadati, ovvero per le informazioni sui dati.

L’RDF Model and Syntax viene presentato in una serie di documenti pubblicati dal W3C¹⁶⁶, esso si basa su tre concetti¹⁶⁷:

- risorse (*resource*): tutto ciò che viene descritto in RDF viene chiamato risorsa, ogni risorsa viene identificata da un URI;
- proprietà (*property*): è un attributo, una relazione, una caratteristica o un aspetto specifico che viene associato alla risorsa;
- asserzioni (*statement*): l’associazione di una proprietà ad una risorsa, un’asserzione produce una tripla, formata da un soggetto (la risorsa), un predicato (la proprietà) e un oggetto (il valore). L’oggetto può essere rappresentato da un’altra risorsa o da un *literal*, se si tratta di una risorsa essa sarà identificata da un URI mentre se si tratta di un *literal* si avrà una semplice espressione, una sequenza di caratteri.

L’RDF esprime dunque la sua struttura basilare in triple, ognuna formata da *soggetto*, *predicato*, *oggetto*; l’insieme degli *statement* viene denominato *grafo RDF*¹⁶⁸. La frase “questo libro ha per titolo Pian della Tortilla”, in RDF verrebbe scomposta nei seguenti tre elementi:

soggetto: questo libro

predicato: ha per titolo

oggetto: Pian della Tortilla

Questo linguaggio si configura come una sovrastruttura che serve per descrivere risorse e per codificare e scambiare metadati nell’ambito del web semantico¹⁶⁹. Esso permette di definire relazioni tra entità, dichiarandole come link qualificati. L’RDF quindi stabilisce in che modo i singoli elementi descrittivi devono essere composti per creare delle “proposizioni”, secondo la modalità soggetto-predicato-oggetto. L’XML ad un livello più

¹⁶⁶ «Resource Description Framework (RDF) Model and Syntax Specification», <https://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>; «RDF 1.1 XML Syntax», <https://www.w3.org/TR/rdf-syntax-grammar/>.

¹⁶⁷ Oreste Signore, «RDF per la rappresentazione della conoscenza», (2002), <http://www.w3c.it/papers/RDF.pdf>.

¹⁶⁸ «RDF - Semantic Web Standards».

¹⁶⁹ Signore, «RDF per la rappresentazione della conoscenza».

alto fornisce invece una griglia che struttura l'insieme di queste proposizioni, cioè dei grafi RDF e li presenta secondo una visualizzazione gerarchica ad albero.

Essendo un modello di dati, RDF non permette di utilizzare direttamente nodi e grafi, che devono essere espressi in un formato processabile da parte delle macchine. A questo scopo il W3C ha indicato come standard due formati di serializzazione¹⁷⁰: RDF/XML e RDFa (Resource Description Framework in Attributes). RDF/XML definisce l'XML come sintassi per i grafi RDF¹⁷¹. RDFa invece è un formato di serializzazione che permette di inserire informazioni semantiche in pagine HTML, arricchendole con elementi di marcatura¹⁷².

RDF consente semplicemente di creare una struttura a grafo (soggetto – predicato – oggetto, ovvero risorsa – proprietà – valore)¹⁷³ ma non specifica come dichiarare queste proprietà, né descrive la semantica. RDF Schema risponde a questa necessità in quanto fornisce alcune linee guida sull'utilizzo formalizzato di una struttura a grafo e offre una modalità per costruire i vocabolari che possono essere utilizzati nelle triple RDF¹⁷⁴. Come afferma Oreste Signore:

RDF Schema permette di definire significato, caratteristiche e relazioni di un insieme di proprietà, compresi eventuali vincoli sul dominio e sui valori delle singole proprietà. Inoltre, implementando il concetto (transitivo) di classe e sottoclasse, consente di definire gerarchie di classi, con il conseguente vantaggio che agenti software intelligenti possono utilizzare queste relazioni per svolgere i loro compiti.¹⁷⁵

Le triple RDF però non sono composte dalle parole che utilizziamo nel linguaggio naturale ma da URI. Gli indirizzi URI sono simili agli indirizzi che troviamo nel web, URL (Uniform Resource Locator), ma nel web semantico essi vengono utilizzati per identificare cose (*thing*) come le persone, i luoghi, le epoche storiche e anche i concetti

¹⁷⁰ La serializzazione consente di avere delle rappresentazioni testuali adatte all'elaborazione automatica. Esistono diverse serializzazioni di RDF, questo procedimento prende i set di triple e crea un file che esprime i contenuti in un formato leggibile dalla macchina. I formati più conosciuti, oltre a quelli raccomandati dal W3C, sono Turtle, N-Triples e RDF/JSON.

¹⁷¹ «RDF/XML Syntax Specification (Revised)», <https://www.w3.org/TR/REC-rdf-syntax/>.

¹⁷² «RDFa 1.1 Primer - Third Edition», <https://www.w3.org/TR/xhtml-rdfa-primer/#the-basics-of-rdfa-lite>.

¹⁷³ Allemang e Hendler, *Semantic Web for the Working Ontologist*.

¹⁷⁴ Ibid.

¹⁷⁵ Signore, «RDF per la rappresentazione della conoscenza», 3.

astratti¹⁷⁶. Le singole parole che vanno a costituire una frase RDF sono così composte da identificativi univoci. In linea di principio tutto può essere rappresentato da URI, sia i soggetti, che i predicati e gli oggetti.

2.2.3. URI

Gli URI servono per nominare i concetti e le relazioni, sono unici al livello globale e consentono di creare reticoli significativi tra gli elementi del web.¹⁷⁷

L'URI rappresenta l'innovazione più importante del web “perché è l'unica specifica che ogni programma, client o server della rete ovunque si trovi usa quando segue un link”¹⁷⁸. Questi identificatori servono a rendere reperibili gli elementi nel web, seguono una sintassi uniforme e sono estensibili¹⁷⁹. Il grande successo del web dipende in larga parte proprio dall'utilizzo di URI. Grazie a questi identificatori il WWW è riuscito a individuare risorse accessibili tramite il suo protocollo, HTTP (HyperText Transfer Protocol)¹⁸⁰, e anche altri sistemi di gestione informatica dei documenti. Questo meccanismo ha consentito quindi di individuare un linguaggio universale, facilmente memorizzabile o interscambiabile per identificare le risorse nella rete.

Esistono due tipologie di URI¹⁸¹:

- URL (Uniform Resource Locator): un indirizzo che identifica una risorsa tramite la sua localizzazione, e che quindi contiene le informazioni per accedere alla risorsa;
- URN (Uniform Resource Name): una sintassi che permette di identificare in modo permanente la risorsa indipendentemente dal riportare informazioni

¹⁷⁶ Baker, «Designing Data for the Open World of the Web».

¹⁷⁷ Antonella Trombone, «Cataloghi, modelli concettuali, data model: gli orientamenti della ricerca e gli ordinamenti tematici nella Library and information science», *Bibliothecae.it* V, n. 1 (2016): 95–129, doi:10.6092/issn.2283-9364/6108.

¹⁷⁸ Berners-Lee e Fischetti, *L'architettura del nuovo Web*, 46.

¹⁷⁹ «URIs, URls, and URNs: Clarifications and Recommendations 1.0», <https://www.w3.org/TR/uri-clarification/>.

¹⁸⁰ L'HTTP è un protocollo che serve a trasmettere le informazioni sul web, le sue specifiche sono gestite dal W3C. «HTTP - Hypertext Transfer Protocol Overview», <https://www.w3.org/Protocols/>.

¹⁸¹ «URIs, URls, and URNs: Clarifications and Recommendations 1.0».

sull'accesso, ad esempio l'ISBN (International Standard Bibliographic Number) di un libro.

In pratica l'URN identifica l'oggetto mentre l'URL localizza il luogo in cui esso è contenuto, va detto che un URN può anche essere un URL, cioè identificare un oggetto e allo stesso tempo fornire informazioni sulla localizzazione¹⁸². Ad esempio il DOI (Digital Object Identifier) <http://dx.doi.org/10.4403/jlis.it-30> identifica l'oggetto in modo univoco, cioè l'articolo di Maria Cassella pubblicato su Jlis nel 2010, e allo stesso tempo ne permette il recupero e l'accesso.

I metadati, nella visione del web semantico, possono diventare *machine-understandable* attraverso l'uso combinato di XML, RDF e URI. I dati vengono atomizzati con RDF, che li esprime in forma di triple. Affinchè queste siano processabili dalle macchine la loro sintassi viene espressa con XML. L'URI infine “con la sua funzione di nome globalmente unico per una determinata “cosa” [...] serve ad ancorare i fili delle reti del significato (web semantico). Ogni filo della rete è una frase grammaticale. Ogni frase esprime un'idea semplice”¹⁸³. Ciò che manca però è la semantica dei metadati, ovvero fornire strumenti che consentano di definire il significato dei termini che vengono associati ai dati delle triple. Il web semantico prevede infatti l'utilizzo di ontologie e vocabolari che forniscano un “sostrato concettuale che leghi i termini tra loro, mostrandone le relazioni e facendone quindi emergere il significato”¹⁸⁴.

2.2.4. Ontologie e vocabolari

Non c'è una distinzione chiara tra i termini “vocabolario” e “ontologia” a livello di web semantico. La tendenza che si è imposta è quella di utilizzare *ontologia* per sistemi di organizzazione della conoscenza più complessi e con una struttura formale più rigorosa, *vocabolario* per un insieme di termini a cui non viene applicato questo formalismo in senso stretto¹⁸⁵. Ontologie e vocabolari derivano dalla logica descrittiva ed offrono la possibilità di rappresentare in modo più dettagliato un dominio conoscitivo rispetto a RDF

¹⁸² Antonella De Robbio, «URI, URN e URL, una questione di definizioni: universal versus uniform», *Biblioteche oggi*, n. 1 (2002): 30–7, <http://www.bibliotecheoggi.it/2002/20020103001.pdf>.

¹⁸³ Baker, «Designing Data for the Open World of the Web»: 66.

¹⁸⁴ Bouquet e Ferrario, «Il Semantic Web»: iv.

¹⁸⁵ «Ontologies - W3C», <https://www.w3.org/standards/semanticweb/ontology>.

Schema. Gli standard promossi dal W3C come modelli di dati per il web semantico sono OWL (Web Ontology Language) e SKOS (Simple Knowledge Organisation).

In particolare i linguaggi formali caratterizzati da una maggiore espressività sono nati intorno al 2000. Fu l'agenzia del Dipartimento della difesa degli Stati Uniti ARPA (Defense Advance Research Projects Agency) a sviluppare il linguaggio DAML (Darpa Agent Markup Language Program) e in un secondo momento OIL (Ontology Inference Layer), grazie ad una cooperazione con l'Europa¹⁸⁶. Il gruppo di lavoro sulle ontologie del W3C, il Web Ontology Working Group¹⁸⁷, realizzò infine una revisione di DAML ed OIL proponendo OWL, conosciuto per essere un linguaggio molto complesso.

OWL consente di associare, in modo più analitico rispetto a RDF Schema, classi e proprietà alle regole del ragionamento formale, che rafforza i vincoli tra classi e produce nuova conoscenza grazie alle inferenze¹⁸⁸. Un predicato che rappresenta la relazione *has written*, potrebbe essere espresso in un'ontologia nel modo seguente¹⁸⁹:

@prefix ex: <<http://example.org/ontology#>>.

@prefix rdf: <<http://www.w3.org/1999/02/22-rdf-syntax-ns#>>.

@prefix rdfs: <<http://www.w3.org/2000/01/rdf-schema#>>.

Ex:hasWritten rdf:type rdf:Property;

rdfs:domain ex:Person;

rdfs:range ex:LiteraryWork;

rdfs:subPropertyOf ex:HasAuthored.

La proprietà *hasWritten* ha come dominio (*domain*) *Person* e come valore (*range*) *LiteraryWork*, ciò significa che i soggetti che avranno la proprietà *hasWritten* saranno sempre persone, cioè verranno riconosciuti come appartenenti alla classe “persone”, mentre gli oggetti saranno sempre opere letterarie. A partire quindi dalle definizioni delle classi e delle relazioni è possibile creare dei vincoli sui significati dei termini che vengono metadati. I prefissi servono a distinguere *namespace* appartenenti a standard diversi,

¹⁸⁶ Biagetti, «Le ontologie come strumenti per l'organizzazione della conoscenza in rete», *AIDA Informazioni*, n. 1–2 (2010): 9–32.

¹⁸⁷ «W3C Web Ontology (WebOnt) Working Group (OWL) (Closed)», <https://www.w3.org/2001/sw/WebOnt/>.

¹⁸⁸ Hooland e Verborgh, *Linked Data for Libraries, Archives and Museums*.

¹⁸⁹ *Ibid.*, 126.

nell'esempio *ex* rappresenta un *namespace* creato ad hoc, mentre RDF Schema e RDF fanno riferimento agli standard del W3C.

Per comprendere come funzionano le inferenze supponiamo che un computer elabori la seguente tripla:

:HermanMelville ex:hasWritten :MobyDick.

Attraverso l'URI della proprietà *hasWritten* il computer risalirà all'ontologia di riferimento dell'esempio, la quale contiene maggiori informazioni leggibili dalla macchina riguardo questa proprietà e riuscirà ad inferire la seguente tripla in modo automatico.

:HermanMelville a ex:Person.

:MobyDick a ex:LiteraryWork.

:HermanMelville ex:hasAuthored :MobyDick.

Poiché è stato indicato a monte che la proprietà *hasWritten* ha come dominio le persone e come oggetto un elemento appartenente a *LiteraryWork*, il computer arriverà ad inferire che Herman Melville è una persona e che *Moby Dick* è un'opera letteraria. Inoltre se nell'ontologia di riferimento il predicato *ex:hasWritten* viene indicato come una specifica di *ex:hasAuthored* esso eredita la stessa classe e gli stessi valori. Il computer ovviamente non comprende che *Moby Dick* è un'opera letteraria nello stesso modo in cui lo comprendono gli esseri umani, ma l'aspetto importante è che la macchina è in grado di interpretare la relazione *has written* per arrivare ad inferire nuova conoscenza¹⁹⁰. Questo esempio illustra le potenzialità delle ontologie e l'idea di ragionamento che sottostà al web semantico.

SKOS (Simple Knowledge Organisation System)¹⁹¹ invece viene utilizzato per la rappresentazione di vocabolari riguardanti i concetti. Quest'ontologia è basata su RDF Schema e serve a supportare l'uso dei sistemi di organizzazione della conoscenza, come i thesauri, le classificazioni, le tassonomie, e in generale i vocabolari controllati. SKOS si basa su un modello che separa i *concetti* dalle *etichette*, cioè dai termini che li esprimono, seguendo quanto espresso in FRSAD (Functional Requirements for Subject

¹⁹⁰ Hooland e Verborgh, *Linked Data for Libraries, Archives and Museums*.

¹⁹¹ «SKOS Simple Knowledge Organization System», <https://www.w3.org/2004/02/skos/>.

Authority Data)¹⁹². Esso è un modello concettuale, sviluppatosi in seguito a FRBR e focalizzato sui dati di autorità di natura semantica, che vede la separazione tra *thema* e *nomen*. Questo approccio risulta chiaro se si osserva la strutturazione dei termini presenti nel Thesaurus del Nuovo Soggettario, il termine “arte”, ad esempio, presenta delle equivalenze con il LCSH (Library of Congress Subject Headings) e con RAMEAU (Répertoire d'autorité-matière encyclopédique et alphabétique unifié) il thesaurus francese della Biblioteca nazionale di Francia. Definito un concetto, in questo caso quello di arte, esso può essere rappresentato da termini in diverse lingue, cioè da vari *nomen* per usare la definizione di FRSAD, le cui equivalenze vengono stabilite grazie alla relazione di SKOS *closeMatch*¹⁹³. Si crea così un reticolo di dati di autorità espressi anche in diverse lingue, collegati e navigabili.

Le ontologie rivestono una grande importanza all'interno del web semantico, non solo perché sono gli strumenti che consentono di rappresentare un dominio conoscitivo, esprimendone i termini e le relazioni, ma anche perché i mutamenti di approccio verso le ontologie hanno segnato due fasi differenti relative al web semantico. In una prima fase, compresa tra il 1999 e il 2006, predomina una visione *top-down* “forte”, detta anche dell'*ontologia fondazionale*, in cui Tim Berners-Lee aveva previsto l'utilizzo di ontologie interdisciplinari e generali¹⁹⁴, associate alle teorie dell'unificazione della conoscenza, che postulano la possibilità di poter rappresentare universalmente la conoscenza a priori¹⁹⁵. Questo approccio dall'alto si è però scontrato con la natura arbitraria e incontrollabile dei costrutti umani, evidenziando tutti i suoi limiti

nella difficoltà di raggiungere uno standard ontologico riconosciuto universalmente e definito a priori poiché il Web è un ambiente dinamico e in costante evoluzione e perché risultava difficile esprimere in un linguaggio formale la semantica naturale soggetta a diverse convenzioni sociali e individuali¹⁹⁶.

¹⁹² «IFLA -- Functional Requirements for Subject Authority Data», <http://www.ifla.org/node/5849>.

¹⁹³ Giovanni Bergamin e Anna Lucarelli, «The Nuovo Soggettario as a Service for the Linked Data World», *JLIS.it* 4, n. 1 (2012): 213, doi.org/10.4403/jlis.it-5474.

¹⁹⁴ Valeria Lo Castro, «Web Semantico e Linked Open Data: best practices, prospettive e criticità», *Nuovi Annali della Scuola Speciale per Archivisti e Bibliotecari XXVIII* (2014): 207–22.

¹⁹⁵ Paola Di Maio, «Linked Data Beyond Libraries: Towards Universal Interfaces and Knowledge Unification», in *Linked Data and User Interaction: the Road Ahead*, a c. di H. Frank Cervone e Lars G. Svensson, (Berlino ; Boston: De Gruyter Saur, 2015), 12–32.

¹⁹⁶ Lo Castro, «Web Semantico e Linked Open Data: best practices, prospettive e criticità»: 211.

Nella seconda fase, dal 2006 ad oggi, l'approccio al web semantico subisce una revisione¹⁹⁷, Tim Berners-Lee propone un approccio *bottom-up* che prevede il coinvolgimento delle varie comunità del web nella costruzione delle ontologie, secondo lo spirito collaborativo del web 2.0. In una serie di famosi interventi ed articoli, l'inventore del WWW propone i linked data come alternativa ad un controllo centralizzato per la rappresentazione delle informazioni e come modalità per consentire uno scambio di informazioni flessibile ed aperto. La visione di un mondo di informazioni collegate tramite dati di significato univoco permette che le macchine possano inferire conclusioni logiche. Ma questo funziona bene su sistemi chiusi e molto specifici a cui è possibile applicare regole logiche stringenti. Questo approccio non ha però funzionato in un contesto aperto e flessibile come quello del web. I linguaggi logici basati su regole non riescono infatti a lavorare con l'imprevedibilità, con l'imperfezione che connatura i dati presenti in uno spazio globale e universale, per questo su larga scala l'iniziale idea del web semantico non si è realizzata. I linked data, che sono un insieme di pratiche per collegare e pubblicare dati nel web, si presentano come una soluzione affinché il web evolva da web di documenti a web di dati, un luogo globale di dati connessi e provenienti da diversi domini.

2.3. I principi dei linked data

Tim Berners-Lee ha presentato un insieme di *best practices* per pubblicare i dati nel web in modo che essi possano diventare parte di un "single global data space"¹⁹⁸. Queste tecniche sono conosciute come I Principi dei linked data¹⁹⁹:

1. usare URI per identificare cose (*thing*);
2. usare HTTP URI in modo che le cose possano essere individuate da persone o da computer;
3. fornire informazioni sulle cose quando si individua un URI, usando RDF come linguaggio di formalizzazione dei dati e SPARQL per l'interrogazione;

¹⁹⁷ N. Shadbolt, T. Berners-Lee, e W. Hall, «The Semantic Web Revisited», *IEEE Intelligent Systems* 21, n. 3 (2006): 96–101, doi:10.1109/MIS.2006.62.

¹⁹⁸ Bizer, Heath, e Berners-Lee, «Linked Data - The Story So Far»: 2.

¹⁹⁹ Berners-Lee, «Linked Data - Design Issues».

4. includere link ad altri URI per facilitare la scoperta di altre cose.

Questi principi forniscono indicazioni di base per pubblicare e collegare i dati utilizzando l'infrastruttura del web e utilizzando la sua architettura e i suoi standard.

Il primo principio si basa sul meccanismo degli URI. Come visto precedentemente, nel web semantico tutto è identificato da URI, questo permette che le risorse siano sempre rintracciabili e riusabili sia da utenti che da applicazioni.

Il secondo principio sostiene l'uso di HTTP URI per identificare le risorse. Il protocollo HTTP è un meccanismo universale di accesso al web. Nel web dei documenti gli HTTP URI sono usati per combinare un unico sistema di identificazione con un semplice meccanismo di reperimento delle risorse.

Il terzo principio prevede che vengano fornite informazioni sulle risorse. Questo è garantito dall'uso di URI *dereferenziabili*, che presentano una descrizione della risorsa attraverso il suo URL²⁰⁰. L'URI del documento contenente la risorsa e l'URI della risorsa contenuta devono quindi essere diversi e non confusi. L'URI dereferenziabile si può costruire fornendo una rappresentazione della risorsa sia in formato HTML per l'essere umano, che in RDF per la macchina. Inoltre viene consigliato l'uso di standard, come RDF e SPARQL. RDF permette di risolvere uno dei principali problemi di questo ambito: gestire dati distribuiti in diversi ambienti web e garantire le basi per la creazione di un network di dati. SPARQL (Simple Protocol And RDF Query Language) invece è un linguaggio che permette di interrogare diversi dataset rappresentati in RDF o di navigare tra di essi²⁰¹. La sua sintassi è simile a SQL, usato per interrogare basi di dati relazionali. SPARQL si aggiunge ai vari tasselli necessari per la realizzazione del web semantico poiché consente di ricavare informazioni da differenti dataset distribuiti nel web. Mentre SPARQL è il linguaggio di interrogazione di RDF, gli SPARQL Endpoint sono i servizi che accettano *query* e restituiscono risposte utilizzando questo linguaggio. Una buona pratica è infatti quella di fornire URL agli SPARQL Endpoint dei dataset creati in RDF, per consentire l'accesso, l'analisi e la navigazione a questi stessi dati²⁰².

²⁰⁰ Hooland e Verborgh, *Linked Data for Libraries, Archives and Museums*.

²⁰¹ «SPARQL Query Language for RDF», <https://www.w3.org/TR/rdf-sparql-query/>.

²⁰² «SparqlEndpoints - W3C Wiki», <https://www.w3.org/wiki/SparqlEndpoints>.

L'ultimo principio prevede l'inclusione di link ad altri URI, in modo da facilitare la scoperta di nuove informazioni. La creazione di link RDF verso dataset esterni consente di ottenere un singolo grafo globale di dati interconnessi, alla stregua di ciò che avviene nel web classico, dove tutti i documenti HTML sono connessi in un unico spazio globale. Le risorse possono provenire da fonti diverse, creando in questo modo un reticolo di connessioni che permettono di passare da un dataset interno ad uno esterno. Si scoprono così nuove informazioni²⁰³, secondo la logica *follow your nose*²⁰⁴, che viene ben rappresentata dall'idea di surfare nel web lasciandosi in qualche modo guidare dal proprio istinto di esplorazione. Il metodo può essere applicato sia dagli esseri umani che dagli agenti software.

2.3. Dai linked data ai linked open data

Il paradigma dei linked data viene sempre più spesso associato al movimento *Open data*, tanto che si parla sempre più di *linked open data*²⁰⁵. Gli *open data* possono essere utilizzati da chiunque senza però presentare connessioni tra i dataset. I linked data invece presuppongono la creazione di link tra i dati ma senza necessariamente essere disponibili per il riutilizzo e la distribuzione. I *linked open data* sono una potente miscela di *linked data* e *open data* con riferimento alla possibilità di avere dati correlati, interoperabili e anche aperti, cioè privi di restrizioni d'uso o di licenze commerciali²⁰⁶. Come scrivono Antonella De Robbio e Silvia Giacomazzi

è necessario parlare tutti la stessa lingua per creare la collaborazione, lo scambio, il riutilizzo. Non possiamo fermarci ai dati aperti, ma dobbiamo anche concorrere per il loro collegamento in modo interoperabile²⁰⁷.

Il progetto Linking Open Data del W3C nasce proprio con lo scopo di pubblicare i vari *open dataset* come RDF e di potenziare l'estensione del web attraverso la creazione di

²⁰³ Guerrini e Possemato, *Linked data per biblioteche, archivi e musei*.

²⁰⁴ Liyang Yu, «Follow Your Nose: A Basic Semantic Web Agent», in *A Developer's Guide to the Semantic Web*, di Liyang Yu (Berlin, Heidelberg: Springer Berlin Heidelberg, 2014), 711–36, http://link.springer.com/10.1007/978-3-662-43796-4_16.

²⁰⁵ «Linked Open Data - W3C eGovernment Wiki», https://www.w3.org/egov/wiki/Linked_Open_Data.

²⁰⁶ Lo Castro, «Web Semantico e Linked Open Data: best practices, prospettive e criticità».

²⁰⁷ De Robbio e Giacomazzi, «Dati aperti con LODe».

link RDF tra i dati che provengono da diverse fonti²⁰⁸. La *Linking data cloud*, la “nuvola” che raccoglie gli *open dataset* messi a disposizione nella rete, negli anni si è enormemente arricchita. Dai 295 dataset del 2011 si è passati ai 1014 del 2014²⁰⁹. Il *Linking data cloud diagram* è stato recentemente aggiornato al 2017, sono visibili anche i dataset di ambito bibliotecario, come il MARC Codes List, il Library of Congress Subject Headings, la British National Bibliography, Europeana Linked Open Data, data.bnf.fr, RAMEAU, la Dewey Decimal Classification, Sudoc Bibliographic Data, DBLP Bibliography Database e molti altri²¹⁰.

Ma che vantaggi può portare l’applicazione dei linked open data ai dati delle biblioteche? E perché le biblioteche si sono avvicinate non solo a queste tecnologie ma anche ai valori di questo movimento?

²⁰⁸ «Linked Data | Linked Data - Connect Distributed Data across the Web», <http://linkeddata.org/>.

²⁰⁹ «The Linking Open Data cloud diagram», <http://lod-cloud.net/>.

²¹⁰ Ibid.

CAPITOLO TERZO

I linked open data e l'universo bibliografico

I linked open data stanno assumendo un ruolo centrale nelle modalità con cui le istituzioni culturali organizzano e condividono i loro dati nel web. Le iniziative LOD stanno contribuendo a creare un ambiente aperto e unificato per le informazioni. Questo potenziale può essere utilizzato per risolvere alcune problematiche che da sempre si pongono nel settore delle istituzioni culturali, come ad esempio l'interoperabilità e l'integrazione tra dati e la valorizzazione delle informazioni in un contesto più ampio dei singoli siti o portali.

Il cosiddetto approccio *walled garden*²¹¹ alla gestione dei dati nelle biblioteche, negli archivi e nei musei, così come in altre organizzazioni culturali, ha spesso limitato l'accesso e l'utilizzo delle loro risorse. I linked open data invece, basandosi sui principi di apertura, offrono la possibilità di esporre i propri dati, generalmente contenuti nei database, favorendo e migliorando la collaborazione tra gli istituti culturali anche di ambiti diversi²¹².

Negli ultimi anni molti sono stati i dibattiti, le proposte e le iniziative in ambito biblioteconomico volte ad analizzare la tecnologia dei LOD, come possibilità di miglioramento della gestione dei propri dati.

Ma perché le biblioteche dovrebbero interessarsi ai linked data? La maggior parte degli utenti inizia a ricercare informazioni a partire dal web e milioni di dati bibliografici appartenenti a biblioteche di tutto il mondo rimangono invisibili ai motori di ricerca²¹³.

Carlo Bianchini sostiene che:

i motivi di questa situazione sono molti, ma si possono ricondurre a una ragione fondamentale: l'automazione delle biblioteche e la nascita del web sono stati due eventi contemporanei, paralleli ma fino a oggi estranei l'uno all'altro. Al centro della crisi del ruolo delle biblioteche – anche digitali – e

²¹¹ Letteralmente "giardino recintato", l'espressione viene utilizzata per indicare ambienti chiusi in cui i sistemi software sono sottoposti a controlli e restrizioni riguardo i contenuti, i media, le applicazioni e gli accessi.

²¹² M. Cristina Pattuelli e Hilary Thorsen, «Linked Open Data and the Cultural Heritage Landscape», in *Linked Data for Cultural Heritage.*, a c. di Ed Jones e Michele Seikel (London: Facet, 2016), 1.

²¹³ Carlo Bianchini, «Dagli OPAC ai library linked data: come cambiano le risposte ai bisogni degli utenti», *AIB studi* 52, n. 3 (2012), <http://aibstudi.aib.it/article/view/8597>.

dei bibliotecari, c'è la constatazione che i lettori non percepiscono più il catalogo come il punto di partenza delle loro ricerche: i lettori sono e rimangono altrove, sul web, e cercano nel web. Le nostre biblioteche, anche le biblioteche 2.0, compaiono sul web ma non sono nel web.²¹⁴

Se i dati bibliografici non diventano parte del web, gli utenti che non li trovano con una prima semplice ricerca, molto probabilmente non sapranno mai che esistono. In realtà tutti gli ambiti che hanno strutturato i loro dati con formati specifici (giuridico, amministrativo, statistico, scientifico etc.) non vengono processati dalla rete, perché il loro contenuto non è comprensibile ai motori di ricerca.

Uno dei primi studi sulla possibilità di utilizzare i linked data in ambito bibliotecario è stato quello del Library Linked Data Incubator Group²¹⁵. Il Gruppo di lavoro partendo da alcuni casi di studio ha cercato di capire come si possono utilizzare gli standard del web semantico e i principi dei linked data per valorizzare e migliorare l'interoperabilità dei dati, creati, gestiti e curati dalle istituzioni culturali.

²¹⁴ Bianchini, «Library Linked Data e il futuro delle biblioteche», 2012, <https://sites.google.com/site/homepagecarlobianchini/Ricerca/library-linked-data-e-il-futuro-delle-biblioteche>.

²¹⁵ «W3C Library Linked Data Incubator Group», <https://www.w3.org/2005/Incubator/ld/>.

3.1. Il Library Linked Data Incubator Group

Nel 2010 il W3C ha incaricato un gruppo di lavoro, la cui missione dichiarata è

to help increase global interoperability of library data on the web, by bringing together people involved in semantic web activities – focusing on linked data – in the library community and beyond, building on existing initiatives, and identifying collaboration tracks to the future²¹⁶.

Il Gruppo di lavoro, composto da Thomas Baker, Karen Coyle, Gordon Dunsire, Antoine Isaac, Michael Panzer, Ed Summer e Marcia Zeng, pubblica nel 2011 il *Final Report*. Esso contiene delle linee guida per le istituzioni che decidono di avvicinarsi all'utilizzo di queste tecnologie, focalizzando l'attenzione sui bisogni specifici della comunità bibliotecaria.

Il Report spiega che con *library data* si fa riferimento a qualunque tipo di informazione digitale, prodotta dalle biblioteche, che descrive le risorse o facilita la loro scoperta. Nel documento vengono distinte tre tipologie di dati bibliografici²¹⁷, a partire dai principali usi che se ne fanno:

- *dataset*: l'insieme di dati che rappresenta le risorse di una biblioteca, o rete di biblioteche, come ad esempio il catalogo della Biblioteca Nazionale Ungherese o la British National Bibliography;
- *element set*: i metadati che descrivono le risorse, come gli elementi di RDA, SKOS o DCMI Metadata Terms;
- *value vocabularies*: i termini controllati, come quelli del Thesaurus del Nuovo Soggettario, LCSH (Library of Congress Subject Headings), la DDC (Dewey Decimal Classification), VIAF (Virtual International Authority File) o GeoNames.

²¹⁶ Thomas Baker et al., «Library Linked Data Incubator Group Final Report» (W3C Incubator Group Report, 2011), <https://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/>.

²¹⁷ Ibid.

Con linked data ci si riferisce dunque a dati pubblicati nel rispetto dei principi stabiliti per facilitare la creazione di connessioni tra dataset, *element set* e *value vocabularies* (termini controllati).

Nel Report si sostiene inoltre la differenza tra *open data* e *linked data*, come abbiamo già visto, mentre il primo termine si riferisce alla possibilità usare, riutilizzare e distribuire liberamente i dati, il secondo indica le tecnologie che garantiscono una maggiore interoperabilità. I linked data in sostanza non devono necessariamente essere *open*, ma la potenzialità di questa tecnologia si sprigiona al meglio quando i dati sono pubblicati come *linked open data*²¹⁸.

Ma quali sono i vantaggi dell'approccio linked data? E quali benefici può apportare?

I linked data, e in particolare i linked open data, sono condivisibili, estensibili e riutilizzabili. Supportano funzionalità multilingue, grazie alla separazione tra etichette (*label*) e concetti, i quali se identificati da URI diventano agnostici rispetto alla lingua. Inoltre le risorse possono essere descritte con metodi collaborativi, coinvolgendo altre biblioteche o utilizzando i linked data messi a disposizione da altre comunità o da singoli individui. Come i link che oggi vengono creati tra i documenti del web, allo stesso modo i linked data si propongono di estendere a chiunque la possibilità di contribuire condividendo la propria conoscenza in una forma riutilizzabile e ricombinabile con la conoscenza messa a disposizione da altri²¹⁹. L'uso di identificativi consente di riferire diverse descrizioni alla stessa entità, dando vita ad un ambiente informativo pluralistico e più ricco. Le biblioteche, proprio grazie alla creazione di link con dati aggiuntivi e complementari, selezionati da fonti affidabili, possono accrescere il valore dei loro dati, che in questo modo aumenta se paragonato alle descrizioni prese individualmente.

L'utilizzo di identificativi globali univoci applicato alle opere, alle persone, ai luoghi, agli eventi, ai soggetti o altre entità di interesse, fa sì che le risorse possano essere citate in molti ambienti web e di conseguenza valorizza la portata dei metadati creati dalle biblioteche.

²¹⁸ Baker et al., «Library Linked Data Incubator Group Final Report»

²¹⁹ Ibid.

Un altro vantaggio dato dall'utilizzo degli identificativi univoci è che consentono ai *data providers* (fornitori di dati) di condividere porzioni di dati espresse come *statement*, cioè una singola tripla RDF composta da soggetto, predicato e oggetto²²⁰. Oggi la maggior parte dei dati bibliografici viene condivisa sotto forma di interi record, ognuno dei quali rappresenta una descrizione completa. In un ecosistema basato sui grafi RDF una istituzione può invece utilizzare singole informazioni, in forma di *statement*, per aggiungere i dati mancanti. Ad esempio una biblioteca potrebbe aver bisogno solamente dei dati che riguardano una specifica pubblicazione, come i dati del campo 260 del MARC21 con i vari sottocampi, e “catturare” solo le triple che riguardano quella porzione della descrizione, senza dover importare un intero record. Le biblioteche oggi ritengono opportuno integrare le proprie descrizioni con dati creati esternamente, basti pensare alle copertine, agli abstract o anche ai commenti dei lettori e alle recensioni. Con dati atomizzati nella forma proposta da RDF questo potrebbe avvenire in maggior misura. In particolare i dati di autorità delle biblioteche, come i nomi di persona ed enti, i soggetti, le classi e i titoli delle opere, se identificati da URI e strutturati come linked data possono ridurre la ridondanza che spesso è presente nei cataloghi²²¹. In quanti database di biblioteche, o reti di biblioteche, è presente il termine “arte” come dato di autorità? Perché doverlo duplicare continuamente?

Il Report del Library Linked Data Incubator Group analizza anche i benefici che possono avere gli utenti. L'utilizzo di linked data può portare a creare maggiori connessioni tra i dati, nel rispetto degli ICP del 2016 che prescrivono tra gli obiettivi del catalogo la possibilità di navigare all'interno, ma anche al di fuori delle entità del catalogo stesso²²². La navigazione tra risorse che provengono da contesti bibliotecari e non bibliotecari può diventare più sofisticata, perché le relazioni stabilite a livello di atomi informativi possono aumentare enormemente la funzione di scoperta di nuova conoscenza. Ogni singolo dato può essere esplorato all'interno della rete di connessioni che presenta, proprio come avviene quando si comincia a navigare dentro il portale di *data.bnf.fr*²²³, che consente a partire da un'entità ricercata di esplorare tutti i dati che le sono relazionati.

²²⁰ Baker et al., «Library Linked Data Incubator Group Final Report»

²²¹ Ibid.

²²² IFLA Cataloguing Section e IFLA Meetings of Experts on an International Cataloguing Code, «Statement of International Cataloguing Principles (ICP)».

²²³ «index (data.bnf.fr)».

Visualizzando la descrizione di Panizzi²²⁴, l'utente può decidere di cliccare su "Italia" e vedere tutto ciò che è collegato a quel termine, oppure cliccare sulla data di nascita "1797" ed esplorare tutto ciò che è legato a quel dato, soggetti, autori nati o deceduti in quella data, organizzazioni create, opere ideate o terminate, pubblicazioni, performance etc. Il *Final Report* a questo proposito utilizza un'espressione molto efficace per rendere l'idea: "a data-powered form of toURIs²²⁵", in riferimento ai possibili percorsi che l'utente può seguire attraverso le connessioni tra dati identificati tramite URI, una sorta di turismo che promuove la scoperta delle informazioni navigando attraverso le strade degli URI. I linked data consentono infatti di collegare dati che provengono da molte comunità, si pensi a Wikipedia, GeoNames, la BBC, il New York Times, MusicBrainz, immettendo le collezioni locali in un universo informativo più vasto nel web²²⁶.

Anche le istituzioni e i bibliotecari possono trarre vantaggio dall'applicazione dei linked data. Il tradizionale approccio *top-down* ai dati bibliografici, che prevede che le descrizioni bibliografiche siano elementi isolati usati solo nel contesto bibliotecario, presenta dei problemi principalmente di tipo economico. Molte biblioteche infatti non hanno le risorse economiche per creare informazioni con un alto livello di granularità e analiticità, pratica che richiede tempo e professionalità. Con i linked data molte tipologie di dati possono essere create in modo decentralizzato, da attori diversi, e poi aggregati in un singolo grafo RDF²²⁷. Questo approccio può essere visto come un primo passo verso la possibilità di gestire le informazioni culturali in modalità *cloud-based*. Questa soluzione consente anche alle piccole istituzioni o ai singoli progetti di essere più visibili e connessi e di ridurre i costi per le infrastrutture tecnologiche necessarie. Essere visibili nel web, non solo con i propri siti, ma anche con i propri dati, significa essere dove la maggior parte delle persone cerca le informazioni e quindi essere là dove sono gli utenti²²⁸. L'apertura dei dati, come già detto, non è un obbligo ma ricordiamo che "the coolest thing to do to your data will be thought of by someone else"²²⁹, cioè rendere i dati

²²⁴ «Antonio Panizzi (1797-1879) - Author - Resources from the BnF», http://data.bnf.fr/13739722/antonio_panizzi/.

²²⁵ Baker et al., «Library Linked Data Incubator Group Final Report».

²²⁶ Ibid.

²²⁷ Ibid.

²²⁸ Connaway e Dickey, «The Digital Information Seeker: Report of Findings from Selected OCLC, RIN and JISC User Behaviour Projects».

²²⁹ Baker et al., «Library Linked Data Incubator Group Final Report».

liberamente riutilizzabili e condivisibili può far sì che ad altre persone venga in mente un loro utilizzo inaspettato. Ad esempio possono creare delle applicazioni che consentono di esplorarli a partire da nuovi punti di vista. Il valore della condivisione può incidere sulla valorizzazione dei dati, a partire anche dagli usi che attori esterni ne possono fare.

I bibliotecari oggi non riescono a recuperare tutti i dati disponibili quando creano una descrizione, molti dati devono essere creati da zero, pur esistendo già da qualche parte nel web. I linked open data potrebbero aumentare l'interoperabilità consentendo minori sforzi e minore ridondanza²³⁰. Proprio l'uso di identificativi consente di prendere dati da più ambiti culturali, e più in generale dal web, permettendo ai catalogatori di risparmiare tempo sulla creazione di informazioni che esistono già e di concentrarsi eventualmente sull'arricchimento della risorsa aggiungendo e apportando tutte le specifiche previste dal nostro settore. I dati potrebbero quindi essere catturati nel maggior numero possibile, senza togliere specificità, qualità e controllo, tutti elementi previsti dalla disciplina catalografica.

Un altro aspetto molto importante riguarda i formati dei dati. L'evoluzione delle tecnologie utilizzate per gestire l'informazione dimostra la loro transitorietà. I linked data si propongono di esplicitare il significato dei dati separatamente dalle specifiche strutture degli stessi, cioè dalla sintassi per esempio dell'XML o dai formati, come MARC21 o UNIMARC. I formati dei metadati in questo senso dipendono da una particolare struttura e ne sono vincolati. Il fatto di non essere più costretti ad utilizzare solamente i formati specifici dell'ambiente bibliotecario può avere delle ricadute positive anche sui venditori di software e sugli sviluppatori. Avere una modalità di scambio e riutilizzo dei dati supportata da tutti i fornitori di metadati (*metadata providers*) significa ad esempio non dover ricorrere per forza allo Z39.50, un protocollo che come si è visto viene usato quasi solamente dalle biblioteche. Il protocollo HTTP previsto dai linked data è molto più conosciuto²³¹. Usare dati basati HTTP URI nella forma proposta da RDF libera dunque i venditori dalla necessità di usare formati specifici e di adattare di conseguenza i propri prodotti a quei formati, perché i linked data prevedono che i dati possano essere genericamente comprensibili al web ma non li vincolano ad uno specificato formato. Gli

²³⁰ Ibid.

²³¹ Baker et al., «Library Linked Data Incubator Group Final Report»

strumenti che vengono creati potrebbero quindi adattarsi a diversi contesti. Ora invece ogni ambito, anche solo all'interno di quello culturale, utilizza generalmente modelli di dati e normative descrittive differenti, basti pensare agli archivi, ai musei e alle biblioteche. L'interoperabilità prevista dai linked data in questo senso andrebbe nella direzione già prevista dal coordinamento MAB (Musei Archivi Biblioteche)²³² Italia, che si propone di esplorare le convergenze tra gli istituti culturali.

Nel *Final Report* viene dedicata una sezione anche all'analisi della situazione attuale dei dati bibliografici. Nonostante alcune criticità siano già state messe in evidenza nel primo capitolo, è bene soffermarsi di nuovo su alcuni aspetti, che alla luce delle tecnologie proposte dai linked data diventano più comprensibili.

I dati delle biblioteche oggi sono contenuti in database, che presentano delle interfacce di ricerca accessibili dal web, ma non sono ancora del tutto integrati con altre fonti di dati. Le tipologie di dati in comune con le risorse di altri ambiti possono essere moltissime, dalle date, ai luoghi geografici, alle persone e alle organizzazioni. I linked open data prevedono la possibilità di collegare tutte queste informazioni. Oggi però i dati bibliografici sono quasi tutti espressi in uno formato MARC che stabilisce una precisa visualizzazione del record, le cui informazioni vengono date principalmente nel linguaggio naturale. Inoltre il MARC consente l'uso di identificativi, ma questi vengono perlopiù gestiti localmente e non vengono espressi come URI, da qui l'impossibilità di creare link nel web a partire da questi stessi dati²³³.

Il processo vero e proprio di pubblicazione dei linked open data prevede una serie di passaggi oltre alla consapevolezza dei benefici che essi possono portare, sulla quale ancora oggi il dibattito è aperto. Una prima importante raccomandazione riguarda il fatto di riutilizzare il più possibile vocabolari e ontologie standardizzate²³⁴, questo per facilitare l'espansione del web dei dati e dare maggiore visibilità alle risorse. Nel paragrafo

²³² «MAB - Musei Archivi Biblioteche», <http://www.mab-italia.org/>.

²³³ Baker et al., «Library Linked Data Incubator Group Final Report».

²³⁴ «Best Practices for Publishing Linked Data», <https://www.w3.org/TR/ld-bp/>.

seguito a questo scopo verrà fornita una panoramica sulle ontologie e sui vocabolari più importanti per il settore bibliotecario.

3.2. Ontologie e vocabolari per rappresentare i dati bibliografici

Come abbiamo visto, nella documentazione fornita dal W3C si esplicita che non è presente una chiara divisione tra i termini vocabolario e ontologia, e che la pratica comune è quella di utilizzare ontologia per una concettualizzazione più complessa che stabilisce classi e proprietà di un determinato dominio, mentre vocabolario è utilizzato in riferimento ad una lista di termini controllati²³⁵. Il risultato di questa ambiguità, scrive Karen Coyle, è che spesso il termine ontologia viene utilizzato al posto di vocabolario, che a sua volta viene spesso impiegato per indicare non solo i termini dei metadati ma anche i dati che vengono associati a dei metadati²³⁶. Ad esempio è possibile avere un metadato come *dcterms:language* che può essere associato a valori come “eng”, “fre”, “ita”. L’insieme di questi termini (“eng”, “ita”, “fre” etc.) viene chiamato anch’esso vocabolario. Per questo con vocabolario ci si può riferire sia ai termini controllati dei metadati che ai valori che possono esservi associati.

Le ontologie e gli schemi di metadati più importanti per il settore bibliotecario sono le seguenti:

- FRBR: questo modello è stato espresso in RDF in varie versioni. Nel 2011 l’IFLA ha pubblicato la versione ufficiale di FRBR in RDF, pubblicata nell’Open Metadata Registry, chiamata FRBRer poiché “er” indica entità relazioni. Questa ontologia presenta 10 classi (entità) e 206 proprietà (attributi e relazioni)²³⁷. Secondo Karen Coyle FRBRer è però stato progettato senza tener conto delle implicazioni di apertura del web semantico, nel senso che questo modello risulta troppo rigido per poter dialogare con dati strutturati secondo altri modelli²³⁸.

²³⁵ «Ontologies - W3C».

²³⁶ Coyle, «Semantic Web and Linked Data», *Tutorial on Metadata and the Semantic Web*, 2012, <http://kcoyle.net/metadata/7lecture.html>.

²³⁷ Pat Riva, «FRBR Review Group Initiatives and the World of Linked Data», *JLIS.it* 4, n. 1 (2013): 105, doi 10.4403/jlis.it-5482.

²³⁸ Coyle, *FRBR, Before and After*.

FRBRoo invece nasce dall'amornizzazione di FRBRer e CIDOC CRM²³⁹, che si focalizza sulla descrizione di risorse appartenenti al settore bibliotecario e museale;

- RDA in RDF²⁴⁰: presenta circa 1300 elementi, suddivisi solamente per facilità di consultazione secondo le entità FRBR, consente di utilizzare inoltre vocabolari controllati per i valori, alcuni chiusi altri liberamente accessibili²⁴¹. Lo schema è pubblicato su Open Metadata Registry;
- ISBD²⁴²: presenta la traduzione in RDF degli elementi ISBD, riproducendo la struttura delle 8 aree descrittive. Presenta 181 sottoclassi e proprietà, poiché fanno tutte riferimento ad un'unica classe, cioè la manifestazione di FRBR, indicata come "risorsa". Anch'esso pubblicato su Open Metadata Registry;
- BIBFRAME²⁴³: questo modello si ispira a FRBR ma propone tre classi principali, ovvero *work* (l'essenza concettuale di ciò che viene catalogato), *instance* (l'espressione materiale) e *item* (la singola copia), seguite da *agent* (persone, enti o associazioni che hanno a che fare con l'opera e l'istanza), *subject* (i soggetti delle opere) ed *event* (gli eventi e in particolari le registrazioni che possono contenere un'opera). BIBFRAME quindi comprende attributi che vengono associati sia all'opera che all'espressione di FRBR, mentre l'istanza è analoga alla manifestazione di FRBR²⁴⁴;
- FOAF (Friend of a Friend)²⁴⁵: serve per descrivere persone e relazioni, utilizzata dalle biblioteche per descrivere persone, luoghi associati e relazioni. Presenta comunque delle limitazioni rispetto all'analiticità e precisione richiesta per i dati bibliografici. Non consente ad esempio di indicare l'anno di nascita e di morte²⁴⁶;

²³⁹ «Home | CIDOC CRM», <http://www.cidoc-crm.org/>.

²⁴⁰ «RDA Registry | Home», <http://www.rdaregistry.info/>.

²⁴¹ Guerrini e Possemato, *Linked data per biblioteche, archivi e musei*.

²⁴² «IFLA ISBD Vocabularies».

²⁴³ «Overview of the BIBFRAME 2.0 Model (BIBFRAME - Bibliographic Framework Initiative, Library of Congress)», <https://www.loc.gov/bibframe/docs/bibframe2-model.html>.

²⁴⁴ Coyle, *FRBR, Before and After*.

²⁴⁵ «FOAF Vocabulary Specification», <http://xmlns.com/foaf/spec/>.

²⁴⁶ Guerrini e Possemato, *Linked data per biblioteche, archivi e musei*.

- Dublin Core²⁴⁷: uno schema di metadati molto popolare e utilizzato che serve per definire gli attributi di una risorsa, come l'autore, il titolo, l'editore. Moltissimi elementi di Dublin Core sono presenti nelle ontologie e negli schemi qui proposti.
- Schema.org²⁴⁸: nato nel 2011 dall'accordo di Bing, Google e Yahoo!, presenta la versione Schema.RDFS.org che aggiunge la possibilità di stabilire link tra i dati;
- RDF Schema²⁴⁹: strutturato in 6 classi e 7 proprietà, sviluppato all'interno del W3C, è stato usato anche per progetti come *data.bnf.fr*;
- GeoNames Ontology²⁵⁰: diffusissimo nel web e usato per descrivere luoghi geografici;
- CC REL (Creative Commons Rights Expression Language)²⁵¹: per la descrizione dei diritti, utile alle biblioteche che hanno bisogno di indicare i permessi di utilizzo delle risorse;
- SKOS²⁵²: per rappresentare vocabolari controllati e relazioni tra concetti.

Le ontologie esistenti sono moltissime, consultando LOV (Linked Open Vocabularies)²⁵³ ci si rende conto non solo di quanti vocabolari siano utilizzabili ma anche di quali siano i più popolari. Elemento, quest'ultimo, molto importante nella scelta di un'ontologia. I vocabolari che in LOV spiccano sono: Dublin Core, nelle sue diverse varianti, FOAF, GeoNames, Schema.org e SKOS.

La scelta di ontologie e vocabolari riveste un ruolo molto importante, così come la creazione di un profilo di applicazione che si rende necessario quando si riutilizzano classi e proprietà preesistenti.

²⁴⁷ «DCMI Home: Dublin Core Metadata Initiative (DCMI)», <http://dublincore.org/>.

²⁴⁸ «Home - schema.org», <http://schema.org/>.

²⁴⁹ «RDF Schema 1.1», <https://www.w3.org/TR/rdf-schema/>.

²⁵⁰ «GeoNames Ontology - Geo Semantic Web», <http://www.geonames.org/ontology/documentation.html>.

²⁵¹ «CC REL - Creative Commons», https://wiki.creativecommons.org/wiki/CC_REL.

²⁵² «SKOS Simple Knowledge Organization System - home page».

²⁵³ «Linked Open Vocabularies», <http://lov.okfn.org/dataset/lov/>.

3.2.1. La scelta delle ontologie e i profili di applicazione

Le ontologie servono ad inquadrare i dati attraverso una metodologia *bottom-up*, si parte da ciò che si vuole rappresentare e si decide come rappresentarlo. L'istituzione generalmente seleziona gli elementi che si ritiene abbiano più valore. Non è necessario rappresentare tutte le informazioni di cui si dispone, ma è importante che vengano valorizzate quelle funzionali agli usi che vengono stabiliti a monte. Una buona descrizione di un dataset infatti dovrebbe tener conto dei possibili utilizzi che verranno fatti di quei dati, e, in particolare iniziare a valutare ontologie vocabolari a partire dalle necessità degli utenti di quello specifico dominio a cui appartiene il dataset²⁵⁴.

Negli ultimi anni si è assistito al proliferare di ontologie e vocabolari, questo ha reso cruciale l'aspetto riguardante la loro scelta. Il processo di selezione dovrebbe utilizzare un approccio olistico e tenere in considerazione diversi elementi. Alcune indicazioni vengono fornite da Iker Huerga e Michael Lauruhn²⁵⁵, e prevedono l'analisi dei seguenti fattori:

- il dominio di appartenenza dei contenuti delle risorse;
- la quantità di risorse impiegabili nella personalizzazione di un modello;
- la tecnologia alla base del vocabolario o dell'ontologia;
- la sostenibilità a lungo termine delle risorse.

Il primo criterio da seguire, quando si seleziona un'ontologia, è assicurarsi che questa riesca a rappresentare il dominio e gli argomenti relativi ai contenuti delle collezioni e dei dati che si vogliono convertire. Conoscere il settore di appartenenza e il significato dei dati risulta dunque cruciale. Solo così infatti si creano le premesse per scegliere in modo corretto l'ontologia più appropriata a rappresentare il proprio dominio. Un altro fattore importante è la conoscenza dell'utenza che usufruisce di questi dati, chiedersi quali bisogni informativi si vogliono soddisfare con i propri dati, quali esigenze conoscitive hanno le persone che accedono a quelle particolari collezioni e quali esigenze si vogliono in qualche modo anticipare, consente di selezionare in modo più appropriato tra i vari

²⁵⁴ M. Cristina Pattuelli, Alexandra Provo, e Hilary Thorsen, «Ontology Building for Linked Open Data: A Pragmatic Perspective», *Journal of Library Metadata* 15, n. 3-4 (2015): 265-94, doi:10.1080/19386389.2015.1099979.

²⁵⁵ Iker Huerga e Michael P. Lauruhn, «Linked Data Implications for Authority Control and Vocabularies», in *Linked Data for Cultural Heritage*, a c. di Ed Jones e Michele Seikel (London: Facet, 2016), 55-72.

modelli. Un altro criterio importante è la valutazione della granularità descrittiva e della specificità del vocabolario

Un passo successivo nella selezione delle ontologie è la valutazione del contenuto e della qualità del modello. Questo processo include l'analisi dell'accuratezza e dell'attualità delle informazioni, la valutazione dell'affidabilità di chi pubblica quel modello o vocabolario, e anche il grado di appropriatezza del linguaggio utilizzato rispetto all'utenza di riferimento.

Stabilito il contenuto e la qualità dell'ontologia, è importante considerare la gestione dell'ontologia stessa. In particolare bisogna capire quanto spesso quel modello viene aggiornato e che tipo di modifiche vengono fatte, e come si viene a conoscenza di queste stesse modifiche. Bisogna anche capire quanto lavoro serve per modificare il modello e farlo aderire ai nostri bisogni. Le più comuni personalizzazioni di solito comprendono l'aggiunta di elementi ai concetti e l'aggiunta di sinonimi ai termini dei vocabolari.

Va valutata anche la sostenibilità del vocabolario o del modello. Spesso infatti alcuni di essi, normalmente accessibili nel web, vengono creati per supportare singoli progetti e, una volta raggiunto il risultato, non più aggiornati. In altri invece non è chiaro chi sia il proprietario e non si sa per quanto tempo il vocabolario resterà accessibile. Per questo è importante sapere se chi mantiene quel vocabolario è un'organizzazione stabile e affidabile e conoscere la strategia commerciale che c'è alla base.

Spesso le istituzioni hanno necessità specifiche che non combaciano con le funzionalità e la specificità di una determinata ontologia o di un vocabolario. Soprattutto quando si tratta di garantire l'accesso a collezioni di oggetti molto eterogenei o molto specifici, l'utilizzo di un solo schema non è sufficiente, poiché non riesce a rappresentare in modo analitico quella realtà di dati²⁵⁶. Una buona pratica, ampiamente utilizzata nelle comunità che sviluppano linked open data, è il riutilizzo di ontologie e vocabolari esistenti²⁵⁷ e per necessità particolari la creazione dei profili di applicazione.

²⁵⁶ Susanna Peruginelli, «*I profili di applicazione*», Presentazione tenuta a Seminario nazionale sui metadati, ICCU, Roma, 3 aprile 2001, www.iccu.sbn.it/upload/documenti/Peruginelli.ppt.

²⁵⁷ Allemang e Hendler, *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*.

Un profilo di applicazione è composto da elementi presi da schemi di metadati preesistenti e combinati insieme per ottimizzare una determinata applicazione²⁵⁸. Questi schemi, che si basano su RDF e XML, servono a dichiarare gli elementi utilizzati per un progetto, indicando per ciascuna classe e proprietà i *namespace* di provenienza con riferimento agli URI dove risiedono le definizioni degli elementi²⁵⁹.

I profili di applicazione generalmente forniscono documentazione sui seguenti aspetti:

- specifiche degli elementi utilizzati con riferimento agli schemi da cui sono presi;
- informazioni sui responsabili del profilo di applicazione;
- indicazioni sulle regole d'uso e sui vocabolari di riferimento con la possibilità di raffinare alcune definizioni fornendo informazioni semantiche più specifiche o più generiche a seconda dei casi²⁶⁰;
- specifiche relative all'aggiunta di nuovi elementi creati appositamente per quel profilo.

La creazione di questa documentazione ha lo scopo di migliorare l'interoperabilità tra risorse e applicazioni in riferimento alla creazione e alla gestione dei metadati ma anche alla ricerca e al recupero delle informazioni²⁶¹. In un'ottica di scambio e riutilizzo dei dati è infatti importante optare per politiche di gestione dei metadati trasparenti, che consentano anche a terzi di poter eventualmente riutilizzare il lavoro fatto per quel progetto.

Da un punto di vista storico la nascita del concetto di profilo di applicazione è connessa alla DCMI (Dublin Core Metadata Initiative)²⁶². Dublin Core è un progetto nato nel 1994²⁶³ durante la Conferenza Internazionale sul World Wide Web che si tenne a Chicago, quando alcuni membri di OCLC (Online Computer Library Center)²⁶⁴ e della NCSA

²⁵⁸ Rachel Heery e Manjula Patel, «Application Profiles: Mixing and Matching Metadata Schemas», *Ariadne*, n. 25 (2000), <http://www.ariadne.ac.uk/issue25/app-profiles>.

²⁵⁹ Peruginelli, «I profili di applicazione».

²⁶⁰ Heery e Patel, «Application Profiles».

²⁶¹ Peruginelli, «I profili di applicazione».

²⁶² «IFLA -- DCMI library application profile», <http://www.ifla.org/best-practice-for-national-bibliographic-agencies-in-a-digital-age/node/8824>.

²⁶³ «DCMI History», <http://dublincore.org/about/history/>.

²⁶⁴ Online Computer Library Center (OCLC) è un'organizzazione cooperativa di biblioteche, no profit, che propone servizi e prodotti per le biblioteche e le istituzioni culturali. La sede principale è a Dublin in Ohio. OCLC è la rete di biblioteche più ampia al mondo. «OCLC: Worldwide, member-driven library cooperative», <https://www.oclc.org/en/home.html>.

(National Center for Supercomputing Applications)²⁶⁵ cominciarono a discutere sulle implicazioni del web semantico e sulla difficoltà di reperire risorse per i loro progetti. Nel 1995 durante un incontro tenuto a Dublin, guidato sempre da OCLC e da NCSA, emerse chiaramente l'esigenza di avere un set di metadati semantici, per risorse web, che migliorasse e facilitasse i sistemi di ricerca e recupero²⁶⁶. Da allora DC si è imposto come uno dei più utilizzati set di metadati. Le esigenze delle comunità e delle applicazioni relativamente ai metadati sono però molto diversificate e questo ha portato alla proliferazione di formati, anche per applicazioni che presentano necessità simili²⁶⁷. Per ovviare a questo problema DC ha avviato nel 2009 un progetto per la creazione di profili di applicazione che consentano di produrre set di metadati in grado di rispondere ai bisogni specifici. Questo progetto ha così gettato le basi fornendo linee guida generiche sulla costruzione di questi documenti. In particolare si possono ricordare i passaggi principali che servono a creare un profilo di applicazione, ovvero un documento o una serie di documenti che descrivono i metadati che sono stati utilizzati.

Inizialmente vanno definiti i requisiti funzionali dell'applicazione: i metadati servono a supportare un'attività. Il primo passo è capire quali sono gli obiettivi dell'applicazione che si sta creando. Essi fungono da guida per definire quali sono i requisiti necessari ai metadati e individuare i confini del proprio progetto. Alcune domande, prese dalle linee guida DC²⁶⁸, possono aiutare in questa prima fase:

- cosa si vuole ottenere con la propria applicazione?
- quali sono i limiti del proprio progetto, o meglio cosa non si vuole ottenere con esso?
- quali funzioni l'applicazione deve garantire agli utenti?
- quali sono le caratteristiche principali delle risorse prese in considerazione?
- quali sono le caratteristiche principali dell'utenza di riferimento? Gli utenti appartengono ad una specifica istituzione o ci si vuole rivolgere ad un'utenza

²⁶⁵ «Home | National Center for Supercomputing Applications at the University of Illinois», <http://www.ncsa.illinois.edu/>.

²⁶⁶ «DCMI History».

²⁶⁷ Karen Coyle e Thomas Baker, «DCMI Home: Dublin Core Metadata Initiative (DCMI)», 2009, <http://dublincore.org/>.

²⁶⁸ Ibid.

generica? Che lingue parlano? Che tipo di confidenza gli utenti hanno con le risorse che saranno descritte e con i dati che si vogliono gestire e valorizzare?

- esistono già degli standard comuni che possono essere presi in considerazione?

Un secondo aspetto, quando si crea un profilo di applicazione, riguarda la selezione e lo sviluppo di un *domain model*²⁶⁹, ovvero di un modello che riesca a descrivere gli oggetti che saranno rappresentati dai metadati e le relazioni che possono intercorrere tra questi stessi oggetti. I *domain model* possono essere semplici o complessi. Ad esempio *Scholarly Works Dublin Core Application Profile* si basa su FRBR, il modello concettuale utilizzato per organizzare l'universo bibliografico²⁷⁰, con alcune modifiche e personalizzazioni per l'ambito della stampa elettronica relativa ai documenti scientifici.

Una volta definito il modello per i metadati, si scelgono le proprietà per descrivere le risorse²⁷¹, ad esempio un libro può avere un titolo e un autore, a sua volta l'autore può essere una persona con un nome e una data di nascita. In questa fase è necessario valutare i vocabolari RDF già esistenti per capire se le proprietà necessarie esistono già. L'uso di proprietà esistenti, se si rivelano appropriate, richiede sforzi minori e migliora l'interoperabilità dei metadati²⁷².

Ma come si pubblicano i linked open data?

3.3. Library linked open data cookbook

Nel 2014 il W3C ha pubblicato le *Best Practices for Publishing Linked Data* che indicano i passaggi principali che devono essere affrontati per pubblicare linked data.

Si inizia individuando un set di dati a cui si vuole dare la priorità. I linked open data hanno il vantaggio di poter essere scalari, cioè possono essere pubblicati in *tranche* successive andando ad ampliare e arricchire man mano il dataset. Il *Final Report* raccomanda di iniziare dai dati di autorità, in quanto consentono di lavorare su una quantità ridotta di

²⁶⁹ Coyle e Baker, «DCMI Home: Dublin Core Metadata Initiative (DCMI)»

²⁷⁰ Barbara Tillett, «What is FRBR? A Conceptual Model for the Bibliographic Universe» (Library of Congress Cataloging Distribution Service, 2004).

²⁷¹ Coyle e Baker, «DCMI Home: Dublin Core Metadata Initiative (DCMI)».

²⁷² Ibid.

dati e allo stesso tempo garantiscono la presenza nella *linked data cloud*, senza incidere sul normale flusso di lavoro a cui sono sottoposti gli altri dati²⁷³.

Decidere di pubblicare una porzione di dati esistenti, nel caso delle biblioteche si parla di record bibliografici già strutturati contenuti in database generalmente relazionali, può richiedere però una loro preventiva pulizia attraverso operazioni di *data cleaning*. A questo scopo esistono diversi strumenti, come ad esempio OpenRefine²⁷⁴. Non è detto infatti che si vogliano convertire tutte le informazioni contenute in un record MARC, per questo si decide a monte su quali dati lavorare, e se necessario si agisce su di essi per renderli più omogenei ed eliminare le duplicazioni. Successivamente a questa operazione si decide che ontologia o vocabolari utilizzare e si passa alla modellizzazione vera e propria, questa operazione è concettuale ed indipendente dalle applicazioni tecnologiche.

I linked data prevedono un approccio *mix and match*, per cui è consigliato utilizzare vocabolari e dati semantici già esistenti. A questo proposito esistono diversi dispositivi software che permettono di integrare linked data già esistenti, ad esempio Falcons²⁷⁵, Watson²⁷⁶ e Semantic Web Search Engine²⁷⁷. Protégé invece è un vero e proprio editor per creare ontologie e modelli di dati²⁷⁸. Se invece si decide di creare una rappresentazione formale *ad hoc* bisogna definire l'URI che andrà ad identificarla e seguire le pratiche che riguardano la loro costruzione, come indicato nei Principi dei linked data²⁷⁹, e creare un file che contiene tutte le informazioni sugli elementi scelti.

I dati vengono poi convertiti in RDF attraverso ulteriori dispositivi. Attualmente ne esistono diversi e tutti in continua evoluzione, quelli consigliati dal W3C²⁸⁰ per la conversione di record MARC in RDF/XML sono marcmods2rdf, MARiMbA e easyM2R. Le descrizioni RDF possono però essere serializzate anche in formati come Turtle, N-Triple o JSON-LD.

²⁷³ Baker et al., «Library Linked Data Incubator Group Final Report».

²⁷⁴ «OpenRefine», <http://openrefine.org/>.

²⁷⁵ «Falcons Object Search», <http://ws.nju.edu.cn/falcons/objectsearch/index.jsp>.

²⁷⁶ «Watson Semantic Web Search», <http://watson.kmi.open.ac.uk/WatsonWUI/>.

²⁷⁷ «semanticwebsearch.com», <http://www.semanticwebsearch.com/query/>.

²⁷⁸ «protégé», <http://protege.stanford.edu/>.

²⁷⁹ Berners-Lee, «Linked Data - Design Issues».

²⁸⁰ «ConverterToRdf - W3C Wiki», <https://www.w3.org/wiki/ConverterToRdf#MARC>.

Un altro elemento da considerare quando si pubblicano linked data è dato dalle licenze che vengono applicate ai dati. È importante specificare la provenienza dei dati e associarvi una licenza, in modo che essa “viaggi” con i dati stessi, consentendo alle persone che decidono di riutilizzarli di sapere se possono farlo e come devono farlo²⁸¹. Le licenze che vengono consigliate nell’ambito dei dati sono le Creative Commons (CC)²⁸², ormai riconosciute a livello internazionale. Ogni licenza CC comprende tre formati diversi²⁸³:

- i *commons deed*: una sintesi del contratto destinata alle persone;
- un *legal code*: il contratto vero e proprio dove si utilizza un linguaggio tecnico giuridico;
- i metadati: leggibili solo dal computer.

Le principali licenze CC sono date dalla combinazione di quattro elementi: BY implica la titolarità, SA (*share alike*) richiede l’utilizzo della stessa licenza affinché la risorsa possa essere utilizzata; NC (*non commercial*) vieta il riutilizzo per scopi commerciali; ND (*non derivative works*) impedisce che a partire da una risorsa vengano create opere derivate. Le licenze che ne risultano sono le seguenti²⁸⁴:

- CC0 (Creative Commons Zero): non è una vera e propria licenza, dichiara che la risorsa è di pubblico dominio e che il titolare rinuncia a tutti i diritti garantiti dalle leggi sul copyright di tutto il mondo.
- CC BY: esprime un’attribuzione, consente a terzi di distribuire, utilizzare, modificare e usare la risorsa anche per scopi commerciali con il solo vincolo di indicare la titolarità;
- CC BY-SA: la risorsa può essere riutilizzata ma solo con la stessa licenza dell’originale, è la licenza usata da Wikipedia e da molti software *opensource*;
- CC BY-ND: permette la redistribuzione, anche commerciale, ma la risorsa deve essere invariata e presentare la titolarità;
- CC BY-NC: vieta il riutilizzo per scopi commerciali;

²⁸¹ «Best Practices for Publishing Linked Data».

²⁸² «When we share, everyone wins - Creative Commons», <https://creativecommons.org/>.

²⁸³ «Creative Commons FAQ | CreativeCommons.it», <http://www.creativecommons.it/FAQ>.

²⁸⁴ «A proposito delle Licenze - Creative Commons», <https://creativecommons.org/licenses/>.

- CC BY-NC-SA: si può modificare e ridistribuire ma con termini non commerciali e utilizzando la stessa licenza della risorsa;
- CC BY-NC-ND: permette di scaricare e condividere una risorsa, con attribuzione di titolarità, senza apportare modifiche e per scopi non commerciali. È la più restrittiva delle licenze CC.

La OKFn (Open Knowledge Foundation)²⁸⁵, un'organizzazione no profit internazionale focalizzata sulla promozione degli *open data*, ha valutato compatibili con la *open definition*²⁸⁶ la CC0, la CC-BY e la CC-BY-SA. In Italia FormezPA²⁸⁷ ha proposto delle licenze specifiche, IODL (Italian Open Data License), per le amministrazioni pubbliche:

- IODL 2.0²⁸⁸: l'utente può scaricare, copiare, estrarre, trasmettere, pubblicare e distribuire le informazioni, può inoltre creare opere derivate ad esempio attraverso il *mashup* di informazioni, si richiede solamente di indicare la fonte delle informazioni, il nome di chi fornisce la risorsa ed eventualmente l'inclusione di un link alla licenza;
- IODL 1.0²⁸⁹: rispetto alla prima obbliga alla pubblicazione o condivisione delle opere derivate, che devono avere la stessa licenza o una licenza compatibile.

Esistono anche altri tipi di licenze, riconosciute a livello internazionale, che rispettano la *open definition* di OKFn, ad esempio la ODC PDDL (Open Data Commons Public Domain Dedication and License), utilizzabile anche per i database e i dati che contengono e usata ad esempio per il catalogo della biblioteca del CERN²⁹⁰, la ODC-BY (Open Data Commons Attribution) applicata ai dati di OCLC²⁹¹, la ODbL (Open Database License), la GNU FDL (GNU Free Documentation License) e la GNU GPL (GNU General Public License).

²⁸⁵ «Open Knowledge: Home», <https://okfn.org/>.

²⁸⁶ «The Open Definition - Open Definition - Defining Open in Open Data, Open Content and Open Knowledge», <http://opendefinition.org/>.

²⁸⁷ «Formez PA | Centro servizi, assistenza, studi e formazione per l'ammodernamento delle P.A.», <http://www.formez.it/>.

²⁸⁸ «Dati aperti: la nuova licenza italiana | Trasparenza e Integrità», <http://trasparenza.formez.it/content/dati-aperti-nuova-licenza-italiana>.

²⁸⁹ «Italian Open Data License v1.0», <http://www.formez.it/iodl/>.

²⁹⁰ De Robbio e Giacomazzi, «Dati aperti con LODe».

²⁹¹ Ibid.

Dopo aver individuato il modello di dati e i vocabolari di riferimento, aver scelto la tipologia di licenza, aver convertito i dati in RDF e averli serializzati in formati come RDFa, JSON-LD, Turtle o N-Triples e RDF/XML²⁹², è necessario creare link ad altri dataset. L'*interlinking* è infatti il valore aggiunto dei linked open data perché crea dei ponti con i dati di altre comunità e valorizza la portata di un dataset che altrimenti rimarrebbe isolato²⁹³. L'*interlinking* deve essere fatto partendo dall'analisi di dataset affini ai propri dati, valutando la stabilità e la credibilità che li caratterizza. Per capire quali dataset esistono si può consultare *Datahub*²⁹⁴ una piattaforma che rende disponibili dati e servizi per la condivisione di linked data. La creazione di collegamenti tra i linked open data di un dominio specifico e quelli di altre comunità viene chiamato *allineamento semantico*²⁹⁵. In pratica ci si chiede se qualcun altro abbia pubblicato dati sullo stesso soggetto e si crea un link di equivalenza o similarità utilizzando generalmente la proprietà *owl:sameAs*²⁹⁶. Il fine in questo caso non è quello di dichiarare che le risorse del nostro dataset sono identiche a quelle con cui vengono allineate, ma che con fonti e qualità differenti si sta parlando della stessa cosa. L'allineamento consente di dare visibilità ai dati perché garantisce la loro presenza in dataset molto popolari, come DBpedia, Wikidata e VIAF. Questo procedimento può essere fatto sia tra dataset interni che esterni. Anche per queste operazioni sono stati creati *tool* per la creazione di link tra dati²⁹⁷, ad esempio Silk Link Discovery Framework²⁹⁸, un *opensource* che è compatibile con tutti i dataset messi a disposizione con il protocollo SPARQL.

Una volta che i linked open data sono stati creati, essi vanno inseriti in un *triplestore*, un database apposito per conservare e gestire triple RDF. Esistono molti *triplestore*, alcuni prodotti appositamente, altri già esistenti a cui ci si può aggregare. Ad esempio il dataset dell'IBC (Istituto per i beni artistici culturali e naturali dell'Emilia-Romagna) sono conservati in un *triplestore* pubblico della Regione²⁹⁹. Le triple RDF vanno dunque

²⁹² «Best Practices for Publishing Linked Data».

²⁹³ Guerrini e Possemato, *Linked data per biblioteche, archivi e musei*.

²⁹⁴ Datahub si appoggia a CKAN (Comprehensive Knowledge Archive Network), un ambiente che serve a contenere e distribuire i dati aperti e accessibili. «About - the Datahub», <https://datahub.io/about>.

²⁹⁵ Baker et al., «Library Linked Data Incubator Group Final Report».

²⁹⁶ «OWL Web Ontology Language Reference», <https://www.w3.org/TR/2004/REC-owl-ref-20040210/#sameAs-def>.

²⁹⁷ Baker et al., «Library Linked Data Incubator Group Final Report».

²⁹⁸ «Silk - The Linked Data Integration Framework», <http://silkframework.org/>.

²⁹⁹ «Linked open data — E-R Istituto per i beni artistici culturali e naturali».

inserirle in un apposito database e poi, come indicato dai Principi per la creazione dei linked data, si crea uno SPARQL endpoint, per aprire una finestra sul mondo e fare in modo che i dati possano essere interrogati e riutilizzati³⁰⁰.

Non tutti però hanno familiarità con SPARQL, per consentire un accesso più “umano” ai dati, Diego Camarga nel 2014 ha progettato LodView, uno strumento che fornisce una rappresentazione delle risorse in HTML agevolandone la contestualizzazione e la comprensione³⁰¹. LodView è stato implementato nel progetto IBC Linked Open Data³⁰² e una versione simile, LodLive³⁰³, è stata integrata in LinkedJazz³⁰⁴. Quest’ultimo è un progetto di ricerca focalizzato sull’utilizzo dei linked open data nell’ambito del patrimonio culturale, che ha lo scopo di valorizzare le connessioni tra i documenti e i dati appartenenti al mondo del jazz.

Queste sono le principali fasi che portano alla pubblicazione dei linked open data, le istituzioni dovrebbero poi dedicarsi alla valorizzazione dei dati attraverso strumenti semplici ed intuitivi generalmente dati attraverso interfacce di ricerca, come LodView e LodLive. Questo ambito è ancora in fase di sperimentazione e non esistono linee guida o raccomandazioni ma soltanto singoli progetti pionieristici. Ad esempio BILLI (Bibliographic Identifiers for Library Location Information)³⁰⁵ della New York Public Library cerca di valorizzare i dati di collocazione delle risorse attraverso un’interfaccia navigabile dagli utenti. Alla New York Public Library, dal 1899 ad oggi, si sono stratificate cinque tipologie di classificazioni. Con l’utilizzo dei linked open data è stato possibile rappresentare le classi e il loro significato, le relazioni e cambiamenti che le hanno caratterizzate, creando una sorta di visualizzazione storica e dinamica che può servire per conoscere l’istituzione. Altre interfacce di ricerca verranno prese in considerazione nel capitolo dedicato ai casi di studio.

³⁰⁰ «SparqlEndpoints - W3C Wiki».

³⁰¹ «Nasce LodView di diego-valerio-camarda | regista», <http://www.regesta.com/2014/12/19/nasce-lodview-il-web-dei-dati-ha-un-nuovo-volto/>.

³⁰² «IBC Linked Open Data — LodView», <http://dati.emilia-romagna.it/id/ibc/website/html>.

³⁰³ «LodLive - browsing the Web of Data», <http://lodlive.it/>.

³⁰⁴ «Linked Jazz | Revealing the Relationships of the Jazz Community», <https://linkedjazz.org/>.

³⁰⁵ «Billi», <http://billi.nypl.org/>.

3.4. Alcune criticità

I linked open data rappresentano una possibilità, non l'unica, per integrare dati nel web e per renderli maggiormente interoperabili. Essendo una tecnologia ancora sperimentale servirà del tempo affinché le istituzioni che hanno deciso di adottare queste tecnologie condividano in modo più approfondito le principali criticità. Tuttavia alcune perplessità sono già state espresse.

Un problema è dato dalla difficoltà di rappresentare in modo efficace la semantica dei contenuti informativi e in particolare di quelli più complessi e articolati. Un'analisi di OCLC, condotta tra il 2014 e il 2015, ha evidenziato, tra le principali barriere per pubblicare linked open data, proprio la selezione di un'ontologia appropriata³⁰⁶. Inoltre non si è ancora in grado di gestire in modo completamente automatico l'assegnazione di notazioni di significato ai dati, poiché “le componenti architetturali, i linguaggi, i vocabolari, non sono del tutto trasparenti e capaci di comunicare tra di loro”³⁰⁷.

Interrogare i dati con SPARQL non è semplice, per farlo non solo serve conoscere SPARQL ma bisogna anche avere familiarità con lo schema dei dati che si vogliono esplorare³⁰⁸. In pratica se si vogliono creare *query*, ad esempio sullo SPARQL Endpoint di DBpedia, si deve conoscere l'ontologia di DBpedia altrimenti non è possibile richiamare i dati attraverso i metadati applicati alle informazioni. Per questo è molto importante che venga condivisa la documentazione sulla scelta degli elementi e sui vocabolari di riferimento. A questo proposito si rivelano utili i profili di applicazioni, sia per le istituzioni che possono utilizzarli come punto di partenza di un progetto, sia per le persone che fanno ricerca sui dati. Considerare uno SPARQL Endpoint alla stregua di un'API può poi essere fuorviante, come scrive Dimitri van Hees “it's equivalent to giving

³⁰⁶ Karen Smith-Yoshimura, «Analysis of International Linked Data Survey for Implementers», *D-Lib Magazine* 22, n. 7/8 (2016), doi:10.1045/july2016-smith-yoshimura.

³⁰⁷ Lo Castro, «Web Semantico e Linked Open Data: best practices, prospettive e criticità»: 218.

³⁰⁸ Matthew Lincoln, «Linked Open Realities: The Joys and Pains of Using LOD for Research», *matthewlincoln.net*, 2016, <http://matthewlincoln.net/2016/06/06/linked-open-realities-the-joys-and-pains-of-using-lod-for-research.html>.

the whole world real access to your SQL database”³⁰⁹, nel senso che i dati risultano spesso troppo complessi per essere riutilizzati da terzi, ad esempio in nuove applicazioni.

Un altro problema riguarda la qualità dei dataset, spesso soggetti a interruzioni o perdite o con informazioni poco affidabili, questo riguarda la fiducia e la credibilità dei dati che provengono da *repository* remoti³¹⁰, elementi tra l’altro presenti nelle rappresentazioni dell’architettura del web semantico³¹¹ sulla quale si stanno concentrando molte energie in termini di ricerca. “Fiducia” può avere molti significati, ma essa dipende principalmente dalla fonte da cui sono tratte le informazioni e dalle politiche adottate da chi detiene i dati³¹². È anche per questo che i dati delle biblioteche sono molto ambiti dalle comunità del web, sono infatti considerati semanticamente molto ricchi e strutturati, dati che possono in qualche modo aumentare il grado di fiducia generale del web.

Esistono anche problemi legati alla manutenzione degli URI, possono infatti esistere diversi identificativi univoci per una stessa entità o addirittura si può commettere l’errore di duplicare uno stesso URI inserendolo in più *triplestore*. Inoltre non sempre vengono stipulate le dovute relazioni tra i dati e questo incide sulla navigazione e sulla possibilità di scoprire altre risorse³¹³. Spesso poi non viene applicata la dereferenziazione agli URI, cioè non viene predisposta una pagina HTML visibile agli utenti con informazioni sui dati.

Il formato MARC, pur con tutte le sue limitazioni, ha comunque garantito la conservazione di informazioni semanticamente ricchissime, ed è importante che nella conversione dei dati si cerchi di non perdere questa specificità³¹⁴.

Infine non va sottovalutato il rapporto tra costi e benefici, elemento che dovrebbe essere soppesato dalle biblioteche che intendano convertire i propri dati in linked open data.

³⁰⁹ Dimitri Van Hees, «Why APIs are the missing ‘link’ in Linked Open Data», 2015, <https://www.linkedin.com/pulse/why-apis-missing-link-linked-open-data-dimitri-van-hees>.

³¹⁰ Lo Castro, «Web Semantico e Linked Open Data: best practices, prospettive e criticità».

³¹¹ Gordon, «Semantic Web Specification at W3C - Slide list».

³¹² Adis Medic e Adis Golubovic, «Making secure Semantic Web», *Universal Journal of Computer Science and Engineering Technology* 2, n. 1 (2010): 99–104.

³¹³ Trombone, «Cataloghi, modelli concettuali, data model».

³¹⁴ Ibid.

Berners-Lee ha proposto un sistema di valutazione della qualità dei dati aperti³¹⁵. Affinchè essi possano essere considerati linked open data a tutti gli effetti e ottenere il *rating* massimo, devono avere le seguenti caratteristiche:

- essere dati aperti;
- essere strutturati in formati leggibili da una macchina, utilizzando RDF e SPARQL;
- essere connessi ad altri dati attraverso l'*interlinking*.

Ma cosa comporta questo in termini di costi e benefici³¹⁶? Per avere dati aperti basta scegliere quale licenza applicare, convertire i dati in RDF invece comporta un investimento maggiore poiché sono necessarie professionalità che conoscano i linguaggi e le tecnologie del web semantico. Servono inoltre strumenti adeguati e la possibilità di investire il tempo necessario alla realizzazione di un progetto. Sempre nel rapporto di OCLC, il 61% delle istituzioni intervistate ha dichiarato di aver impiegato più di due anni per portare a termine la conversione e la pubblicazione dei dati³¹⁷.

Creare connessioni con altri dataset presuppone la conoscenza del proprio dominio, la capacità di valutare la qualità e la credibilità dei dataset esterni. L'*interlinking* spesso comporta la mappatura tra i diversi vocabolari che sono stati usati, e di conseguenza la gestione della disambiguazione tra metadati, anche in lingue diverse, e il loro allineamento³¹⁸. Va considerato poi il tempo necessario a gestire gli URI, che a volte non sono stabili e persistenti.

I benefici, come abbiamo visto, sono principalmente dati dalla possibilità di avere dati accessibili, riutilizzabili e modificabili, che consentono di poter essere annotati e arricchiti perché chiunque può aggiungere link e aumentare il valore di un'informazione. Per le biblioteche i linked open data rappresentano ancora una terra poco esplorata, valutare la loro reale portata richiede un'attenta analisi da parte della comunità bibliotecaria. Man mano che i progetti e le realizzazioni aumenteranno sarà possibile

³¹⁵ Berners-Lee, «Linked Data - Design Issues».

³¹⁶ «5-star Open Data», <http://5stardata.info/en/>.

³¹⁷ Smith-Yoshimura, «Analysis of International Linked Data Survey for Implementers».

³¹⁸ Ibid.

avere un quadro più chiaro sui vantaggi e sugli svantaggi di questa tecnologia, che appunto è solamente un mezzo, uno strumento, e non il fine delle biblioteche. Le biblioteche non devono semplicemente ridursi a convertire i propri dati in linked open data, ma devono valutare se questa tecnologia può garantire il raggiungimento degli obiettivi specifici di questo ambito. Nel caso dei dati bibliografici è importante capire quali sono i mezzi migliori per concretizzare gli indirizzi degli ICP. Per valutare appieno la portata dei linked open data, ai bibliotecari, ma in particolare di chi si occupa di catalogazione, è richiesto lo sforzo di avvicinarsi ad un linguaggio nuovo, a strumenti nuovi, cioè quelli proposti dal web semantico. Dal momento che tutte le nostre informazioni sono destinate ad essere sempre più gestite da software e ricercate attraverso la rete e saranno più o meno utili in base a sistemi di ricerca informatici, non possiamo ignorare queste tecnologie ³¹⁹.

³¹⁹ Coyle, *FRBR, Before and After*.

CAPITOLO QUARTO

Casi di applicazione dei linked open data

In questo capitolo vengono presi in esame tre casi di studio. La scelta è ricaduta sui progetti più completi e con maggiore visibilità a livello internazionale.

4.1. Europeana

Europeana è un progetto che utilizza linked open data per valorizzare il patrimonio culturale europeo³²⁰. Alla base vi è l'idea di raccogliere le collezioni di biblioteche, musei e archivi, attraverso uno schema comune di metadati, e utilizzare un'unica piattaforma di indicizzazione e accesso ai contenuti. Il portale è il servizio chiave della Fondazione Europeana, il cui obiettivo è l'aggregazione e la pubblicazione dei dati creati dalle istituzioni culturali e dalle comunità scientifiche. Attualmente, questa piattaforma, è uno dei servizi di metadati più maturi e consolidati a livello internazionale. La biblioteca digitale di Europeana contiene oltre cinquanta milioni di risorse, messe a disposizione da circa 2.500 istituti³²¹. Le risorse possono essere delle rappresentazioni digitali, ma anche gli oggetti veri e propri, nel caso essi siano originariamente digitali³²².

Il DEA (*Data Exchange Agreement*)³²³ è l'accordo che governa lo scambio di metadati tra Europeana e i diversi partner. Il DEA stabilisce che tutti i metadati aggregati e condivisi devono essere pubblicati sotto i termini previsti dalla Creative Commons Zero Public Domain Dedication; questo consente un'ampia distribuzione, così come la completa libertà di riutilizzare i metadati del patrimonio culturale. L'attribuzione CC0 permette in aggiunta che i metadati di Europeana vengano utilizzati nel contesto dei

³²⁰ «Europeana Collections», <http://www.europeana.eu/portal/?locale=en>.

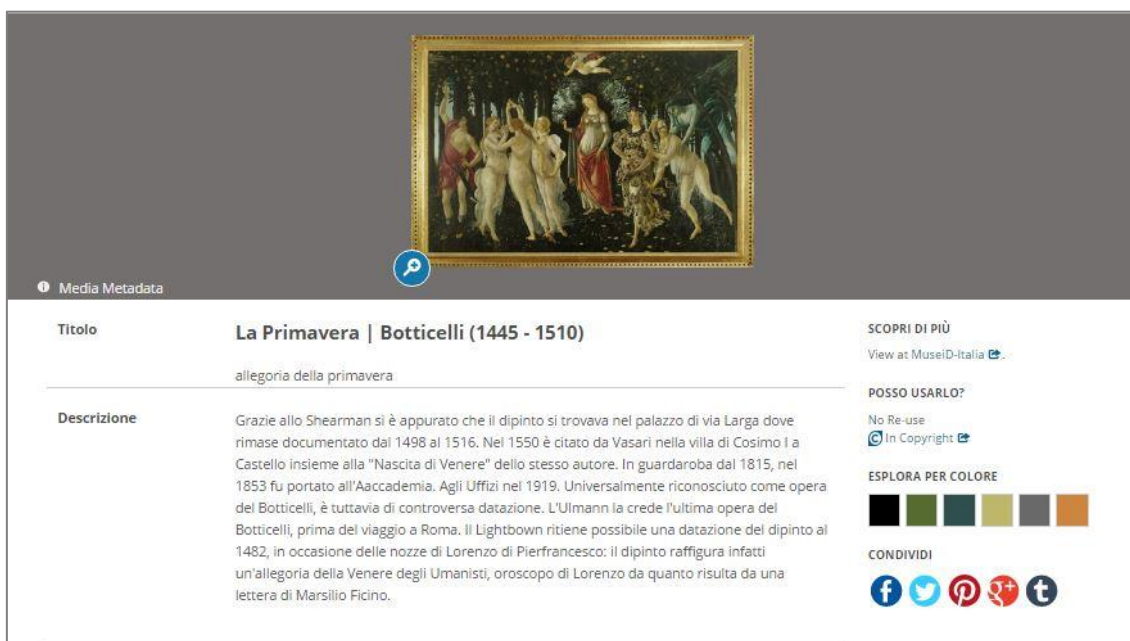
³²¹ «Europeana Strategy 2020: 'We transform the world with culture'», <http://strategy2020.europeana.eu/it/>.

³²² «Glossary of Terms - Europeana Professional», <http://pro.europeana.eu/page/page/glossary>.

³²³ «The Data Exchange Agreement - Europeana Professional», <http://pro.europeana.eu/the-data-exchange-agreement/>.

linked open data³²⁴. Il DEA prescrive che gli oggetti digitali e la loro anteprima, visualizzata nel portale, devono essere contrassegnati da termini d'uso stabiliti dalle istituzioni di provenienza. I metadati di Europeana sono pertanto liberamente fruibili, mentre gli oggetti digitali e le loro anteprime possono essere vincolate da restrizioni d'uso.

Nella Figura 1, sulla destra, si può vedere un'indicazione di copyright, in cui viene dichiarato che l'immagine non può essere riutilizzata.



The image shows a screenshot of a digital object page on the Europeana portal. At the top center is a thumbnail of the painting 'La Primavera' by Sandro Botticelli. Below the thumbnail is a blue circular icon with a white 'p' (Creative Commons license). The page is divided into sections: 'Titolo' (Title) with the text 'La Primavera | Botticelli (1445 - 1510)' and 'allegoria della primavera'; 'Descrizione' (Description) with a detailed text about the painting's history and provenance; 'SCOPRI DI PIÙ' (Discover more) with a link to 'View at MuseiD-Italia'; 'POSSO USARLO?' (Can I use it?) with 'No Re-use' and 'In Copyright' icons; 'ESPLORA PER COLORE' (Explore by color) with a color palette; and 'CONDIVIDI' (Share) with social media icons for Facebook, Twitter, Pinterest, and Tumblr.

Figura 1: Anteprima di un oggetto digitale nel portale di Europeana, tratto da http://www.europeana.eu/portal/it/record/2048011/work_63812.html?q=Sandro+Botticelli.

La biblioteca digitale di Europeana mette a disposizione dei dataset in forma di linked open data. Il repository offre circa cinquanta milioni di record, scaricabili tramite download dei dati e SPARQL endpoint³²⁵.

I metadati sono strutturati sulla base di EDM (Europeana Data Model), che utilizza RDF e gli schemi che si basano su una struttura a grafo. Lo sviluppo di questo profilo applicativo nasce per sopperire alle limitazioni di ESE (Europeana Semantic Elements)³²⁶,

³²⁴ «The Europeana Licensing Framework» (2013), http://pro.europeana.eu/files/Europeana_Professional/Publications/Europeana%20Licensing%20Framework.pdf.

³²⁵ «Data - Europeana Labs», <http://labs.europeana.eu/data>.

³²⁶ «The Europeana Data Model», (2010), <http://www.ifla.org/past-wlic/2010/149-doerr-en.pdf>.

il primo modello di Europeana. ESE, infatti, presentava un approccio piatto alla modellizzazione dei dati, poiché non consentiva l'inclusione di link a risorse esterne presenti nel web, e aveva poca estensibilità verso modelli più specialistici e granulari³²⁷.

Tra i problemi che si volevano risolvere con EDM vi era la difformità tra i metadati relativi a più tipologie di oggetti, provenienti da istituzioni con tradizioni descrittive differenti. Serviva perciò un modello che integrasse ed ospitasse vari standard. L'approccio del web semantico, da questo punto di vista, consente di poter combinare in una stessa descrizione elementi di più ontologie e vocabolari³²⁸.

EDM ha tre principali finalità³²⁹:

- distinguere tra l'oggetto di una istituzione e la sua rappresentazione digitale;
- distinguere l'oggetto dai metadati che lo descrivono;
- consentire l'acquisizione di record multipli afferenti lo stesso oggetto, con la possibilità quindi di raccogliere diverse asserzioni su di esso.

In questo modo è possibile rappresentare le complesse relazioni che intercorrono tra i *providers*, gli oggetti digitali e i metadati che li descrivono.

Europeana fornisce molti documenti relativi al suo modello³³⁰; uno dei più importanti è l'*Europeana Data Model Mapping Guidelines*³³¹, nel quale vengono specificate le classi e le proprietà riprese da ontologie e vocabolari esistenti. In particolare EDM fa riferimento a:

- OAI-ORE (Open Archives Object Reuse and Exchange Model)³³²: per organizzare i metadati degli oggetti e le loro rappresentazioni digitali³³³. ORE è uno standard definito dall'Open Archives Initiative, che serve per la descrizione e lo scambio di insiemi di risorse web. Europeana utilizza due termini di questo modello: *Aggregation*, che aggrega e identifica l'oggetto culturale e la sua rappresentazione

³²⁷ «The Europeana Data Model»

³²⁸ Doerr et al., «The Europeana Data Model».

³²⁹ Isaac, Clayphan, Robina, e Haslhofer, «EUROPEANA: Moving to Linked open data».

³³⁰ «Europeana Data Model Documentation», <http://pro.europeana.eu/page/edm-documentation>.

³³¹ «Europeana Data Model – Mapping Guidelines v2.3», (2016), http://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Mapping_Guidelines_v2.3_112016.pdf.

³³² «Open Archives Initiative - Object Exchange and Reuse», <https://www.openarchives.org/ore/>.

³³³ «Europeana Data Model Documentation».

- digitale³³⁴, e *Proxy*, che consente di mantenere la provenienza di una risorsa, o insiemi di risorse³³⁵;
- Dublin Core³³⁶: per i metadati descrittivi. Il modello EDM presenta diversi elementi ripresi da DC Elements e DC Terms, come le proprietà *dc:subject*, *dc:contributor*, *dcterms:created* e *dcterms:alternative*³³⁷;
 - SKOS³³⁸: per la rappresentazione di vocabolari riguardanti i concetti, di cui EDM usa la classe principale, ovvero *skos:Concept*³³⁹;
 - FOAF³⁴⁰: per descrivere persone;
 - CIDOC CRM (Conceptual Reference Model)³⁴¹: per gli eventi e le relazioni tra oggetti. Il CIDOC CRM è uno standard ISO che fornisce una definizione delle entità e una struttura formale multi-livello, per creare connessioni tra oggetti fisici, eventi e agenti (persone o organizzazioni). EDM allinea alcune delle sue classi e proprietà a quelle di CIDOC CRM, ovvero la classe *edm:Event* come equivalente di *E4 Period*, la classe *edm:InformationResource* come sottoclasse di *E73 Information Object*, e la proprietà *edm:wasPresentAt* come equivalente di *P12 occurred in the presence of (was presenta t)*³⁴².
 - FRBR³⁴³: utilizzato in modo molto limitato, nonostante si dichiara esplicitamente di farne uso³⁴⁴. I riferimenti a FRBR si trovano per la classe *edm:InformationResource* che viene definita come un'unione di *FRBR Work*, *FRBR Expression* e *FRBR Manifestation*, schiacciando in questo modo la

³³⁴ «The Europeana Data Model».

³³⁵ «Europeana Data Model Primer», (2013),

http://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf.

³³⁶ «DCMI Home: Dublin Core Metadata Initiative (DCMI)», <http://dublincore.org/>.

³³⁷ Silvio Peroni, Francesca Tomasi, e Fabio Vitali, «Reflecting on the Europeana Data Model», in *Digital Libraries and Archives*, a c. di Maristella Agosti et al., vol. 354 (Berlin: Springer, 2013), 228–40, http://link.springer.com/10.1007/978-3-642-35834-0_23.

³³⁸ «SKOS Simple Knowledge Organization System», <https://www.w3.org/2004/02/skos/>.

³³⁹ «Definition of the Europeana Data Model v5.2.7», (2016),

http://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation//EDM_Definition_v5.2.7_042016.pdf.

³⁴⁰ «FOAF Vocabulary Specification».

³⁴¹ «CIDOC CRM», <http://www.cidoc-crm.org/>.

³⁴² Peroni, Tomasi, e Vitali, «Reflecting on the Europeana Data Model»; «Definition of the Europeana Data Model v5.2.7».

³⁴³ «IFLA -- Functional Requirements for Bibliographic Records»,

<http://www.ifla.org/publications/functional-requirements-for-bibliographic-records>.

³⁴⁴ «Definition of the Europeana Data Model v5.2.7»; «Europeana Data Model Primer».

gerarchia di FRBR. Le classi *edm:Event* ed *edm:Place* sono indicate invece come equivalenti di *FRBR Event* e *FRBR Place*.

Il modello EDM è strutturato in modo che i diversi partecipanti non debbano modificare la struttura e le funzioni originarie dei propri dati. Europea infatti agisce a livello interno, aggregando ed arricchendo le informazioni che riceve. Il modello prevede tre classi principali:

- *edm:ProvidedCHO* : l'oggetto culturale (ad esempio un dipinto);
- *edm:WebResource*: una o più rappresentazioni digitali accessibili di questo oggetto (l'immagine digitale del dipinto);
- *ore:Aggregation*: un'aggregazione che rappresenta l'insieme degli elementi forniti.

I primi due elementi servono a distinguere l'*opera* dalla sua *rappresentazione digitale*. Il terzo elemento invece stabilisce che l'oggetto fornito e la sua rappresentazione digitale formano un'unica entità logica, che può essere vista come un tutto composto di singole parti.

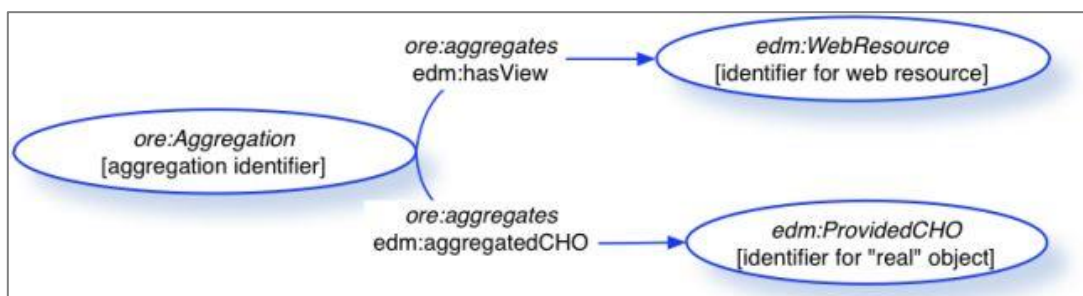


Figura 2: Classi principali di EDM, tratto da http://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation//EDM_Primer_130714.pdf.

Un aspetto importante di Europea è l'arricchimento dei dati ricevuti. Un procedimento automatico associa delle stringhe di testo, trovate nei metadati, con dei termini appartenenti a dataset o vocabolari strutturati come linked open data. La valorizzazione dei metadati con termini aggiuntivi viene chiamata *arricchimento semantico*. In pratica si crea un *match* tra i metadati della risorsa di Europea e dati semantici esterni, che si concretizza in un link che punta a risorse esterne³⁴⁵. La generazione di linked open data

³⁴⁵ «Europeana semantic enrichment - Europea Professional», <http://pro.europeana.eu/page/europeana-semantic-enrichment>.

riguarda la creazione di collegamenti a diversi vocabolari. I nomi geografici vengono associati a GeoNames³⁴⁶, i soggetti a GEMET³⁴⁷, un thesaurus multilingue dedicato all'ambiente, DBpedia³⁴⁸ invece fornisce informazioni su persone, luoghi e organizzazioni. In aggiunta Europeana crea link anche con Wikidata³⁴⁹, i cui contenuti rientrano nel pubblico dominio³⁵⁰.

Europeana rende possibile l'accesso a delle API, per cercare e recuperare i dati. Le interrogazioni sui linked open data vengono fatte attraverso uno SPARQL endpoint, che restituisce i dati in RDF/XML o JSON³⁵¹. Questo tipo di apertura, definita "sperimentale", permette lo scarico completo di interi dataset e ricerche semantiche più complesse³⁵². Lo scopo è quello di coinvolgere attori esterni nella creazione di applicazioni, così da valorizzare le risorse culturali. La sezione *Incubation*, presente sul sito di Europeana, è nata proprio per facilitare il processo di sviluppo di nuovi prodotti, idee o strumenti, sempre basati sul riutilizzo del patrimonio culturale³⁵³. Le applicazioni realizzate sono moltissime³⁵⁴ e dimostrano il beneficio che può apportare la condivisione di dati aperti, facilmente utilizzabili attraverso l'uso di API.

Art Faces ad esempio è un'applicazione *mobile* sviluppata da una start-up italiana³⁵⁵. Questo gioco educativo si rivolge a bambini tra i 5 e i 10 anni e cerca di promuovere la conoscenza dell'arte attraverso l'interazione con famosi dipinti, i cui metadati sono stati messi a disposizione da Europeana.

³⁴⁶ «GeoNames», <http://www.geonames.org/>.

³⁴⁷ «Themes list», <http://www.eionet.europa.eu/gemet/>.

³⁴⁸ «DBpedia», <http://wiki.dbpedia.org/>.

³⁴⁹ «Wikidata», https://www.wikidata.org/wiki/Wikidata:Main_Page.

³⁵⁰ «Europeana Data Sources - Europeana Collections», <http://www.europeana.eu/portal/it/rights/data-sources.html>.

³⁵¹ «SPARQL Endpoint - Europeana Labs», <http://labs.europeana.eu/api/linked-open-data-SPARQL-endpoint>.

³⁵² «Introduction to the Europeana APIs - Europeana Labs», <http://labs.europeana.eu/api>.

³⁵³ «Incubation - Europeana Labs», <http://labs.europeana.eu/incubation>.

³⁵⁴ «Apps showcase - Europeana Labs», <http://labs.europeana.eu/apps>.

³⁵⁵ «Art Faces - Europeana Labs», <http://labs.europeana.eu/apps/art-faces>.



Figura 3: Art Faces, tratto da <http://labs.europeana.eu/apps/art-faces>.

Gli utenti possono esplorare il patrimonio culturale anche attraverso il portale, il principale punto di accesso alle informazioni di Europeana. La ricerca semplice è supportata dalle faccette, che dovrebbero aiutare a filtrare le risorse. In realtà non è presente la possibilità di navigare o selezionare elementi semantici. I filtri disponibili sono infatti: collezione, *media* dell'oggetto, tipologia di restrizione, paese fornitore, lingua, aggregatore e istituzione. Una volta scelto il record, la visualizzazione presenta diversi problemi. I metadati vengono tradotti in base alla lingua che si seleziona dal portale, mentre i dati conservano la lingua d'origine. Nel record della Figura 4, tutti gli elementi descrittivi e semantici non presentano una funzione multilingue.

Titolo	Kaj pa ti pobič, waltz, ljudska
Persone	Contributore: France Štrukelj (pripreditelj) , Slovenski moški kvartet Prešeren (izvajalec)
Classificazioni	Tipo: zvočni posnetki Soggetto: narodnozabavna glasba , ok. 1940 , slovenija , slovenski izseljenci , valčki
Proprietà	Lingua: slv
Tempo	Data: http://semium.org/time/19xx_2_third Data Di Creazione: 1940 , http://semium.org/time/1940 Periodo: 1940 Rilasciato: 1940 , http://semium.org/time/1940 Temporale: http://semium.org/time/19xx
Provenienza	Provenienza: zvočni posnetki Editore: Columbia broadcasting system Identificatore: URN:URN:NBN:SI:snd-Z71V01M1 Fornitori Di Contenuti Or Dati: National and University Library of Slovenia Fornitore: Slovenian National E-content Aggregator Paese Fornitore: Slovenia Pubblicato In Europeana: 2016-09-29 L'ultimo Aggiornamento Da Europeana: 2016-09-29

Figura 4: Visualizzazione di un record nel portale di Europeana, tratto da http://www.europeana.eu/portal/it/record/92056/URN_NBN_SI_snd_Z71V01M1.html?q=.

Riguardo i dati semantici, EDM prevede la creazione di relazioni tra un concetto e le varie etichette che lo possono rappresentare³⁵⁶, come tra *skos:Concept* e *skos:CloseMatch*, in modo da garantire la presentazione di un concetto in diverse lingue. Ma probabilmente questa implementazione non è ancora matura. Sempre nella Figura 4, sono visibili dati espressi come linked open data, cioè HTTP URI, ad esempio quelli presenti nella sezione “tempo” del record. Se l’utente dovesse provare a cliccare sul link presente nel campo “data”, si ritroverebbe davanti l’immagine della Figura 5, un ambiente che non presenta nessuna descrizione dell’elemento e nessuna possibilità di continuare a navigare in modalità *follow your nose*.

³⁵⁶ «EDMObjectTemplatesProviders · europeana/corelib Wiki · GitHub», <https://github.com/europeana/corelib/wiki/EDMObjectTemplatesProviders#skosConcept>.

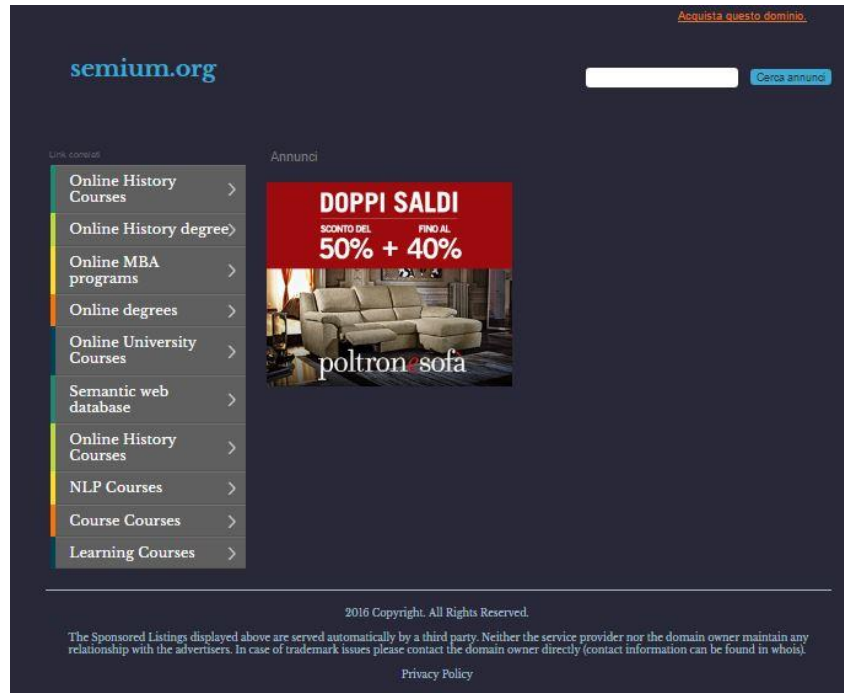


Figura 5: Elemento espresso in HTTP URI in un record di Europeana, tratto da http://semium.org/time/19xx_2_third.

Gli HTTP URI in un documento RDF/XML non devono essere confusi con l'HTTP URI di un indirizzo web. In un documento RDF/XML essi sono solamente nomi per le entità. Le informazioni che vengono fornite attraverso la dereferenziazione dovrebbero però essere presentate in una forma comprensibile e utile anche per gli esseri umani, altrimenti si rischia di presentare agli utenti dati predisposti solo per le macchine. Questo può disorientare o creare frustrazione in chi naviga tra le risorse.

Riprendiamo l'analisi delle possibilità di ricerca: le relazioni di similarità tra record, che dovrebbero rispondere alla domanda "che cosa c'è di simile a questo oggetto nel catalogo?", sono principalmente basate sull'appartenenza ad uno stesso dataset. Non sono presenti legami interni più profondi e complessi, come tra opere originali ed opere derivate. La descrizione del record che abbiamo preso in esame, presente nella Figura 4, è relativa ad una registrazione sonora che proviene dal dataset 92056_Ag_Slovenia_music_records. I suggerimenti che vengono forniti a partire da questa descrizione appartengono infatti tutti allo stesso dataset, come visibile nella Figura 6. I suggerimenti non prendono quindi in considerazione l'ambito semantico, e la conseguente possibilità di navigare in modo facilitato a partire dai contenuti, così come tutte le relazioni che si possono instaurare tra le varie entità.



Figura 6: Relazioni di similarità tra record di Europeana, tratto da http://www.europeana.eu/portal/it/record/92056/URN_NBN_SI_snd_Z71V01M1_.html?q=.

Nonostante Europeana arricchisca le proprie risorse con elementi esterni, in questo momento mancano relazioni semanticamente significative per l'utente, sia tra i dati e le descrizioni interne, che tra questi elementi e il web. La navigazione spesso si interrompe e l'utente non riesce ad esplorare agevolmente i contenuti culturali e concettuali, a partire dai dati. Gestire importanti quantità di informazioni, uniformarne la presentazione e migliorare le modalità di scoperta, sono processi complessi che richiedono di frequente una continua revisione dei modelli di dati e della loro struttura.

Europeana si è imposta come uno dei più importanti progetti di implementazione di linked open data. Molte applicazioni sono state create grazie a questa scelta, coinvolgendo attori esterni che sono riusciti a valorizzare i dati in modo originale. Le applicazioni si avvicinano realmente alle nuove esigenze informative degli utenti, che spesso consultano il web in modalità *mobile*. Ma la tecnologia che viene usata per migliorare la fruizione delle risorse dovrebbe essere invisibile, proprio come avviene nella maggior parte degli OPAC di biblioteca, in cui la visualizzazione della notizia "nasconde" la struttura dei campi MARC.

4.2. Digital Public Library fo America

La visione di una biblioteca nazionale americana nasce intorno al 1990³⁵⁷ grazie alle proposte di bibliotecari, studenti, educatori e società private. Gli sforzi fatti da istituzioni come la Library of Congress, la HathiTrust e Internet Archive, hanno consentito l'accesso, tramite internet, a milioni di risorse. Inoltre, negli anni, molte università, biblioteche pubbliche e altre istituzioni culturali, hanno digitalizzato i propri materiali, con l'inconveniente però che queste preziose risorse risiedevano in ambienti separati, obbligando gli utenti ad accedere di volta in volta a portali diversi³⁵⁸. La Digital Public Library of America (DPLA) è nata con lo scopo di convogliare tutti questi differenti punti di vista, esperienze e collezioni, in un unico portale.

Il progetto della DPLA comincia nel 2010 quando, durante un incontro a Cambridge, 40 responsabili di biblioteche, fondazioni, università e società tecnologiche decidono di lavorare insieme per costruire

an open, distributed network of comprehensive online resource that would draw on the nation's living heritage from libraries, universities, archives, and museums in order to educate, inform, and empower everyone in current and future generations³⁵⁹.

Riguardo la gestione delle licenze, anche la Digital Public Library of America distingue tra l'oggetto digitale, la sua anteprima e i metadati che lo descrivono. Mentre i primi due elementi possono essere soggetti a restrizioni, i metadati che le istituzioni decidono di inviare alla piattaforma di aggregazione, rientrano nel pubblico dominio, come espresso dalla CC0 Public Domain Dedication³⁶⁰.

I principali prodotti di questa iniziativa sono un portale, per la scoperta e l'accesso alle risorse, e una piattaforma che consente l'interrogazione e il recupero dei dati. Il sito è orientato agli utenti, mentre le API sono l'ingresso ai dati aggregati, pensato per chi vuole compiere operazioni più complesse, attraverso l'utilizzo di specifiche tecnologie.

³⁵⁷ «Digital Public Library of America » History», <https://dp.la/info/about/history/>.

³⁵⁸ Ibid.

³⁵⁹ Ibid.

³⁶⁰ «The Digital Public Library of America Policy Statement on Metadata», (2013), <https://dp.la/info/wp-content/uploads/2013/04/DPLAMetadataPolicy.pdf>.

DPLA ha sviluppato un profilo di applicazione, relativo ai propri metadati, il DPL MAP (Metadata Application Profile), ispirato all'esperienza di Europeana³⁶¹. In aggiunta al modello EDM, il gruppo di lavoro della Digital Public Library of America ha tenuto in considerazione i suggerimenti provenienti dalla comunità di riferimento, per creare un ambiente che consentisse di ospitare schemi già esistenti ed emergenti nel settore delle biblioteche, degli archivi e dei musei.

L'ultima versione del DPL MAP incorpora diversi *namespace*, con classi relative ai luoghi geografici, ai concetti e agli agenti. Elementi che traggono beneficio dall'uso dei linked open data³⁶², in particolare per la creazione dell'*interlinking*.

DPL MAP oltre a basarsi su EDM è allineata a DC Metadata Terms³⁶³. Questa scelta ha consentito di creare maggiore interoperabilità con gli schemi di metadati utilizzati dalla maggior parte dei partner di DPLA.

I principali *namespace* di riferimento del modello sono:

- DPLA: classi o proprietà create appositamente per DPLA;
- RDF e RDFS: Resource Description Framework e RDF Schema;
- DCTerm: Dublin Core Metadata Terms;
- DC: Dublin Core Metadata Element Set;
- DCMITYPE: Dublin Core Media Type;
- EDM: Europeana Data Model;
- GN: Geonames;
- ORE: Open Archive Initiative Object Reuse and Exchange;
- SKOS: Simple Knowledge Organisation System.

La descrizione delle entità avviene tramite l'individuazione di classi. Ogni classe può avere diverse proprietà, basate su elementi preesistenti di Dublin Core o Europeana Data Model. Nella Figura 7 viene illustrato il modello.

³⁶² «Digital Public Library of America» Policies», <https://dp.la/info/about/policies/>.

³⁶³ «DCMI Metadata Terms», <http://dublincore.org/documents/dcmi-terms/>.

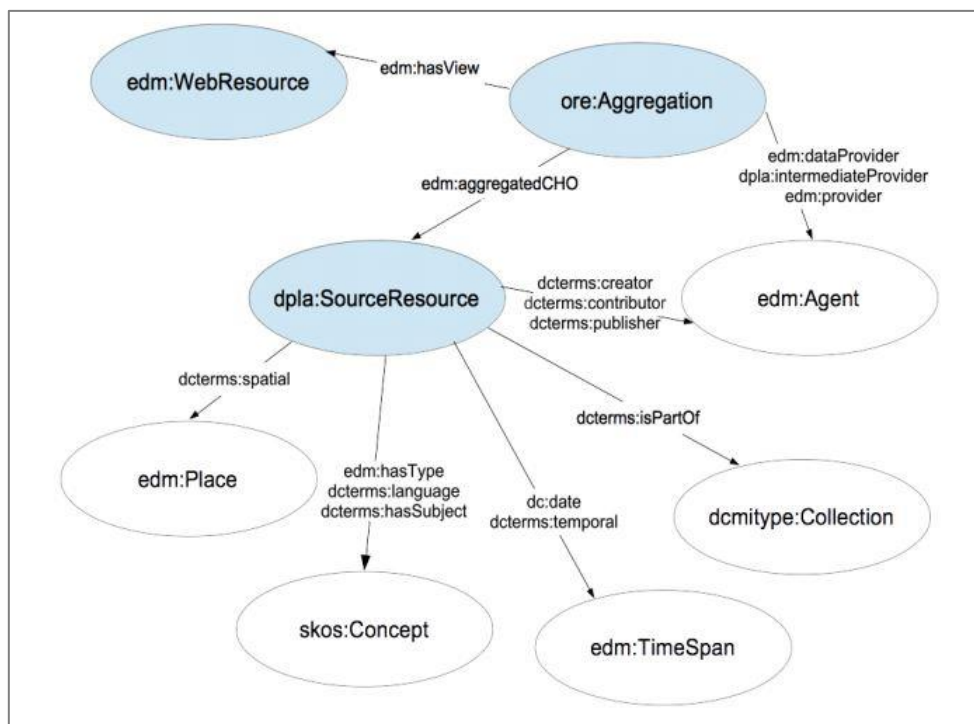


Figura 7: DPLA Metadata Model, tratto da [https://dp.la/info/wp-content/uploads/2017/01/Intro to DPLA metadata model 01062017.pdf](https://dp.la/info/wp-content/uploads/2017/01/Intro%20to%20DPLA%20metadata%20model%2001062017.pdf).

La classe *SourceResource* contiene molte proprietà di tipo descrittivo (titolo, data, formato, etc.) afferenti alla risorsa originale, cioè la fonte da cui vengono tratti questi dati. Questa è l'entità che presenta la maggior parte dei dati visualizzati da un utente sul portale. Quasi tutte queste proprietà utilizzano elementi di Dublin Core. La *SourceResource* è collegata alla sua versione digitale, la *WebResource*, grazie al tramite fornito dalla *ore:Aggregation*. Quest'ultima classe contiene informazioni importanti sui partner di DPLA, sulle istituzioni che inviano i dati e sulla localizzazione del record di origine. In essa si trova anche un link al dataset di provenienza dei metadati.

Il modello di metadati è una rappresentazione astratta, di conseguenza può essere espresso in diversi standard processabili dalle macchine. In particolare JSON-LD e RDF/XML³⁶⁴. Questi standard sono stati scelti proprio per la loro capacità di gestire linked open data. Tali tecnologie consentono infatti di sincronizzare i dati con record di autorità preesistenti. Utilizzando gli URI di riferimento, presenti in file di autorità come quelli di VIAF (Virtual Authority Record) e della Library of Congress, si ereditano le informazioni sulle risorse, che possono essere in altre lingue oppure consentire la disambiguazione di

³⁶⁴ «An introduction to the DPLA Metadata Model».

omonimi. Ad esempio se si inserisce in un record l'URI <http://id.loc.gov/authorities/subjects/sh97009149>, invece della stringa testuale *Snowmen* (uomini di neve), si recuperano le annotazioni semantiche predisposte dalla Library of Congress, che consentono di distinguere questo termine dal film di animazione *The snowman* del 1982.

Anche la Digital Public Library of America prevede l'arricchimento semantico dei dati. Generalmente viene stabilito un link alle forme standardizzate di luoghi, presenti in GeoNames, e di persone, presenti in VIAF. Attualmente la DPLA crea collegamenti solamente ai nomi di luoghi, mentre sono i partner che partecipano al progetto, a fornire dati linkati a soggetti o nomi di persona ed ente, in forma di record di autorità identificati da HTTP URI. In aggiunta è presente un servizio di pulizia dei dati, che comprende la rimozione di punteggiatura estranea e la normalizzazione dei formati delle datazioni³⁶⁵.

I dati sono conservati internamente nel formato RDF/XML, ma vengono resi disponibili attraverso API in JSON-LD. Tutti gli interessati possono presentare idee per valorizzare le risorse e creare applicazioni e strumenti, che vengono poi presentati in una apposita pagina del portale. Le applicazioni sono moltissime³⁶⁶. Color Browse, ad esempio, consente di esplorare le collezioni digitali a partire dalle diverse sfumature o tonalità contenute nelle immagini.

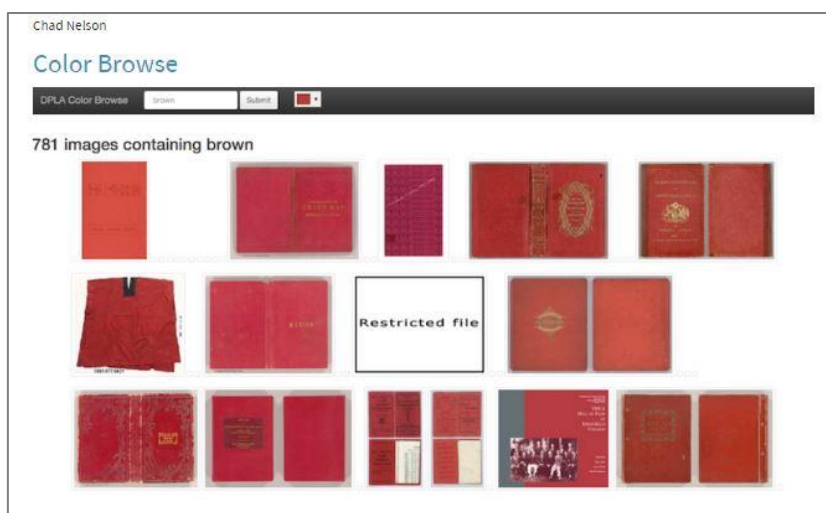


Figura 8: Color Browse, tratto da <https://dp.la/apps/33>.

³⁶⁵ «An introduction to the DPLA Metadata Model»

³⁶⁶ «App Library · Digital Public Library of America», <https://dp.la/apps>.

Metadata Explorer consente di navigare in diverse biblioteche digitali, tra cui DPLA, Europeana, Harvard e la Digital New Zealand. Dopo aver digitato nell'apposita stringa i termini di ricerca, tutti i metadati che vi sono associati appaiono come una nuvola formata da varie etichette, contrassegnate da colori diversi in base alla tipologia.

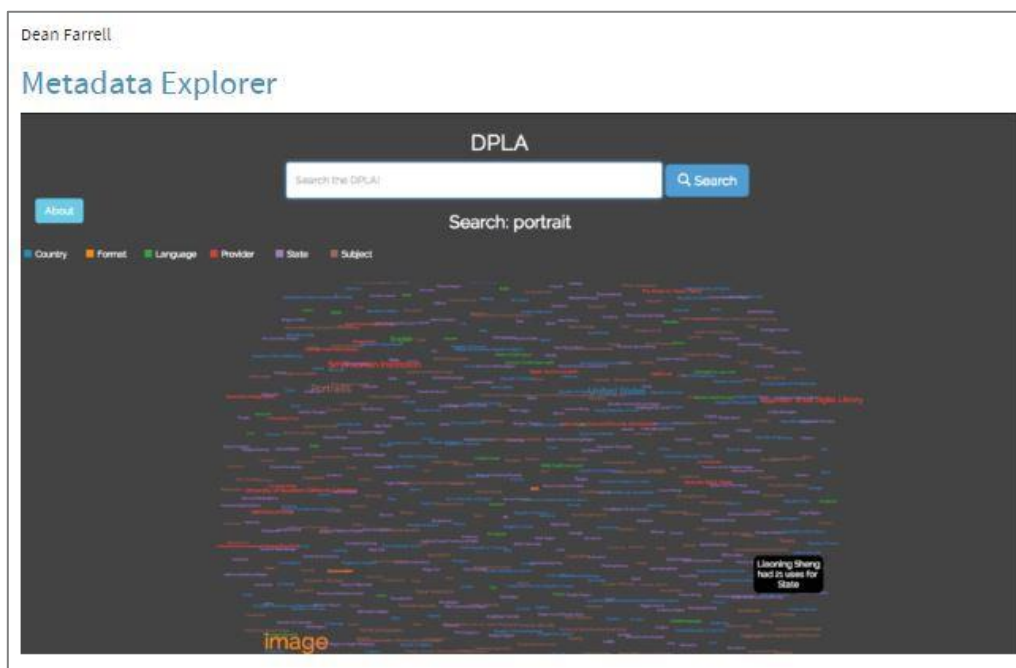


Figura 9: Metadata Explorer, tratto da <https://dp.la/apps/32>.

Cliccando sul metadato di interesse si accede direttamente alle risorse delle biblioteche.

Il principale punto di accesso ai dati è costituito dal portale, come per Europeana. La disposizione delle informazioni è infatti simile. Le risorse possono essere visionate attraverso una ricerca semplice, oppure esplorando le mappe, le *timeline*, le applicazioni o le mostre virtuali. I record presentano informazioni più omogenee rispetto ad Europeana, in quanto gli elementi in forma di URI vengono nascosti all'utente: questo agevola la comprensione e la lettura delle notizie. I campi che si visualizzano sono quelli classici (autore, titolo, soggetto, lingua, tipo di oggetto, restrizioni, etc.) e cliccando sugli elementi di autorità si resta all'interno della biblioteca digitale. A partire dai dati di un record non sembra possibile navigare anche all'esterno.

4.3. Online Computer Library Center

OCLC (Online Computer Library Center) è una rete cooperativa di biblioteche senza fini di lucro, nata nel 1976 e con sede a Dublin in Ohio³⁶⁷. Attualmente comprende 16.937 membri, afferenti a 120 paesi, con biblioteche di varia natura e tipologia³⁶⁸. I principali servizi si concentrano su soluzioni destinate alla gestione bibliotecaria, alla catalogazione, e agli strumenti di accesso alle risorse, comprese biblioteche digitali e sistemi di condivisione dell'informazione³⁶⁹.

All'interno di OCLC vi è il dipartimento *Research*³⁷⁰, uno dei più importanti centri di ricerca al mondo, che si focalizza sull'evoluzione di biblioteche e archivi in riferimento ai cambiamenti degli ambienti informativi. Tra gli ambiti di cui OCLC *Research* si occupa vi è l'applicazione dei linked open data ai dati bibliografici. La strategia adottata si concretizza nella creazione di nuovi servizi.

Uno degli strumenti più conosciuti è VIAF. Un aggregatore che combina diversi record di autorità in un unico servizio. Il sistema estrae e collega gli *authority file* delle biblioteche nazionali e li raggruppa in un "super" authority record, che tiene insieme i diversi nomi di quell'entità. OCLC ha elaborato un algoritmo specifico (*matching algorithm*) che "consente di riconoscere i record relativi alla medesima entità per effettuare in modo quasi automatico i collegamenti tra i dati estratti"³⁷¹.

Attualmente VIAF presenta dati relativi a persone, enti, opere, espressioni, luoghi geografici, e titoli di pubblicazioni. L'estensione delle entità, inizialmente limitate ai nomi di persona, rientra nel processo di FRBRizzazione del sistema. Il funzionamento dell'algoritmo dipende da molti fattori, ma in prima battuta dalla qualità dei dati di provenienza³⁷². Nel caso manchino dati aggiuntivi, VIAF collega alcuni record utilizzando solamente il nome dell'autore, pratica che produce *match* inaffidabili. Inoltre se i partner uniscono in unico record di autorità due omonimi si possono avere ricadute

³⁶⁷ «About OCLC», <https://www.oclc.org/en/about.html>.

³⁶⁸ Ibid.

³⁶⁹ Ibid.

³⁷⁰ «Research», <http://www.oclc.org/research.html>.

³⁷¹ Paul Gabriele Weston, «Authority data», in *Biblioteche e biblioteconomia: principi e questioni* (Roma: Carocci, 2015), 298.

³⁷² «Annual report: VIAF Council meeting 2015», (2015), <http://www.oclc.org/content/dam/oclc/viaf/viafc-2015-annual-report.pdf>.

su tutti i dati legati a quelle entità³⁷³. Le istituzioni che partecipano a questo progetto devono quindi fare attenzione alla qualità dei dati, poiché informazioni insufficienti inficiano il funzionamento dell'algoritmo e di conseguenza la presentazione di dati corretti.

Il dataset di VIAF è disponibile con licenze ODC-by (Open Data Commons Attribution License)³⁷⁴, in questo caso chi utilizza i dati deve dichiararne la fonte ed esplicitarne la titolarità. Attraverso un'API i dati possono essere scaricati sia in forma di *cluster* che di singoli record; il *cluster* aggrega tutti gli *autohority file* di una determinata entità, mentre i singoli record sono quelli inviati dalle istituzioni, espressi in MARC21 o UNIMARC³⁷⁵.

Il modello RDF di VIAF si è evoluto negli anni presentando diversi aggiornamenti e modifiche. Inizialmente era stato adottato esclusivamente SKOS, che però non possiede la nozione di persona³⁷⁶, questa mancanza ha portato all'aggiunta di alcune classi di FOAF, come *foaf:Person* e *foaf:Organization*. Successivamente OCLC, dopo un'attenta valutazione dei modelli disponibili per la creazione di linked open data, decide di utilizzare principalmente Schema.org, nato dall'accordo tra Google, Yahoo!, Bing e Yandex, per codificare in RDF i dati contenuti nelle pagine HTML³⁷⁷. Quasi tutte le classi e le proprietà inizialmente utilizzate sono state dunque modificate con quelle proposte da Schema.org.

³⁷³ «Annual report: VIAF Council meeting 2015»

³⁷⁴ «VIAF (Virtual International Authority File) | OCLC Developer Network», <http://www.oclc.org/developer/develop/web-services/viaf.en.html>.

³⁷⁵ Ibid.

³⁷⁶ Giulia Manzotti, «Analisi e riflessioni sul VIAF, Virtual International Authority File», *JLIS.it* 1, n. 2 (2010): 357–81.

³⁷⁷ Giovanni Bergamin, «Metadati e open data: nuovi paradigmi per vecchie professioni», (2012), https://www.openstarts.units.it/dspace/bitstream/10077/9360/1/Bergamin_MAB_2012.pdf.

Class Replacements	
Current	Changing to
http://dbpedia.org/ontology/Place	http://schema.org/Place
http://rdvocab.info/uri/schema/FRBRentitiesRDA/CorporateBody	http://schema.org/Organization
http://xmlns.com/foaf/0.1/Organization	http://schema.org/Organization
http://rdvocab.info/uri/schema/FRBRentitiesRDA/Person	http://schema.org/Person
http://xmlns.com/foaf/0.1/Person	http://schema.org/Person

Class Additions
http://schema.org/CreativeWork
http://bibliograph.net/Agent
http://purl.org/vocab/frbr/core#Expression
http://purl.org/vocab/frbr/core#Work
http://www.w3.org/2006/gen/ont#InformationResource

Figura 10: Aggiornamenti del vocabolario VIAF RDF, tratto da <https://www.oclc.org/developer/news/2014/viaf-rdf-vocabulary-changes-and-additions.en.html>.

In un post del 2014 OCLC pubblica le principali modifiche del vocabolario di VIAF in RDF. Diverse classi prese dall'ontologia di DBpedia, dal vocabolario di RDA e da FOAF vengono sostituite con elementi di Schema.org. In particolare, l'introduzione di questo schema allarga il panorama dei possibili fruitori dei dati e migliora l'integrazione con Wikidata³⁷⁸, un database che contiene i dati strutturati di Wikipedia in forma di linked open data.

OCLC utilizza linked open data anche per altri servizi. WorldCat, ad esempio, è uno strumento di accesso alle raccolte delle biblioteche aderenti. Tutte le informazioni create vengono condivise all'interno di questo portale. Nel 2012 OCLC decide di convertire in RDF circa trecento milioni di record catalografici in formato MARC³⁷⁹. Il principale vocabolario che viene scelto è, anche in questo caso, Schema.org. Un interessante sviluppo è la creazione di servizi di arricchimento dei dati, a partire dalle informazioni che si trovano nei record bibliografici. Avendo come base concettuale il modello FRBR, il gruppo di ricerca di OCLC ha estratto dai metadati delle notizie bibliografiche gli elementi che si riferiscono alle entità *opera*, *persona*, *ente*, *soggetto*, *luogo* ed *evento*. In

³⁷⁸ «Wikidata: Introduzione - Wikidata», <https://www.wikidata.org/wiki/Wikidata:Introduction/it>.

³⁷⁹ Carol Jean Godby, «A Division of Labor: The Role of Schema.org in a Semantic Web Model of Library Resources», in *Linked Data for Cultural Heritage* (London: Facet, 2016), 73–103.

pratica attraverso un processo di *data mining* si estraggono questi riferimenti e si creano delle descrizioni arricchite degli oggetti che rappresentano³⁸⁰. Questo progetto prende il nome di *WorldCat Entities*. Di un autore, ad esempio, vengono fornite informazioni sulle opere che ha scritto e i loro generi di appartenenza, ma anche sulle opere che riguardano quell'autore, il tipo di *audience*, e le entità che sono correlate in qualità di traduttori, curatori o autori.

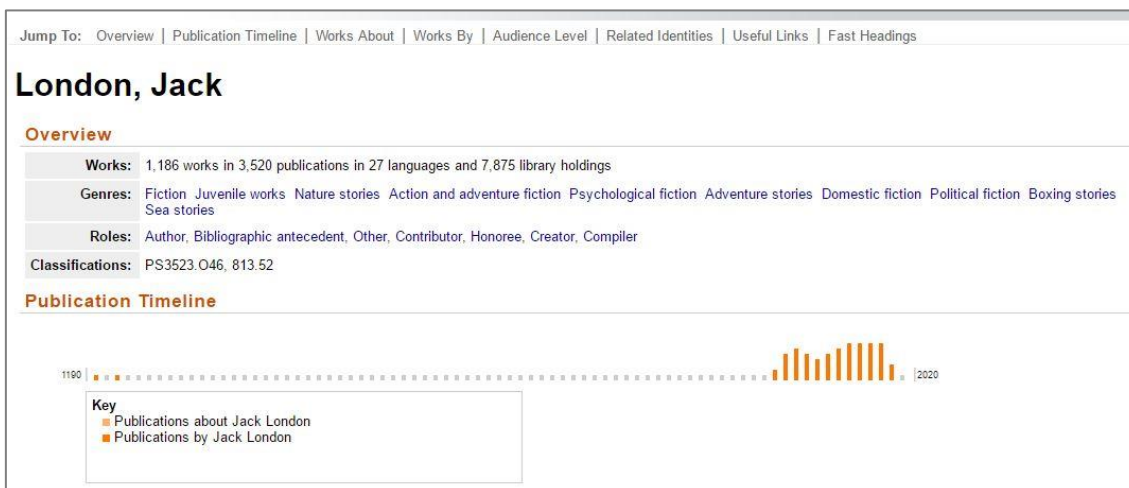


Figura 11: WorldCat Entities, tratto da <https://www.worldcat.org/wcidentities/np-london,%20jack#identitieslinks>.

Nella Figura 12 è presente anche una *timeline*, che consente di visualizzare graficamente la cronologia delle pubblicazioni di quell'autore.

OCLC sta inoltre sperimentando la realizzazione *WorldCat Works*, un modo per restituire tutte le connessioni che possono essere stabilite tra le entità di FRBR. L'*opera* indica un livello alto di descrizione di una risorsa, con informazioni come l'autore, il titolo, un riassunto e i soggetti. Questo livello può essere condiviso con tutte le edizioni di quell'espressione concettuale³⁸¹, consentendo di ereditare tutte le informazioni dell'*opera*. Attualmente ciò che si può vedere è una descrizione che riunisce i vari link relativi a quell'entità.

³⁸⁰ «WorldCat Entities | OCLC Developer Network», <https://www.oclc.org/developer/develop/linked-data/worldcat-entities.en.html>.

³⁸¹ Ibid.

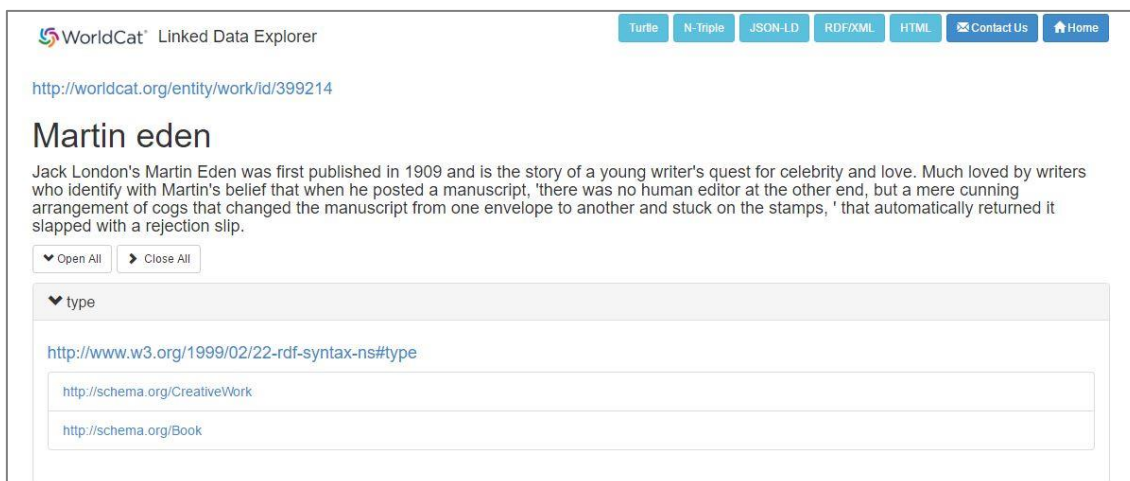


Figura 12: WorldCat Works, tratto da <http://experiment.worldcat.org/entity/work/data/399214>.

OCLC si propone di estendere questa realizzazione anche alle altre entità di FRBR, con l'obiettivo di realizzare un web dei dati che stimoli la condivisione delle informazioni e la creazione di applicazioni e interfacce innovative.

L'organizzazione si sta anche occupando di fornire accesso ai dati bibliografici in diverse lingue. Questa attività cerca di sfruttare i dati di WorldCat, che contiene molte lingue, anche in caratteri non latini. Le informazioni potrebbero così essere presentate nella lingua e nella scrittura preferite dall'utente³⁸². Gandhi, ad esempio, ha scritto la sua biografia in guajarati. I record di WorldCat presentano spesso solamente la forma in caratteri latini. Wikipedia, invece, che è connessa a Wikidata, mostra il titolo originale nei caratteri originali. OCLC ha pensato di sfruttare i vantaggi dei linked open data collegando ad esempio i record alle informazioni nella scrittura originale fornite da Wikidata. L'idea è quella di stipulare delle relazioni tra WorldCat, VIAF, Wikidata, ed altri dataset, che contengono informazioni su una stessa entità, utilizzando la proprietà *owl:SameAs*³⁸³.

³⁸² «Multilingual Bibliographic Structure», <http://www.oclc.org/research/themes/data-science/multilingual-bib-structure.html>.

³⁸³ Karen Smith-Yoshimura, «*Moving Towards True Multilingualism: Leveraging Global Cooperation Through WorldCat*», Presentazione tenuta a OCLC Asia Pacific Regional Council Conference, 4 December 2015, <http://www.slideshare.net/oclc/moving-towards-true-multilingualism-leveraging-global-cooperation-through-worldcat2015>.

Linked data opportunity

The Story of My Experiments with Truth

WIKIPEDIA
The Free Encyclopedia

GANDHI
AN AUTOBIOGRAPHY

Cover page of 1969 reprint by Beacon Press.

Author: Mohandas Karamchand Gandhi

Original title સત્યના પ્રયોગો અથવા આત્મકથા

Translator: Mahatma Gandhi

Country: India

Language: Gujarati

Series: none

ISBN: India - ISBN 81-7229-000-X
United States - authorized edition with forward by Beata Bok, Beacon Press 1969 reprint ISBN 0-8070-0800-9
Dover Publications 1983 reprint of 1948 Public Affairs Press edition ISBN 0-486-24050-2

Satyanā prayogo, athavā, ātmakathā

Author: [Gandhi, Mahatma](#)

Publisher: Amadāvāda : Navajvana Prakāśana Mandra, 2002.

Edition/Format: Print book : Biography : Gujarati [View all editions and formats](#)

Database: WorldCat

Rating: (not yet rated) [0 with reviews - Be the first](#)

Subjects: [Gandhi, Mahatma, 1869-1948](#), [Statesmen - India - Biography](#), [Nationalists - India - Biography](#), [View all subjects](#)

More like this [Similar items](#)

Ingest original script (Gujarati here) for readers who can read the original

સત્યના પ્રયોગો અથવા આત્મકથા

Figura 13: Multilinguismo per i record di WorldCat, tratto dalla presentazione di Karen Smith-Yoshimura, Moving Towards True Multilingualism, OCLC Asia Pacific Regional Council Conference Dec. 2015.

Per poter presentare i dati nella scrittura preferita dagli utenti, è necessario conoscere la lingua dei singoli elementi di un record, distinguendo tra la lingua della risorsa e quella usata per catalogare. Quando i singoli elementi della descrizione bibliografica vengono convertiti in linked open data, si assegna un tag che ne indica la lingua.

Markup for the semantic web

Original Work (in Chinese)
<<http://worldcat.org/entity/work/id/1215997>>
a [schema:CreativeWork](#);
schema:creator <<http://viaf.org/viaf/102266649>> ; # "Gao, Xingjian"
schema:inLanguage "zh";
schema:name "靈山"@zh-hant .

Translated Work (in English)
<<http://worldcat.org/entity/work/id/145209748>>
a [schema:CreativeWork](#);
schema:creator <<http://viaf.org/viaf/102266649>> ; # "Gao, Xingjian"
schema:translator <<http://viaf.org/viaf/81663420>> ; # "Lee, Mabel"
schema:inLanguage "en";
schema:name "Soul Mountain"@en ;
schema:translationOfWork <<http://worldcat.org/entity/work/id/1215997>>

Figura 14: Markup di entità di WorldCat, tratto dalla presentazione di Karen Smith-Yoshimura, Moving Towards True Multilingualism, OCLC Asia Pacific Regional Council Conference Dec. 2015.

Nella Figura 15 viene presentata la visualizzazione di linked open data relativi ad un'opera di Gao Xingjian, scritta originariamente in cinese, e indicata dall'elemento *schema:CreativeWork*. La sua traduzione viene espressa da *schema:translationOfWork*. Ognuno di questi livelli presenta dei tag, *@zh-hant* e *@en*, che indicano la scrittura di quel dato. Il nome dell'autore è in caratteri cinesi tradizionali, mentre il titolo dell'opera tradotta in inglese è in lingua inglese, quindi nello *script* latino. Il progetto, in fase ancora sperimentale, prevede la possibilità taggare gli elementi della descrizione a partire dalle diverse lingue di catalogazione, e di associare dati provenienti da altri dataset espressi in linked open data. Lo scopo è di rendere possibile la costruzione di schermate differenti in base alla lingua e alla scrittura preferita dall'utente. Questo grazie all'utilizzo di tutti i dati disponibili in WorldCat, in VIAF e in altri dataset, strutturati come linked open data.

Conclusioni

Cosa sono i dati bibliografici e qual è il panorama odierno di questo ambito? Cosa sono i linked open data e in che modo possono valorizzare i dati bibliografici? Quali criticità presenta la loro applicazione?

Queste sono le domande a cui la tesi ha cercato di rispondere.

La contestualizzazione dei dati bibliografici ha messo in luce quali sono oggi gli obiettivi della catalogazione. L'apertura dei dati, l'interoperabilità e l'accessibilità sono i principali elementi di novità che caratterizzano i *Principi Internazionali di Catalogazione* del 2016. Questi sono gli elementi che dovrebbero guidare la strutturazione delle informazioni e la costruzione dei cataloghi. Il recepimento del modello FRBR ha contribuito ad una maggiore granularità degli elementi bibliografici. L'individuazione delle diverse entità, che possono essere desunte da una risorsa, ha infatti portato alla scomposizione delle informazioni in unità sempre più piccole. Questo perché l'atomo informativo è potenzialmente più riutilizzabile e di conseguenza più interoperabile, rispetto ai blocchi di testo che caratterizzano le descrizioni del formato MARC. Il formato di marcatura più utilizzato attualmente dalle biblioteche resta però vincolato a questo ambito, cosa che ostacola il dialogo con altre comunità. La sua struttura inoltre non consente di realizzare appieno tutte le relazioni tra entità previste da FRBR. I nuovi standard, come BIBFRAME e RDA, nascono con l'obiettivo di aprire il mondo delle biblioteche all'ecosistema del web. L'interesse tra il mondo delle biblioteche e la comunità del web è reciproco. I dati bibliografici sono spesso informazioni di qualità caratterizzate da un alto livello di standardizzazione e da contenuti semanticamente molto ricchi. Rendere questi dati parte del web e liberarli dai silos dei database, può giovare a tutti coloro che cercano queste informazioni, per studio, per lavoro o semplicemente per poter riutilizzare dati che qualcuno ha già creato. Allo stesso tempo far emergere i dati nel web significa renderli più visibili e avvicinarsi alla principale infrastruttura di ricerca degli utenti. È importante sottolineare che questo potrebbe avere un ritorno di visibilità per le biblioteche. A fronte di queste possibilità, la comunità catalografica ha cominciato ad interessarsi ai linked open data in quanto prospettiva che può interessare gli strumenti di recupero e accesso all'informazione, e trasformare il catalogo in un insieme di dati collegati. Proprio il

collegamento, cioè la nozione di relazione semanticamente qualificata, può migliorare l'esperienza di navigazione tra le risorse, come previsto dai *Principi Internazionali di Catalogazione*.

Nel secondo capitolo si è cercato quindi di delineare una panoramica su queste nuove tecnologie che presentano la possibilità di aumentare l'interoperabilità semantica e la creazione di un web di dati. Avvicinarsi a questo ambito comporta uno sforzo intellettuale da parte di chi si appropria con una prospettiva biblioteconomica. Si è comunque ritenuto opportuno offrire uno sguardo sulle tecnologie principali e sull'impostazione teorica che ne ha segnato lo sviluppo. I linked open data sono infatti un insieme di buone pratiche e la loro creazione prevede la compresenza di diverse tecnologie. In particolare RDF, l'uso di HTTP URI, linguaggi come SPARQL e formati di serializzazione quali XML.

Ma quali sono effettivamente le possibili applicazioni dei linked open data ai dati bibliografici? A questa domanda si è tentato di rispondere nel capitolo terzo e quarto.

Le linee guida del Library Linked Data Incubator Group rappresentano il primo tentativo di dialogo tra l'ambito bibliotecario e l'ambito del web semantico. In particolare queste tecnologie consentono di creare dati condivisibili e riutilizzabili, incidendo di conseguenza sull'economia della catalogazione. Ciò non significa che i catalogatori debbano demandare la creazione di notizie bibliografiche ad attori esterni ma che essi potrebbero avvalersi ancora di più di dati già creati, dedicandosi ad arricchire le descrizioni, a fare gli opportuni controlli per i dati di autorità e creare le ricche relazioni tra entità previste da FRBR. L'utilizzo di RDF riesce ad esprimere ogni informazione individuando ciascun dato, che può essere espresso con un identificativo univoco. Questo processo consente di recuperare singole porzioni di informazioni in fase di cattura. Quindi possono essere recuperate ad esempio una data o il nome di un editore, senza dover per forza catturare un intero record, come avviene adesso. I linked open data supportano anche funzionalità multilingue, grazie alla separazione tra concetti, identificati da URI, e l'applicazione di etichette, le quali vanno a rappresentare le diverse forme linguistiche di quel concetto. Il valore dell'*interlinking*, ovvero della creazione di relazioni qualificate tra i dati interni e dataset esterni, può essere applicato anche agli elementi bibliografici dei cataloghi. I *Principi Internazionali di Catalogazione* prevedono, tra gli obiettivi del catalogo, la navigazione sia all'interno di un OPAC che all'esterno di esso, in ambienti

non necessariamente legati alle biblioteche. Ecco che questo obiettivo può essere potenziato dall'utilizzo dei linked open data, che proprio nei loro Principi, prevedono la creazione di link tra dataset di diversi ambiti. L'utente potrebbe esplorare i dati in modalità *follow your nose*, lasciandosi guidare dalla propria voglia di scoprire nuove informazioni. I linked data non devono necessariamente essere caratterizzati da licenze che ne consentano l'apertura ma il loro valore aggiunto risiede proprio nella possibilità del libero riutilizzo. In questo modo chiunque può accedere ai dati, scaricarli e modificarli, per creare nuove applicazioni o nuove modalità di fruizione.

La creazione dei linked open data richiede però una serie di operazioni, sia concettuali che tecnologiche. È necessario infatti scegliere in che modo rappresentare il proprio dominio conoscitivo. A questo scopo sono state sviluppate ontologie appositamente per il settore bibliotecario. I linked open data sono comunque caratterizzati da un approccio *mix and match*, nel senso che è una buona pratica riutilizzare ontologie e vocabolari preesistenti, in modo da poter dialogare con più ambiti possibili. Lo strumento che di solito viene utilizzato è il profilo di applicazione, che si concretizza in un documento nel quale vengono esplicitati gli obiettivi dei progetti e i *namespace* utilizzati.

La pubblicazione vera e propria di dati strutturati come HTTP URI, richiede diversi passaggi tecnici e in particolare l'ausilio di strumenti di conversione, il recupero o la creazione di URI e l'utilizzo di *triplestore* per la loro conservazione. In aggiunta vanno predisposte delle "finestre" che consentano a chiunque di interrogare e recuperare i dati. Questo è il compito degli SPARQL Endpoint.

Le principali criticità nell'applicazione dei linked open data ai dati bibliografici derivano dall'aspetto tecnologico e dalle sue ricadute economiche. Tutto il ciclo di creazione e conversione richiede la presenza di professionisti, e di conseguenza investimenti economici, sia da parte delle biblioteche, sia da parte di chi crea software per questo ambito. Inoltre le ontologie e i vocabolari disponibili sono soggetti ad una continua revisione e richiedono ancora un pesante intervento umano per il loro aggiustamento o allineamento. I dataset di linked open data presentano spesso interruzioni o perdite, dovute anche a problemi di persistenza degli URI, duplicazioni di un identificativo per una stessa entità o mancanza di aggiornamento.

I benefici e le criticità emergono anche dall'analisi dei casi di applicazione.

Tutti i progetti presi in considerazione hanno sviluppato un profilo di applicazione. I modelli hanno anche messo in luce quali siano gli elementi più comuni. SKOS, Dublin Core, Schema.org, FOAF e GeoNames. Europeana e la Digital Public Library of America hanno scelto di fornire un'apertura totale dei metadati, utilizzando la CC0. OCLC invece ha optato per l'ODC-by, che prevede la clausola dell'attribuzione. Europeana e la Digital Public Library of America sono riuscite a promuovere la creazione di applicazioni, che sono molte e presentano spesso punti di vista interessanti ed inconsueti sulla scoperta del patrimonio culturale. OCLC invece ha sviluppato strumenti come VIAF, WorldCat e WorldCat Entities, resi possibili grazie all'adozione dei linked open data. Lo scopo è quello di aggregare i dati e di creare servizi più ricchi a livello informativo, implementando il modello concettuale FRBR e le entità da esso individuate. Così in sostanza possono essere raccolti i dati che provengono dai vari record bibliografici per creare delle descrizioni di livello alto, su elementi come *opere, espressioni, persone, enti e soggetti*. Inoltre OCLC sta cercando di sviluppare strumenti multilingue, che consentano agli utenti di visualizzare le descrizioni nella lingua e nella scrittura che essi preferiscono. Ma ciò che è piuttosto visibile, principalmente in Europeana, è la disomogeneità dei dati e la loro presentazione in forma di URI. In alcuni casi la presenza di informazioni predisposte principalmente per le macchine disturba la navigazione e la ricerca. Inoltre sia Europeana che la Digital Public Library of America presentano sistemi di ricerca e di visualizzazione dei dati che ricalcano gli OPAC classici. Non sono ancora presenti modalità innovative di navigazione e manca, in qualche modo, la creazione di una ricca rete semantica tra le informazioni.

Per concludere, le ricerche condotte per questa tesi cercano di dimostrare che i linked open data rappresentano una possibilità per raggiungere gli obiettivi posti dai *Principi Internazionali di Catalogazione*, ma comportano attualmente diverse difficoltà applicative. È comunque importante che la comunità bibliotecaria si avvicini a questo ambito, per capire effettivamente quali benefici esso possa portare ma anche quali debolezze esso in realtà presenti. Il panorama attuale ricorda il periodo in cui cominciarono ad essere sperimentati i primi cataloghi elettronici: una fase di tentativi, inventiva e progetti più o meno riusciti. Gli obiettivi posti da Cutter non sono ancora pienamente realizzati dai cataloghi, così come la profondità delle relazioni tra entità catalografiche prevista da FRBR. Gli strumenti di accesso all'informazione dovrebbero

fornire dati utili, omogenei, qualitativamente ricchi e navigabili per gli utenti delle biblioteche. Il linguaggio dei metadati che utilizziamo è tuttora poco comprensibile, si pensi solo a indicazioni come “titolo uniforme”. Inoltre i cataloghi non hanno pienamente sviluppato strumenti di navigazione semantica, a partire dai thesauri o dalle classificazioni. Dunque l’ambito catalografico ha ancora molto lavoro davanti a sé. L’obiettivo non deve essere affidarsi ciecamente ad una tecnologia, ma valutare quale sia in grado di realizzare al meglio i bisogni specifici delle biblioteche e dei loro utenti. Restano da approfondire l’evoluzione dei cataloghi che adottino come base di dati RDF e lo sviluppo di sistemi di ricerca intelligente. Inoltre anche i linked open data andranno conservati, mantenuti, e resi stabili, e bisognerà capire quali possano essere le strategie per garantire una loro maggiore persistenza. Riguardo la rappresentazione formale delle risorse bibliografiche invece andrà valutato se è opportuna la definizione di un’ontologia condivisa da tutto l’ambito bibliotecario e se la riduzione ad un unico modello può confliggere con il metodo *bottom-up* prevalente nell’ambito dei linked open data.

Bibliografia

- Allemang, Dean, e James A. Hendler. *Semantic Web for the working ontologist: effective modeling in RDFS and OWL*. 2nd ed. Waltham: Morgan Kaufmann, 2011.
- Bacchi, Cristian. «Marc2Wikibase: ipotesi di caricamento dei dati bibliografici in Wikibase». Presentazione tenuta a *SBN: un'eredità per il futuro, Ravenna, 16 settembre 2016*. Consultato 10 gennaio 2017.
http://www.bibliotecheromagna.it/admin/PagPar.php?op=fg&id_pag_par=1683&fld=file.
- Baker, Thomas. «Designing Data for the Open World of the Web». *JLIS.it* 4, n. 1 (2013): 63-6. doi 10.4403/jlis.it-6308.
- Baker, Thomas, Emanuelle Bermès, Karen Coyle, Gordon Dunsire, Antoine Isaac, e Peter Murray. «Library Linked Data Incubator Group Final Report». W3C Incubator Group Report, 2011. Consultato 20 dicembre 2016. <https://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/>.
- Barbanti, Ilaria. «La codifica informatica dei dati di archivio: EAD ed EAC». *Scrinia*, n. 2–3 (2005): 1-15.
http://media.regesta.com/dm_0/regesta/regestacms/generic/documentazione/Barbanti-EADEAC2005.pdf.
- Bassi, Maria Cristina. «Formati MARC-XML e servizi derivati: resoconto sulla sessione al congresso IFLA 2005». *Bibliotime* VIII, n. 3 (2005).
<http://www.aib.it/aib/sezioni/emr/bibtime/num-viii-3/bassi.htm>.
- Bellingeri, Luca, e Maria Chiara Giunti. «BNI Open and Collaborative: How and Why». *JLIS.it*, n. 1 (2017): 67-76. doi:10.4403/jlis.it-12167.
- Bergamin, Giovanni. «Metadati e open data: nuovi paradigmi per vecchie professioni». 2012. Consultato 14 dicembre 2016.
https://www.openstarts.units.it/dspace/bitstream/10077/9360/1/Bergamin_MAB_2012.pdf.
- Bergamin, Giovanni, e Anna Lucarelli. «The Nuovo Soggettario as a Service for the Linked Data World». *JLIS.it* 4, n. 1 (2012): 213-26. doi.org/10.4403/jlis.it-5474.

- Berners-Lee, Tim. «Linked Data - Design Issues». 2006. Consultato 28 dicembre 2016.
<https://www.w3.org/DesignIssues/LinkedData.html>.
- Berners-Lee, Tim, e Mark Fischetti. *L'architettura del nuovo Web: dall'inventore della rete il progetto di una comunicazione democratica, interattiva e intercreativa*. Milano: Feltrinelli, 2001.
- Berners-Lee, Tim, James A. Hendler, e Ora Lassila. «The Semantic Web. A new form of web content that is meaningful to computers will unleash a revolution of new possibilities». *Scientific American magazine*, (2001): 1-4. https://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf.
- Biagetti. «Le ontologie come strumenti per l'organizzazione della conoscenza in rete». *AIDA Informazioni*, n. 1-2 (2010): 9-32.
- Bianchini, Carlo. «Dagli OPAC ai library linked data: come cambiano le risposte ai bisogni degli utenti». *AIB studi* 52, n. 3 (2012): 303-23. <http://aibstudi.aib.it/article/view/8597>.
- . «Dai cataloghi alla navigazione semantica». In *Linked data per biblioteche, archivi e musei: perché l'informazione sia del web e non solo nel web*, 173-204. Milano: Editrice bibliografica, 2015.
- . «Library Linked Data e il futuro delle biblioteche». 2012. Consultato 2 gennaio 2017.
<https://sites.google.com/site/homepagecarlobianchini/Ricerca/library-linked-data-e-il-futuro-delle-biblioteche>.
- . «Universo bibliografico, descrizione e accesso». In *Biblioteche e biblioteconomia: principi e questioni*, a cura di Giovanni Solimine e Paul Gabriele Weston, 229-54. Roma: Carocci, 2015.
- Bianchini, Carlo, e Mauro Guerrini. «Quis custodiet ipsos custodes?: osservazioni sulle relazioni fra FRBR, ICP, ISBD e RDA». *Il bibliotecario* 3, n. 1 (2009): 59-85.
- Bizer, Christian, Tom Heath, e Tim Berners-Lee. «Linked Data - The Story So Far». *International Journal on Semantic Web and Information Systems* 5, n. 3 (2009): 1-22.
doi:10.4018/jswis.2009081901.

- Bizer, Christian, Tom Heath, Kingsley Idehen, e Tim Berners-Lee. «Linked Data on the Web», In *Proceedings of the 17th international conference on World Wide Web, Beijing China, 2008*, 1265-6. New York: ACM Press, 2008. doi:10.1145/1367497.1367760.
- Bouquet, Paolo, e Roberta Ferrario. «Il Semantic Web». *Networks. Rivista di filosofia dell'intelligenza artificiale e scienze cognitive*, n. 2 (2003): i–viii.
- Calhoun, Karen, Joanne Cantrell, Peggy Gallagher, e Janet Hawk. *Online Catalogs: What Users and Librarians Want: An OCLC Report*. Dublin: OCLC, 2009.
- Connaway, Lynn Silipigni, e Timothy J. Dickey. «The Digital Information Seeker: Report of Findings from Selected OCLC, RIN and JISC User Behaviour Projects». In *The Library in the Life of the User: Engaging with People Where They Live and Learn*, a cura di Lynn Silipigni Connaway, 102–9. Dublin: OCLC Research, 2015.
- Coyle, Karen. *FRBR, before and after: a look at our bibliographic models*. Chicago: ALA Editions, 2016.
- . «Linked Data: An Evolution». *JLIS.it* 4, n. 1 (2013): 53-61. doi.org/10.4403/jlis.it-5443.
- . «On the Web, of the Web». 2011. Consultato 2 gennaio 2017 <http://www.kcoyle.net/presentations/lita2011.html>.
- . «Semantic Web and Linked Data». *Tutorial on Metadata and the Semantic Web*. 2012. Consultato 28 dicembre 2016. <http://kcoyle.net/metadata/7lecture.html>.
- Coyle, Karen, e Thomas Baker. «DCMI Home: Dublin Core Metadata Initiative (DCMI)». 2009. Consultato 10 gennaio 2017. <http://dublincore.org/>.
- Cutter, Charles Ammi. *Rules for a Dictionary Catalog*. Washington: Government Printing Office, 1904. Per una versione online: <https://archive.org/stream/rulesforadictio06cuttgoog#page/n10/mode/1up>.
- De Robbio, Antonella. «URI, URN e URL, una questione di definizioni: universal versus uniform». *Biblioteche oggi*, n. 1 (2002): 30–7. <http://www.bibliotecheoggi.it/2002/20020103001.pdf>

- De Robbio, Antonella, e Silvia Giacomazzi. «Dati aperti con LODe». *Bibliotime* XIV, n. 2 (2011). <http://www.aib.it/aib/sezioni/emr/bibtime/num-xiv-2/derobbio.htm>.
- Di Giammarco, Fabio. «Le biblioteche nella rete dei linked data». *Digitalia* 1, (2012): 138–41. <http://digitalia.sbn.it/article/view/544/383>.
- Digital Public Library of America. «The Digital Public Library of America Policy Statement on Metadata». 2013. Consultato 2 gennaio 2017. <https://dp.la/info/wp-content/uploads/2013/04/DPLAMetadataPolicy.pdf>.
- . «An introduction to the DPLA Metadata Model». 2015. Consultato 4 gennaio 2017. https://dp.la/info/wp-content/uploads/2015/03/Intro_to_DPLA_metadata_model.pdf.
- Di Maio, Paola. «Linked Data Beyond Libraries: Towards Universal Interfaces and Knowledge Unification». In *Linked data and user interaction: the road ahead*, a cura di H. Frank Cervone e Lars G. Svensson, 12–32. Berlin ; Boston: De Gruyter Saur, 2015.
- DocumentaryStorm. *The Internet's Own Boy: The Story of Aaron Swartz*. Video YouTube, 1:44:59, pubblicato il 30 giugno 2014. Consultato il 20 dicembre 2016. <https://www.youtube.com/watch?v=M85UvH0TRPc>.
- Domanovszky, Akos. *Functions and objects of author and title cataloguing*. Budapest: Akademiai Kiado, 1974.
- Esposito, Renzo. «Regole per la compilazione del catalogo». In *Atti del convegno di studi su Antonio Panizzi, Roma, 21-22 aprile 1980*, 225-41. Galatina: Editrice salentina, 1980.
- Europeana. «The Data Exchange Agreement - Europeana Professional». Consultato 8 febbraio 2017. <http://pro.europeana.eu/the-data-exchange-agreement/>.
- . «Definition of the Europeana Data Model v5.2.7». 2016. Consultato il 3 gennaio 2017. http://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation//EDM_Definition_v5.2.7_042016.pdf.
- . «The Europeana Data Model». 2010. Consultato il 3 gennaio 2017. <http://www.ifla.org/past-wlic/2010/149-doerr-en.pdf>
- . «Europeana Data Model – Mapping Guidelines v2.3». 2016. Consultato il 3 gennaio 2017.

http://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Mapping_Guidelines_v2.3_112016.pdf.

———. «Europeana Data Model Primer». 2013. Consultato il 3 gennaio 2017.

http://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Primer_130714.pdf.

———. «The Europeana Licensing Framework». 2013. Consultato il 4 gennaio 2017.

http://pro.europeana.eu/files/Europeana_Professional/Publications/Europeana%20Licensing%20Framework.pdf.

Galeffi, Agnese. «Standard di catalogazione». In *Biblioteche e biblioteconomia: principi e questioni*, a cura di Giovanni Solimine e Paul Gabriele Weston, 255–80. Roma: Carocci, 2015.

Galeffi, Agnese, e Lucia Sardo. *FRBR*. Roma: Associazione italiana biblioteche, 2013.

Gambari, Stefano, e Matteo Barucci. «Formati di registrazione per record bibliografici leggibili dalla macchina (MACHINE-Readable Cataloging, MARC)». In *Biblioteconomia: guida classificata*, 416–20. Milano: Editrice Bibliografica, 2007.

Genetasio, Giuliano. «Analisi e controllo bibliografico». In *Biblioteconomia: guida classificata*, a cura di Mauro Guerrini, 355–61. Milano: Editrice Bibliografica, 2007.

———. «I Principi Internazionali di Catalogazione (ICP) e il loro futuro». *JLIS.it* 3, n. 1 (2012): 1-18. doi:10.4403/jlis.it-4936.

Godby, Carol Jean. «A Division of Labor: The Role of Schema.org in a Semantic Web Model of Library Resources». In *Linked Data for Cultural Heritage*, 73–103. London: Facet, 2016.

Gordon, Eric. «Semantic Web Specification at W3C - Slide list». 2004. Consultato 2 febbraio 2017. <https://www.w3.org/2004/Talks/0412-RDF-functions/Overview.html>.

Gorman, Michael. «Changes in cataloguing codes: rules for entry and heading». *Library Trends* 25, n. 3 (1977): 587–601.

- . «Seymour Lubetzky uomo di principi: discorso tenuto al Seymour Lubetzky 100. Birthday Symposium, 18 aprile 1998, University of California, Los Angeles». *Biblioteche oggi* 18, n. 6 (2000): 6–11.
- Guerrini, Mauro. «Le funzioni del catalogo dall'ICCP a FRBR». 2000. Consultato 3 gennaio 2017. <http://www.aib.it/aib/sezioni/toscana/conf/frbr/guerrini.htm>.
- Guerrini, Mauro, e Carlo Bianchini. «Catalogazione». In *Biblioteconomia: guida classificata*, 372–90. Milano: Editrice Bibliografica, 2007.
- Guerrini, Mauro, e Tiziana Possemato. «Linked Data: A New Alphabet for the Semantic Web». *JLIS.it* 4, n. 1 (2013): 67-90. doi 10.4403/jlis.it-6305.
- . *Linked data per biblioteche, archivi e musei: perché l'informazione sia del web e non solo nel web*. Milano: Editrice Bibliografica, 2015.
- Heath, Tom, e Christian Bizer. «Linked Data: Evolving the Web into a Global Data Space». *Synthesis Lectures on the Semantic Web: Theory and Technology* 1, n. 1 (2011): 1–136. doi:10.2200/S00334ED1V01Y201102WBE001.
- Heery, Rachel, e Manjula Patel. «Application Profiles: Mixing and Matching Metadata Schemas». *Ariadne*, n. 25 (2000). <http://www.ariadne.ac.uk/issue25/app-profiles>.
- Hooland, Seth van, e Ruben Verborgh. *Linked Data for Libraries, Archives and Museums: How to Clean, Link and Publish Your Metadata*. Chicago: Neal-Schuman, 2014.
- Huerga, Iker, e Michael P. Lauruhn. «Linked Data Implications for Authority Control and Vocabularies». In *Linked Data for Cultural Heritage*, a cura di Ed Jones e Michele Seikel, 55–72. London: Facet, 2016.
- Iacono, Antonella. «Dal record al dato. Linked data e ricerca dell'informazione nell'OPAC.». *JLIS.it* 5, n. 1 (2013): 77–102. doi 10.4403/jlis.it-9095
- ICCU. «ICCU | Istituto Centrale per il Catalogo Unico - Gruppo di lavoro Linked Open Data SBN». Consultato 12 febbraio 2017. http://www.iccu.sbn.it/opencms/opencms/it/main/attivita/gruppilav_commissioni/pagina_0005.html.

IFLA, Barbara B. Tillett, e Ana Lupe Cristán, a c. di. *IFLA cataloguing principles: the Statement of International Cataloguing Principles (ICP) and its glossary: in 20 languages*. München: K.G. Saur, 2009.

———. «IFLA -- DCMI library application profile». Consultato 26 dicembre 2016. <http://www.ifla.org/best-practice-for-national-bibliographic-agencies-in-a-digital-age/node/8824>.

———. «IFLA -- Functional Requirements for Bibliographic Records». Consultato 23 dicembre 2016. <http://www.ifla.org/publications/functional-requirements-for-bibliographic-records>.

———. «IFLA -- Functional Requirements for Subject Authority Data». Consultato 20 dicembre 2016. <http://www.ifla.org/node/5849>.

———. «IFLA -- ISBD Linked Data Study Group». Consultato 29 dicembre 2016. <http://www.ifla.org/node/1795>.

———. «IFLA Code of Ethics for Librarians and other Information Workers (full version)». Consultato 28 gennaio 2017. <http://www.ifla.org/publications/node/11092>.

———. «IFLA ISBD Vocabularies». Consultato 30 gennaio 2017. <http://iflstandards.info/ns/isbd/>.

———. «IFLA Statement on open access». 2011. Consultato 28 gennaio 2017. <http://www.ifla.org/files/assets/hq/news/documents/ifla-statement-on-open-access.pdf>.

IFLA Cataloguing Section, e IFLA Meetings of Experts on an International Cataloguing Code. «Statement of International Cataloguing Principles (ICP)». 2016. Consultato 20 dicembre 2016. http://www.ifla.org/files/assets/cataloguing/icp/icp_2016-en.pdf.

IFLA Study Group on the Functional Requirements for Bibliographic Records, e International Federation of Library Associations and Institutions, a c. di. *Functional requirements for bibliographic records: final report*. München: K.G. Saur, 1998.

International Conference on Cataloguing Principles. «Statement of Principles Adopted by The International Conference on Cataloguing Principles». 1961. Consultato 20 dicembre 2016. http://www.ifla.org/files/assets/cataloguing/IMEICC/IMEICC1/statement_principles_paris_1961.pdf.

- Isaac, Antoine, Clayphan, Robina, e Bernhard Haslhofer. «EUROPEANA: Moving to Linked Open Data». *Information standards quarterly* 24, n. 2–3 (2012): 34–40.
- Joint Steering Committee for Development of RDA. «RDA: Resource Description and Access», 2010. Consultato 15 febbraio 2017. <http://www.rdatoolkit.org>
- Le Boeuf, Patrick. «Brave new FRBR world». 2003. Consultato 27 dicembre 2016. http://www.ifla.org/files/assets/cataloguing/IMEICC/IMEICC1/papers_leboeuf.pdf.
- Library of Congress Working Group on the Future of Bibliographic Control. «On the Record». Washington: Library of Congress, 2008. Consultato 15 gennaio 2017. <https://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf>.
- Lincoln, Matthew. «Linked Open Realities: The Joys and Pains of Using LOD for Research». *matthewlincoln.net*, 2016. Consultato 18 gennaio 2017. <http://matthewlincoln.net/2016/06/06/linked-open-realities-the-joys-and-pains-of-using-lod-for-research.html>.
- Lo Castro, Valeria. «Web Semantico e Linked Open Data: best practices, prospettive e criticità». *Nuovi Annali della Scuola Speciale per Archivisti e Bibliotecari* XXVIII (2014): 207–22.
- Madison, Olivia M. A. «The Origins of the IFLA Study on Functional Requirements for Bibliographic Records». *Cataloging & Classification Quarterly* 39, n. 3–4 (2005): 15–37. doi:10.1300/J104v39n03_02.
- Manzotti, Giulia. «Analisi e riflessioni sul VIAF, Virtual International Authority File». *JLIS.it* 1, n. 2 (2010): 357–81. doi 10.4403/jlis.it-4520
- Marchitelli, Andrea. *Orientarsi tra le informazioni in biblioteca: cataloghi, banche dati, motori di ricerca*. Milano: Editrice Bibliografica, 2015.
- Marchitelli, Andrea, e Giovanna Fragimelica. *OPAC*. Roma: Associazione italiana biblioteche, 2012.
- Medic, Adis, e Adis Golubovic. «Making secure Semantic Web». *Universal Journal of Computer Science and Engineering Technology* 2, n. 1 (2010): 99–104.

- Miller, Paul. «Interoperability: What Is It and Why Should I Want It?» *Ariadne*, n. 24 (2000).
<http://www.ariadne.ac.uk/issue24/interoperability>.
- «Overview of the BIBFRAME 2.0 Model (BIBFRAME - Bibliographic Framework Initiative, Library of Congress)». Consultato 4 febbraio 2017.
<https://www.loc.gov/bibframe/docs/bibframe2-model.html>.
- Panizzi, Antonio. *Catalogue of Printed Books in the British museum: Volume I*. London: Order of the Trustees, 1841. Per una versione online:
<https://catalog.hathitrust.org/Record/001761602>.
- Pasqui, Valdo. «Xml: una introduzione. Nozioni di base e possibili aree di sviluppo nel contesto bibliotecario». *Biblioteche oggi* 19, n. 6 (2001): 18–24.
- Pattueli, M. Cristina, Alexandra Provo, e Hilary Thorsen. «Ontology Building for Linked Open Data: A Pragmatic Perspective». *Journal of Library Metadata* 15, n. 3–4 (2015): 265–94.
doi:10.1080/19386389.2015.1099979.
- Pattueli, M. Cristina, e Hilary Thorsen. «Linked Open Data and the Cultural Heritage Landscape». In *Linked Data for Cultural Heritage.*, a cura di Ed Jones e Michele Seikel, 1–22. London: Facet, 2016.
- Peroni, Silvio, Francesca Tomasi, e Fabio Vitali. «Reflecting on the Europeana Data Model». In *Digital Libraries and Archives*, a cura di Maristella Agosti, Floriana Esposito, Stefano Ferilli, e Nicola Ferro, 228–40. Berlin: Springer, 2013.
http://link.springer.com/10.1007/978-3-642-35834-0_23.
- Peruginelli, Susanna. «I profili di applicazione». Presentazione tenuta a *Seminario nazionale sui metadati, ICCU, Roma, 3 aprile 2001*. Consultato 18 gennaio 2017.
www.iccu.sbn.it/upload/documenti/Peruginelli.ppt.
- Revelli, Carlo, e Giulia Visintin. *Il catalogo*. Milano: Editrice Bibliografica, 2008.
- Ridi, Riccardo. *La biblioteca come ipertesto: verso l'integrazione dei servizi e dei documenti*. Milano: Editrice Bibliografica, 2007.
- Riva, Pat. «FRBR Review Group Initiatives and the World of Linked Data». *JLIS.it* 4, n. 1 (2013): 105-117. doi 10.4403/jlis.it-5482.

- . «On the new conceptual model of the bibliographic universe: the FRBR Library Reference Model». *AIB studi*, n. 2 (2016): 265-75. doi:10.2426/aibstudi-11480.
- Shadbolt, N., T. Berners-Lee, e W. Hall. «The Semantic Web Revisited». *IEEE Intelligent Systems* 21, n. 3 (2006): 96–101. doi:10.1109/MIS.2006.62.
- Signore, Oreste. «Il ruolo centrale di XML nell’evoluzione del Web», 2001. Consultato 30 dicembre 2016. <http://www.w3c.it/papers/SignorePerXMLDays2001.pdf>
- . «RDF per la rappresentazione della conoscenza», 2002. Consultato 30 dicembre 2016. <http://www.w3c.it/papers/RDF.pdf>.
- . «Strutturare la conoscenza: XML, RDF e Semantic Web», 2003. Consultato 30 dicembre 2016. <http://www.w3c.it/papers/ck2003.pdf>
- Smith-Yoshimura, Karen. «Analysis of International Linked Data Survey for Implementers». *D-Lib Magazine* 22, n. 7/8 (2016). doi:10.1045/july2016-smith-yoshimura.
- . «Moving Towards True Multilingualism: Leveraging Global Cooperation Through WorldCat». Presentazione tenuta a *OCLC Asia Pacific Regional Council Conference, 4 December 2015*. Consultato 2 febbraio 2017. <http://www.slideshare.net/oclc/moving-towards-true-multilingualism-leveraging-global-cooperation-through-worldcat>
- Svenonius, Elaine. *The Intellectual Foundation of Information Organization*. Cambridge: MIT Press, 2009.
- Svenonius, Elaine, Council on Library Resources, e Conference on the Conceptual Foundations of Descriptive Cataloging, a c. di. *The Conceptual Foundations of Descriptive Cataloging*. San Diego: Academic Press, 1989.
- Tennant, Roy. «MARC Must Die». *Library Journal*, (2002): 26-8. <http://lj.libraryjournal.com/2002/10/ljarchives/marc-must-die/>.
- The MARC Pilot Project: Final Report on a Project Sponsored by the Council on Library Resources, Inc.* Washington: Government Printing Office, 1968.
- Tillett, Barbara. *The bibliographic universe and the new IFLA cataloging principles: lectio magistralis in library science = L’universo bibliografico e i nuovi principi di*

- catalogazione dell'IFLA: lectio magistralis di biblioteconomia*. Fiesole: Casalini libri, 2008.
- . «What is FRBR? A Conceptual Model for the Bibliographic Universe». Washington: Library of Congress Cataloging Distribution Service, 2004.
- Trombone, Antonella. «Cataloghi, modelli concettuali, data model: gli orientamenti della ricerca e gli ordinamenti tematici nella Library and information science». *Bibliothecae.it* V, n. 1 (2016): 95–129. doi:10.6092/issn.2283-9364/6108.
- . «Il progetto BIBFRAME della Library of Congress: come stanno cambiando i modelli strutturali e comunicativi dei dati bibliografici». *AIB studi* 55, n. 2 (2015).
<http://aibstudi.aib.it/article/view/11100>.
- Van Hees, Dimitri. «Why APIs are the missing 'link' in Linked Open Data». 2015. Consultato 28 dicembre 2016. <https://www.linkedin.com/pulse/why-apis-missing-link-linked-open-data-dimitri-van-hees>.
- VIAF Council meeting. «Annual report: VIAF Council meeting 2015». 2015. Consultato 3 febbraio 2017. <http://www.oclc.org/content/dam/oclc/viaf/viafc-2015-annual-report.pdf>.
- Weston, Paul Gabriele. «Authority data». In *Biblioteche e biblioteconomia: principi e questioni*, 281–313. Roma: Carocci, 2015.
- . *Gli strumenti della cooperazione in rete. Dal catalogo elettronico ai sistemi della ricerca interdisciplinare*. Napoli: ClioPress, 2003.
- . *Il catalogo elettronico: dalla biblioteca cartacea alla biblioteca digitale*. Roma: Carocci, 2002.
- . «La gestione elettronica delle biblioteche». In *Biblioteconomia: principi e questioni*, a cura di Giovanni Solimine e Paul Gabriele Weston. Roma: Carocci, 2007.
- Yu, Liyang. «Follow Your Nose: A Basic Semantic Web Agent». In *A Developer's Guide to the Semantic Web*, di Liyang Yu, 711–36. Berlin: Springer, 2014.
http://link.springer.com/10.1007/978-3-662-43796-4_16.

Sitografia

- «5-star Open Data». Consultato 5 febbraio 2017. <http://5stardata.info/en/>.
- «A proposito delle Licenze - Creative Commons». Consultato 5 febbraio 2017.
<https://creativecommons.org/licenses/>.
- «About - the Datahub». Consultato 5 febbraio 2017. <https://datahub.io/about>.
- «About OCLC». Consultato 22 gennaio 2017. <https://www.oclc.org/en/about.html>.
- «About W3C». Consultato 21 gennaio 2017. <https://www.w3.org/Consortium/>.
- «Accessibilità | AgID». Consultato 29 gennaio 2017. <http://www.agid.gov.it/agenda-digitale/pubblica-amministrazione/accessibilita>.
- «Algoritmi – Dentro la Ricerca – Google». Consultato 31 gennaio 2017.
<https://www.google.it/intl/it/insidesearch/howsearchworks/algorithms.html>.
- «Antonio Panizzi (1797-1879) - Author - Resources from the BnF». Consultato 4 febbraio 2017.
http://data.bnf.fr/13739722/antonio_panizzi/.
- «App Library · Digital Public Library of America». Consultato 20 gennaio 2017.
<https://dp.la/apps>.
- «Apps showcase - Europeana Labs». Consultato 9 febbraio 2017. <http://labs.europeana.eu/apps>.
- «Art Faces - Europeana Labs». Consultato 9 febbraio 2017. <http://labs.europeana.eu/apps/art-faces>.
- «Banca dati cumulativa - Museo Galileo». Consultato 12 febbraio 2017.
<http://www.museogalileo.it/esplora/biblioteche/biblioteca/bancadaticumulativa.html>.
- «Base di dati». *Wikipedia*, 22 gennaio 2017.
https://it.wikipedia.org/w/index.php?title=Base_di_dati&oldid=85497469.
- «Best Practices for Publishing Linked Data». Consultato 4 febbraio 2017.
<https://www.w3.org/TR/ld-bp/>.
- «Billi». Consultato 5 febbraio 2017. <http://billi.nypl.org/>.

- «CC REL - Creative Commons». Consultato 5 febbraio 2017.
https://wiki.creativecommons.org/wiki/CC_REL.
- «ConverterToRdf - W3C Wiki». Consultato 4 febbraio 2017.
<https://www.w3.org/wiki/ConverterToRdf#MARC>.
- «Creative Commons FAQ | CreativeCommons.it». Consultato 5 febbraio 2017.
<http://www.creativecommons.it/FAQ>.
- «Data - Europeana Labs». Consultato 9 febbraio 2017. <http://labs.europeana.eu/data>.
- «Dati aperti: la nuova licenza italiana | Trasparenza e Integrità». Consultato 5 febbraio 2017.
<http://trasparenza.formez.it/content/dati-aperti-nuova-licenza-italiana>.
- «datos.bne.es». Consultato 12 febbraio 2017. <http://datos.bne.es/inicio.html>.
- «DBpedia». Consultato 9 febbraio 2017. <http://wiki.dbpedia.org/>.
- «DCMI History». Consultato 20 gennaio 2017. <http://dublincore.org/about/history/>.
- «DCMI Home: Dublin Core Metadata Initiative (DCMI)». Consultato 2 febbraio 2017.
<http://dublincore.org/>.
- «DCMI Metadata Terms». Consultato 20 gennaio 2017. <http://dublincore.org/documents/dcmi-terms/>.
- «Deutsche Nationalbibliothek - Linked Data Service». Consultato 12 febbraio 2017.
http://www.dnb.de/EN/Service/DigitaleDienste/LinkedData/linkeddata_node.html.
- «Dewey Web Services | OCLC Developer Network». Consultato 12 febbraio 2017.
<https://www.oclc.org/developer/develop/web-services/dewey-web-services.en.html>.
- «Digital Public Library of America» Policies». Consultato 20 gennaio 2017.
<https://dp.la/info/about/policies/>.
- «Dublin Core Metadata Element Set, Version 1.1». Consultato 20 gennaio 2017.
<http://dublincore.org/documents/dces/>.

«EDMObjectTemplatesProviders · europeana/corelib Wiki · GitHub». Consultato 9 febbraio 2017.
<https://github.com/europeana/corelib/wiki/EDMObjectTemplatesProviders#skosConcept>.

«Europeana Collections». Consultato 7 gennaio 2017.
<http://www.europeana.eu/portal/?locale=en>.

«Europeana Data Model Documentation». Consultato 18 gennaio 2017.
<http://pro.europeana.eu/page/edm-documentation>.

«Europeana Data Sources - Europeana Collections». Consultato 9 febbraio 2017.
<http://www.europeana.eu/portal/it/rights/data-sources.html>.

«Europeana semantic enrichment - Europeana Professional». Consultato 13 gennaio 2017.
<http://pro.europeana.eu/page/europeana-semantic-enrichment>.

«Europeana Strategy 2020: ‘We transform the world with culture’». Consultato 7 gennaio 2017.
<http://strategy2020.europeana.eu/it/>.

«Falcons Object Search». Consultato 14 febbraio 2017.
<http://ws.nju.edu.cn/falcons/objectsearch/index.jsp>.

«FOAF Vocabulary Specification». Consultato 8 gennaio 2017. <http://xmlns.com/foaf/spec/>.

«Formez PA | Centro servizi, assistenza, studi e formazione per l’ammodernamento delle P.A.»
Consultato 5 febbraio 2017. <http://www.formez.it/>.

«GeoNames». Consultato 9 febbraio 2017. <http://www.geonames.org/>.

«GeoNames Ontology - Geo Semantic Web». Consultato 5 febbraio 2017.
<http://www.geonames.org/ontology/documentation.html>.

«Glossary of Terms - Europeana Professional». Consultato 9 febbraio 2017.
<http://pro.europeana.eu/page/page/glossary>.

«Home | CIDOC CRM». Consultato 8 gennaio 2017. <http://www.cidoc-crm.org/>.

«Home | National Center for Supercomputing Applications at the University of Illinois».
Consultato 20 gennaio 2017. <http://www.ncsa.illinois.edu/>.

«Home - LC Linked Data Service (Library of Congress)». Consultato 12 febbraio 2017.
<http://id.loc.gov/>.

«Home - schema.org». Consultato 5 febbraio 2017. <http://schema.org/>.

«HTTP - Hypertext Transfer Protocol Overview». Consultato 3 febbraio 2017.
<https://www.w3.org/Protocols/>.

«IBC Linked Open Data — LodView». Consultato 5 febbraio 2017. <http://dati.emilia-romagna.it/id/ibc/website/html>.

«Incubation - Europeana Labs». Consultato 9 febbraio 2017.
<http://labs.europeana.eu/incubation>.

«index (data.bnf.fr)». Consultato 4 febbraio 2017. <http://data.bnf.fr/>.

«Introduction to the Europeana APIs - Europeana Labs». Consultato 9 febbraio 2017.
<http://labs.europeana.eu/api>.

«Italian Open Data License v1.0». Consultato 5 febbraio 2017. <http://www.formez.it/iodl/>.

«LIBRIS». Consultato 12 febbraio 2017. <http://libris.kb.se/>.

«Linee guida per l'accessibilità dei contenuti Web (WCAG) 2.0». Consultato 29 gennaio 2017.
<https://www.w3.org/Translations/WCAG20-it/>.

«Linked Data | Linked Data - Connect Distributed Data across the Web». Consultato 3 febbraio 2017. <http://linkeddata.org/>.

«Linked Jazz | Revealing the Relationships of the Jazz Community». Consultato 5 febbraio 2017. <https://linkedjazz.org/>.

«Linked open data — E-R Istituto per i beni artistici culturali e naturali». Consultato 5 febbraio 2017. <http://ibc.regione.emilia-romagna.it/servizi-online/lod>.

«Linked Open Data - W3C eGovernment Wiki». Consultato 3 febbraio 2017.
https://www.w3.org/egov/wiki/Linked_Open_Data.

«Linked Open Vocabularies». Consultato 4 febbraio 2017. <http://lov.okfn.org/dataset/lov/>.

«LodLive - browsing the Web of Data». Consultato 5 febbraio 2017. <http://lodlive.it/>.

«MAB - Musei Archivi Biblioteche». Consultato 4 febbraio 2017. <http://www.mab-italia.org/>.

«MARC in XML». Consultato 29 gennaio 2017. <https://www.loc.gov/marc/marcxml.html>.

«Multilingual Bibliographic Structure». Consultato 11 febbraio 2017.
<http://www.oclc.org/research/themes/data-science/multilingual-bib-structure.html>.

«Nasce LodView di diego-valerio-camarda | regista». Consultato 5 febbraio 2017.
<http://www.regesta.com/2014/12/19/nasce-lodview-il-web-dei-dati-ha-un-nuovo-volto/>.

«OCLC: Worldwide, member-driven library cooperative». Consultato 20 gennaio 2017.
<https://www.oclc.org/en/home.html>.

«Ontologies - W3C». Consultato 3 febbraio 2017.
<https://www.w3.org/standards/semanticweb/ontology>.

«Open Archives Initiative - Object Exchange and Reuse». Consultato 8 gennaio 2017.
<https://www.openarchives.org/ore/>.

«Open Knowledge: Home». Consultato 5 febbraio 2017. <https://okfn.org/>.

«OpenRefine». Consultato 5 febbraio 2017. <http://openrefine.org/>.

«OWL Web Ontology Language Reference». Consultato 5 febbraio 2017.
<https://www.w3.org/TR/2004/REC-owl-ref-20040210/#sameAs-def>.

«protégé». Consultato 14 febbraio 2017. <http://protege.stanford.edu/>.

«RDA Registry | Home». Consultato 5 febbraio 2017. <http://www.rdaregistry.info/>.

«RDF - Semantic Web Standards». Consultato 29 gennaio 2017. <https://www.w3.org/RDF/>.

«RDF 1.1 XML Syntax». Consultato 21 gennaio 2017. <https://www.w3.org/TR/rdf-syntax-grammar/>.

«RDF Schema 1.1». Consultato 5 febbraio 2017. <https://www.w3.org/TR/rdf-schema/>.

«RDFa 1.1 Primer - Third Edition». Consultato 15 febbraio 2017.
<https://www.w3.org/TR/xhtml-rdfa-primer/#the-basics-of-rdfa-rdfa-lite>.

«RDF/XML Syntax Specification (Revised)». Consultato 15 febbraio 2017.
<https://www.w3.org/TR/REC-rdf-syntax/>.

«Research». Consultato 22 gennaio 2017. <http://www.oclc.org/research.html>.

«Resource Description Framework (RDF) Model and Syntax Specification». Consultato 21 gennaio 2017. <https://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>.

«ricerca per termine». Consultato 3 febbraio 2017. <http://thes.bncf.firenze.sbn.it/ricerca.php>.

«Semantic web - NektarWiki». Consultato 12 febbraio 2017.
http://nektar.oszk.hu/wiki/Semantic_web.

«semanticwebsearch.com». Consultato 14 febbraio 2017.
<http://www.semanticwebsearch.com/query/>.

«Share Catalogue». Consultato 12 febbraio 2017. <http://catalogo.sharecat.unina.it/sharecat/clusters>.

«Silk - The Linked Data Integration Framework». Consultato 5 febbraio 2017.
<http://silkframework.org/>.

«SKOS Simple Knowledge Organization System - home page». Consultato 8 gennaio 2017.
<https://www.w3.org/2004/02/skos/>.

«SPARQL Endpoint - Europeana Labs». Consultato 9 febbraio 2017.
<http://labs.europeana.eu/api/linked-open-data-SPARQL-endpoint>.

«SPARQL Query Language for RDF». Consultato 3 febbraio 2017. <https://www.w3.org/TR/rdf-sparql-query/>.

«SparqlEndpoints - W3C Wiki». Consultato 3 febbraio 2017.
<https://www.w3.org/wiki/SparqlEndpoints>.

«The Linking Open Data cloud diagram». Consultato 3 febbraio 2017. <http://lod-cloud.net/>.

«The Open Definition - Open Definition - Defining Open in Open Data, Open Content and Open Knowledge». Consultato 5 febbraio 2017. <http://opendefinition.org/>.

«Themes list». Consultato 9 febbraio 2017. <http://www.eionet.europa.eu/gemet/>.

«Turbomarc, faster XML for MARC records | Index Data». Consultato 2 febbraio 2017.
<http://www.indexdata.com/news/2014/06/turbomarc-faster-xml-marc-records>.

«URIs, URLs, and URNs: Clarifications and Recommendations 1.0». Consultato 21 gennaio 2017. <https://www.w3.org/TR/uri-clarification/>.

«VIAF». Consultato 18 gennaio 2017. <https://viaf.org/>.

«VIAF (Virtual International Authority File) | OCLC Developer Network». Consultato 18 gennaio 2017. <http://www.oclc.org/developer/develop/web-services/viaf.en.html>.

«W3C HTML». Consultato 2 febbraio 2017. <https://www.w3.org/html/>.

«W3C Library Linked Data Incubator Group». Consultato 19 gennaio 2017.
<https://www.w3.org/2005/Incubator/lld/>.

«W3C WebApps Working Group». Consultato 2 febbraio 2017.
<https://www.w3.org/2008/webapps/>.

«Watson Semantic Web Search». Consultato 14 febbraio 2017.
<http://watson.kmi.open.ac.uk/WatsonWUI/>.

«When we share, everyone wins - Creative Commons». Consultato 20 gennaio 2017.
<https://creativecommons.org/>.

«Wikidata». Consultato 9 febbraio 2017. https://www.wikidata.org/wiki/Wikidata:Main_Page.

«Wikidata:Introduzione - Wikidata». Consultato 11 febbraio 2017.
<https://www.wikidata.org/wiki/Wikidata:Introduction/it>.

«WikiLeaks». Consultato 27 gennaio 2017. <https://wikileaks.org/>.

«World Wide Web Consortium (W3C)». Consultato 28 gennaio 2017. <https://www.w3.org/>.

«WorldCat Entities | OCLC Developer Network». Consultato 11 febbraio 2017.
<https://www.oclc.org/developer/develop/linked-data/worldcat-entities.en.html>.

«XML Essentials - W3C». Consultato 2 febbraio 2017. <https://www.w3.org/standards/xml/core>.

«Z39.50 Maintenance Agency Page». Consultato 28 gennaio 2017.
<https://www.loc.gov/z3950/agency/>.

«Zeri&LODE». Consultato 12 febbraio 2017. <http://data.fondazionezeri.unibo.it/>.

Ringraziamenti

Ringrazio di cuore le due meravigliose persone con cui vivo, la mia cara nonna e i miei genitori, per l'affetto e il supporto incondizionati.

Sono molto riconoscente a Rosanna Chinaglia, Agnese Galeffi, Lucia Sardo, Dario Rigolin e Paolo Pezzolo, per i suggerimenti, i consigli e il prezioso confronto. Con voi la biblioteconomia è sempre avvincente.