

Biblioteche e linked data

Un convegno internazionale a Firenze

Il 18 e 19 giugno si è svolto a Firenze presso l'aula magna dell'Università degli studi il convegno internazionale "Global interoperability and linked data in libraries", il primo in Italia sul tema dei linked data. Il convegno ha suscitato un interesse enorme tra gli addetti ai lavori (bibliotecari, docenti di scienze dell'informazione, *software analysts*) e ha registrato la partecipazione di oltre 300 persone. Tra gli argomenti affrontati: i linked data, i linked open data, il web semantico, l'interoperabilità, le licenze, gli standard delle biblioteche e il loro *data mapping* in RDF.

Il convegno è stato aperto dai saluti di Mauro Guerrini, docente di biblioteconomia presso l'Università di Firenze, organizzatore del convegno, che ai presenti ha spiegato come i linked data forniscano un nuovo linguaggio di comunicazione in grado di far dialogare e interagire mondi diversi: dalla pubblica amministrazione al settore privato, agli istituti culturali: archivi, biblioteche e musei.

Il paradigma concettuale dei linked data e del web semantico, ha ribadito Guerrini, scaturisce dalla mente di Tim Berners-Lee, inventore del World Wide Web, che nel 2001, a dieci anni dalla nascita ufficiale della rete delle reti, sottolineava come la sua idea originaria del web fosse quella di un modello in cui la rete è vista come un grafo etichettato dove "i nodi rappresentano documenti o meglio dati e sono collegati tramite

relazioni espressive e non piatte".¹ Il web semantico non è in contrapposizione al web tradizionale, ma è un'estensione di quello attuale "in which information is given in well-defined meaning, better enabling computers and people to work in cooperation".²

Nel web semantico l'accento è posto non più sui documenti, ma sui dati. I dati devono essere strutturati e messi in relazione, le relazioni sono esplicitate attraverso regole (*rules*) che rendono i contenuti accessibili a processi automatizzati. Lo stesso Berners-Lee propone una scala crescente di valori e di complessità per mettere a disposizione i dati sul web:

- mettete a disposizione i vostri dati sul web (in qualsiasi formato "così come sono");
- metteteli a disposizione come dati strutturati (per esempio in formato Excel, anziché come immagini scannerizzate di una tabella);
- utilizzate formati non proprietari/aperti (ad esempio CSV anziché Excel);
- usate degli URL (stabili) per identificare gli oggetti, così che le persone possano creare puntatori alle vostre risorse;
- collegate i vostri dati ai dati di altre persone per permettere di capire in modo automatico il contesto (cioè qual è il loro contenuto).

La crescita dei linked data sul web è stata negli ultimi anni esponenziale;³ anche le biblioteche, in quanto forte-

mente interessate dall'integrazione dei dati e delle risorse in rete, stanno sperimentando in modo concreto il paradigma dei linked data e molte delle esperienze realizzate sono state oggetto delle relazioni del convegno. La relazione introduttiva di Karen Coyle, bibliotecaria consulente di biblioteche digitali, ha offerto la vera chiave di lettura per l'intero convegno. La Coyle ha introdotto alcuni concetti basilari per il tema dei linked data applicato al mondo delle biblioteche. La rete, ha detto la Coyle, ha rivoluzionato il mondo dell'informazione e le biblioteche devono affrontare il cambiamento per continuare ad esistere. Nonostante gli sforzi compiuti dai bibliotecari le biblioteche appaiono ancora lontane dal web, i loro dati sono isolati dal web, gli utenti sono sempre più remoti e soddisfano le loro esigenze informative attraverso fonti diverse e sempre meno attraverso le biblioteche. In questo scenario per entrare realmente nel web i cataloghi di biblioteca devono cessare di essere un'entità staccata, un database separato e divenire un insieme di dati strutturati del web. Si tratta di un cambiamento di paradigma per i cataloghi delle biblioteche il cui compito è di creare un nuovo sistema di accesso e d'uso dei dati bibliografici che sia compatibile con il funzionamento della rete. Due sono le strade percorribili per sostenere questo cambiamento:

- a) rendere disponibili e riutilizzabili i propri dati bibliografici in rete;
- b) riutilizzare i dati per creare nuovi servizi per gli utenti delle biblioteche.

La Coyle ha portato come esempio di linked data il soggettario della Library of Congress che è già disponibile nella forma di linked data così come anche i soggettari delle biblio-

teche nazionali di Francia, Germania e Giappone; i dati di autorità relativi ai nomi si trovano collegati in un sistema chiamato Virtual International Authority File, o VIAF. VIAF riceve i record di autorità da circa venti biblioteche nel mondo; sono già stati tradotti in standard *semantic web* anche ISBD, FRBR, FRAD e RDA. Queste esperienze di linked data in ambito biblioteconomico sono importanti anche perché permettono alle biblioteche di aprirsi verso l'esterno e di arricchire i propri dati con i dati e/o metadati generati da altre comunità e risorse, sia che provengano dalla ricerca scientifica, da enti governativi o da imprese commerciali e for-profit. La strada da percorrere è, quindi, ancora lunga.

Michele Barbera (Net7) ha introdotto alcuni concetti tecnici basilari del web semantico e del mondo dei linked data: ad esempio il concetto di tripla che è un'asserzione costituita da soggetto, predicato e valore (in inglese spesso identificati come *subject, verb e object*). Barbera ha spiegato come i dati da soli nel web siano scarsamente significativi ed abbiano poco valore, mentre assumono un significato se divengono *machine-readable*, sono standardizzati e messi tra loro in relazione. Le triple vengono utilizzate concretamente per creare le relazioni tra i dati; i linked data sono a loro volta una tecnologia per la realizzazione dell'idea di web semantico.

Grazie al web semantico e al paradigma dei linked data è possibile pensare ai dati in un modo concettuale differente: non più in formato tabellare (excel o csv), ma con una struttura relazionale. Le sfide da affrontare sono molteplici come, ad esempio, quella di essere in grado di tracciare sempre l'origine del dato e l'evoluzione dei grafi.

Sulle caratteristiche di RDF si è soffermata la relazione di Tom Baker (Dublin Core Metadata Initiative). RDF, ha detto Baker, è la grammatica del web semantico e definisce una struttura per esprimere frasi secondo il modello tripartito delle triple. Le triple possono descrivere un rapporto tra due entità o tra un'entità e delle annotazioni testuali che la descrivono. Una frase RDF è composta da URI, indirizzi che nella lingua del web semantico vengono utilizzati come parole per nominare persone, luoghi, epoche storiche, servizi di ogni tipo. L'assegnazione di indirizzi URI agli oggetti della descrizione bibliografica (opere, pubblicazioni, persone ma anche proprietà, intestazioni di soggetto) agevola la traduzione del lavoro delle biblioteche. Queste ultime possono e devono diventare nodi centrali nella crescente rete dei linked data. Una diversa visione del mondo dei linked data è stata portata dagli editori, attraverso la relazione di Paola Mazzucchi, di mEDRA, l'agenzia europea che fornisce servizi agli editori (ad esempio il DOI). Mazzucchi ha discusso in modo particolare del concetto di dati aperti che di per sé non è un aspetto problematico per gli editori a patto che si riconoscano le dovute differenze e si affronti il tema dei diritti che resta estremamente delicato e problematico. Gli editori chiedono di trovare un approccio cooperativo a queste tematiche. Sono possibili infatti diversi modelli di business e *licensing agreement*. Dal momento che i dati sono di diversa natura alcuni saranno accessibili apertamente, altri tramite accordi, altri ancora saranno messi a disposizione a pagamento.

Alex Kaschte (Ex Libris) ha, invece, proposto il punto di vista di un agente commerciale sul mondo dei

linked data. Presentando l'ultimo prodotto di Ex Libris, il discovery PRIMO, Kaschte ha descritto alcune applicazioni di linked data in PRIMO, ad esempio il mash da Wikipedia (attraverso DBpedia che espone una grande porzione dei dati di Wikipedia come linked data) relativo ai nomi degli autori. Kaschte ha anche portato alcuni esempi di applicazioni di linked data in biblioteche: il VIAF, il thesaurus AGROVOC basato su SKOS, Europeaana. Fonti diverse che espongono i loro dati in modo diverso ed in formati diversi e per la maggior parte non seguono il modello RDF che è lo standard di riferimento per la codifica, lo scambio ed il riutilizzo dei dati. La sfida per chi voglia riutilizzare i dati esposti come linked data è appunto questa. Ex Libris partecipa al progetto *Digitised Manuscripts to Europeaana* attraverso il Work Package 2 che è quello dedicato all'interoperabilità e si propone di aggregare in Europeaana i differenti set di metadati prodotti da diversi data provider.

Un particolare caso applicativo di linked data è stato proposto da Tommaso Agnoloni dell'ITTIG-CNR, Istituto di teorie e tecniche dell'informazione giuridica.⁴ Agnoloni ha parlato delle potenzialità dei linked data per l'informazione giuridica, che è costituita da risorse eterogenee e distribuite tra Giurisprudenza, Legislazione e Dottrina. Esiste, infatti, in ambito giuridico, una forte esigenza di interconnessione tra le diverse tipologie di dati giuridici. Proprio per questa esigenza e per questa natura peculiare dell'informazione giuridica sono già stati sviluppati standard per l'identificazione dei documenti giuridici (URN, European Case Law Identifier, CELEX, UK Registration Iden-

tifier), standard per l'annotazione XML di documenti giuridici (NIR (IT), Metalex (NL), Metalex.eu (EU) ecc.) e risorse semantiche giuridiche come il Thesaurus EUROVOC. L'idea è quella di utilizzare la tecnologia linked data per esporre in rete i dati della banca dati DOGI che offre abstract di articoli pubblicati da riviste giuridiche italiane. A tal fine l'ITTIG ha realizzato uno schema DOGI in RDF. L'obiettivo è una conversione dei dati che preservi la struttura e il contenuto originale della banca dati giuridica ma che sia allo stesso tempo aperta all'internazionalizzazione e all'interconnessione con dataset esterni. L'approccio è quello di riutilizzare vocabolari e identificatori standard, di creare delle estensioni specializzate di dominio, di mappare e convertire i dati giuridici DOGI.

Tiziana Possemato (Atcult) e Mauro Guerrini hanno ripreso alcuni concetti chiave del web semantico e dei linked data. Presentando la relazione, Tiziana Possemato ha esposto i motivi che avvicinano il mondo delle biblioteche, degli archivi e dei musei agli sviluppi dei linked data. Sia il mondo delle biblioteche, infatti, che il mondo di internet sono interessati dall'integrazione in rete. Per le biblioteche è importante, tuttavia, chiedersi se i metadati e i formati MARC utilizzati fino ad oggi siano ancora funzionali, se nonostante la loro qualità i dati delle biblioteche non restino un'esposizione di nicchia circoscritta ad un dominio altamente professionalizzato. Il report pubblicato dal Library of Congress Working Group on the Future of the Bibliographic Control fornisce una *roadmap* per trasformare i record tradizionali di un catalogo elettronico in linked data:

- trasformare la descrizione testua-

le in set di dati usabili per processi ed elaborazioni automatiche;

- rendere gli elementi di dati univocamente identificabili all'interno del contesto informativo del web;
- far sì che i dati siano compatibili con le tecnologie e gli standard del web;
- usare un linguaggio trasversale e interoperabile nella realtà del web.

Un aspetto particolarmente rilevante nel mondo dei linked data è l'identificazione univoca degli oggetti digitali e dei metadati loro associati a mezzo di identificativi persistenti (*persistent identifiers*) grazie ai quali i dati possono essere riutilizzati in contesti di diverso tipo. Di identificativi persistenti hanno discusso Maurizio Lunghi della Fondazione Rinascimento Digitale e Jan Brase della German National Library of Science and Technology. Lunghi ha descritto le caratteristiche del National Bibliography Number (NBN:IT) che permette di identificare in modo univoco una risorsa digitale. Il sistema NBN è associato, da un lato al deposito legale, dall'altro alla conservazione a lungo termine delle risorse digitali che vengono "harvestizzate" dalla Biblioteca nazionale centrale di Firenze nel server di Magazzini Digitali che funge da *trusted repository* per le risorse e i metadati associati. Le istituzioni italiane che vogliono adottare NBN devono firmare un apposito contratto con la Biblioteca nazionale centrale di Firenze per usufruire del servizio di deposito legale delle risorse digitali da loro prodotte e non associate ad un circuito commerciale. NBN non è alternativo rispetto agli altri identificativi persistenti quali il DOI, adottato dagli editori, o l'Handle System, ma li affianca. È necessario che una risorsa in rete sia dotata di almeno un identificativo

persistente ma, laddove ci sia l'opportunità, è possibile associare ad una stessa risorsa più di un identificativo dal momento che ciascuno porta con sé una propria specificità. Jan Brase ha presentato, invece, il progetto DataCite, un consorzio nato nel 2009 e finanziato dal governo federale tedesco, che ha realizzato un differente tipo di identificativo persistente concepito allo scopo di identificare e citare i dati del web. Grazie al progetto DataCite i dati acquisiscono una nuova visibilità e sono riutilizzabili; si annullano le duplicazioni; i risultati della ricerca sono più completi e diventano realmente confrontabili.

È percepibile, in questo scenario nel quale i dati assumono una rilevanza particolare, il cambiamento di prospettiva indotto dai linked data e dal paradigma del web semantico il cui focus non è più sui documenti di per sé statici e non interoperabili, ma sui dati strutturati e tra loro collegati. DataCite è mantenuto dalla German National Library of Science and Technology. Tra i membri del consorzio, per l'Italia, la Conferenza dei rettori delle università italiane (CRUI).

Un nutrito numero di relazioni ha descritto i progetti applicativi e il *data mapping* in RDF di standard e formati del mondo bibliotecario.

Pat Riva, Elena Escolano Rodriguez, Alan Danskin, Joan Mitchell e Kevin Ford hanno affrontato, rispettivamente, i temi della mappatura in RDF di FRBR, ISBD, RDA, CDD e Library of Congress Classification. Pat Riva (Bibliothèques et Archives Nationales du Québec) ha parlato dell'evoluzione del progetto "Namespaces for FRBR enties/elements in RDF" (Spazi di nomi per le entità /elementi di FRBR in RDF) lanciato dal FRBR Working Group

dell'IFLA, che dal 2009 ha tra i suoi compiti il mandato di mantenere, sviluppare e consolidare i tre modelli FRBR, FRAD e FRSAD. I requisiti per la gestione dei *namespaces* dell'IFLA sono la loro disponibilità in un ambiente aperto, la possibilità di offrire servizi di deferenza e la loro gestione all'interno di un quadro comune. Dopo il Congresso IFLA 2009 è stato registrato il dominio <iflstandards.info> e l'IFLA Professional Committee ha creato il gruppo di lavoro denominato IFLA Namespaces Task Group che ha predisposto il modello per ricavare domini di base per ogni *namespace*. Dal 2009 FRBR ha sviluppato il contenuto dei *namespaces* ed il set di elementi FRBRer è stato pubblicato a maggio 2011. Contiene 10 classi (entità) e 206 proprietà (attributi e relazioni). Gli ulteriori vincoli semantici sono espressi da circa 200 triple in TTL.

Elena Escolano Rodriguez, ex membro IFLA ISBD Review Group, ha discusso ancora una volta brevemente delle sfide del web semantico. Le esperienze applicative del web semantico si sono concentrate fino ad oggi in prevalenza sulle ontologie formali, ha detto la Escolano Rodriguez. Le disposizioni e gli standard dell'IFLA assicurano la qualità, la velocità di scambio e la sostenibilità dei dati prodotti dalle biblioteche e costituiscono, quindi, un terreno fertile per il web semantico. Da alcuni anni l'IFLA ha deciso di inserire la dichiarazione dei propri modelli e standard in RDF. Tra gli standard mappati: ISBD e il modello relazionale FRBR. Al fine di esporre i dati nel web è necessario creare un *namespace* che identifichi senza ambiguità gli URI delle disposizioni in RDF. Gli URI creati devono essere corti, chiari, espandibili e applicabi-

li a ciascun modello e standard. Per rispondere alla necessità di una neutralità linguistica gli URI devono, inoltre, essere opachi. La base del web semantico, infatti, è in inglese, mentre IFLA considera importanti gli aspetti del multilinguismo per garantire la diversità linguistica e culturale. All'interno del piano di lavoro del gruppo ISBD/XML dell'IFLA al primo punto si trova la decisione di sostenere la traduzione sia dello standard ISBD sia del set di elementi ISBD registrati nell'Open Metadata Registry, lo spazio che il W3C ha creato e nel quale si trovano descritte molte ontologie dei più diversi ambiti.

Alan Danskin (British Library) ha parlato dell'evoluzione di RDA, il nuovo standard per la descrizione delle risorse che sostituisce le AACR2. RDA è uno standard aperto, nel senso che ogni parte interessata può contribuire al suo sviluppo. RDA è aperto ma non è gratuito perché deve essere sostenibile nel lungo termine. Una parte del suo contenuto è stata resa disponibile gratuitamente come linked data. L'insieme di elementi e i vocabolari di RDA sono in corso di pubblicazione nell'Open Metadata Registry. RDA è stato creato per essere indipendente da ogni schema o formato di metadati specifico. È basato sulla famiglia FR dell'IFLA.

Joan Mitchell e Michael Panzer (OCLC) hanno descritto i lavori di sviluppo che hanno condotto all'esposizione del sistema di classificazione Dewey come linked data. Gli obiettivi iniziali del progetto erano di:

- fornire un URI processabile per ogni classe;
- codificare la semantica della classificazione in RDF/SKOS;
- fornire rappresentazioni leggibili dalla macchina;

- rendere i dati riutilizzabili grazie a licenze di tipo aperto.

Il dominio Dewey.info contiene rappresentazioni Dewey in RDF utili per la macchina. La versione per la macchina ha tre serializzazioni RDF: RDF/XML, Turtle e JSON.

Quanto alle licenze, i dati sono stati resi disponibili con licenza Creative Commons BY-NC-ND.

Kevin Ford (Library of Congress) ha, invece, affrontato il tema del data mapping in RDF della Library of Congress Classification, che ha di recente pubblicato alcune delle 21 classi che la compongono come linked data.

Estremamente interessante per l'Italia la relazione di Giovanni Bergamin e Anna Lucarelli (Biblioteca nazionale centrale di Firenze), che hanno descritto gli sviluppi del Nuovo Soggettario (NS) di Firenze, che contiene 46.000 termini e accoglie e gestisce terminologia proposta da altre biblioteche ed istituzioni culturali. I termini sono organizzati secondo categorie e relazioni previste dagli standard ISO sui thesauri. Dal 2010 sono disponibili i metadati in RDF/SKOS del NS per favorirne l'utilizzazione come linked data.

SKOS (Simple Knowledge Organization System) è un *data model* sviluppato dal W3C Semantic Web Deployment Working Group per favorire l'interoperabilità tra i sistemi di organizzazione della conoscenza. Bergamin e Lucarelli hanno descritto alcuni problemi della conversione del Nuovo Soggettario in RDF come, ad esempio, il fatto che il modello SKOS non preveda sistemi intermedi di divisione e non stabilisca un collegamento tra i membri di un raggruppamento e il concetto generale che definisce il raggruppamento. Alcune difficoltà ha comportato anche la traduzio-

ne della relazione di variante storica e di scomposizione. Dal 2011 è in corso il lavoro sperimentale per l'impiego del NS nell'indicizzazione semiautomatica delle risorse digitali acquisite per deposito legale dalla Biblioteca nazionale centrale di Firenze.

Federico Morando, del Centro Nexa su internet & società del Politecnico di Torino, ha introdotto il tema delle licenze per i dati aperti, chiarendo come creare un'architettura di open data significhi confrontarsi con le difficoltà dell'accessibilità e del riuso del dato, che deve essere definita attraverso l'attribuzione di specifiche licenze. Le licenze Creative Commons, ampiamente utilizzate nel mondo delle pubblicazioni per definire la fruibilità di documenti, sono unanimemente considerate non idonee per un'applicazione agli open data, con l'eccezione della licenza CC0, corrispondente alla licenza PD (Pubblico Dominio).

Il quadro normativo delle licenze per Open Data è caratterizzato dalla loro frammentazione dovuta all'annoso problema della differenziazione dei sistemi legislativi nazionali. Se, da un lato, organismi quali la Open Knowledge Foundation hanno elaborato specifiche licenze, quali Open Data Commons, con lo scopo di uniformare e normalizzare il panorama delle licenze per i dati aperti, molti enti governativi nazionali hanno elaborato propri sistemi di licenza che rispondono al quadro legislativo del paese di riferimento. È il caso delle licenze Open Government Licence (Gran Bretagna), Licence Ouvert (Francia), Italian Open Data License (Italia). Morando ha messo in evidenza come un panorama così frastagliato comporti una forte limitazione all'interoperabilità dei dati, garantita uni-

camente da una licenza completamente aperta come la CC0.

Gli aspetti fin qui delineati definiscono un quadro in evoluzione del mondo dei cataloghi bibliografici la cui esigenza è conciliare gli aspetti identificativi del dato con la sua diffusione e condivisione. Se fino alla fine dello scorso secolo il compito principale del catalogo era infatti quello di identificare e localizzare in modo univoco il dato bibliografico, ora nel mondo del web 2.0 essere "in rete" diventa un imperativo. Il web ha cambiato le modalità di fruizione delle informazioni e delle risorse, così che il catalogo, che aveva fatto del controllo e della validazione del dato il suo punto di forza, deve ora aprire i suoi dati al web consentendo agli utenti la navigazione fra le informazioni.

Su questa consapevolezza si sono mossi i vari progetti che le biblioteche, in particolare quelle nazionali, stanno portando avanti nell'aprire i loro cataloghi e metterli in rete.

Le tre esperienze di Open Bibliographic Data avviati dalla Biblioteca nazionale di Francia, dalla Biblioteca nazionale della Baviera e dalla Biblioteca nazionale di Svezia, si muovono tutte su paradigmi ormai comuni e condivisi. Il primo è l'RDf come linguaggio del *semantic web* e il conseguente lavoro di mappatura del catalogo in RDf, un lavoro spesso lungo e complesso come ha spiegato Gabriel Messmer della Biblioteca nazionale della Baviera, ma fondamentale per rendere i propri dati interoperabili, comunicabili e ricercabili via web. Il secondo è l'attribuzione di un *persistent identifier* agli oggetti bibliografici, siano questi dati di authority, di descrizione bibliografica, di natura semantica, amministrativi o di immagini, che consenta un'identificazione uni-

voca del dato permettendo la costruzione di un'architettura di *linking* che, come ha affermato Martin Malmsten della Biblioteca nazionale di Svezia, annulli la differenza fra "interno e esterno", tra il posseduto e l'informazione ad esso relativa presente in rete. Il terzo paradigma è l'aspetto legale di attribuzione delle licenze, sempre in bilico tra il concetto di apertura e protezione, quindi tra il riuso totale del dato, garantito da licenze di pubblico dominio come la CC0 e la scelta di licenze che consentono un riuso parziale, trattenendo per sé alcuni diritti che garantiscono uno sfruttamento anche di tipo commerciale.

Roman Wenz ha presentato <data.bnf.fr>⁵ il catalogo LOD della Biblioteca nazionale di Francia che integra i dati del catalogo della BNF con quello Archivi e manoscritti e con Gallica per le opere digitalizzate, collegando così fra loro i dati bibliografici provenienti da varie fonti interne con il mondo del web, in particolare Europea e Wikipedia. Martin Malmsten ha illustrato LIBRIS⁶ evoluzione dello Swedish Union Catalogue nel mondo del web semantico. Malmsten ha sottolineato l'importanza dei *related link* per collegare, partendo da un unico punto di accesso, informazioni diverse ma interconnesse. Ciò consente di partire dal titolo di un'opera o dal nome di un autore e da qui navigare raggiungendo altre opere affini per tema, manifestazioni diverse della stessa opera, altre opere di un autore o a lui collegate, e ancora raggiungere pagine di Google e Wikipedia.

Gabriel Messmer ha presentato il progetto Culturegraph⁷ sviluppato con lo scopo di fornire un servizio per la progettazione, implementazione e gestione di Linked Open Data services per il mondo delle bi-

biblioteche tedesche. Il lavoro di analisi e studio del progetto ha portato alla redazione di una guida sugli Open Data,⁸ dalla quale emergono tre tematiche principali: l'attenzione alle soluzioni e architetture di mappatura e *linking* dei metadati; l'importanza del concetto di *catalog enrichment* nel quale le informazioni provenienti dall'esterno (il mondo del web) si integrano con quelle delle catalogo e l'adesione del progetto a Europeana Libraries; il tema delle licenze con la scelta finale di adozione della licenza CC0 per garantire la massima interoperabilità.

L'esempio della Biblioteca Vaticana portato da Paola Manoni riguarda l'integrazione dei cataloghi bibliografici con quelli di opere digitalizzate. Il progetto ha previsto due fasi, la prima relativa alla realizzazione di un catalogo integrato in grado di gestire metadati di formato differente: MARC21 per il materiale bibliografico, TEI-MS per i manoscritti e EAD per il materiale archivistico. La seconda fase, ancora *in fieri*, prevede la creazione di una biblioteca di oggetti digitali gestiti con lo standard METS al quale viene associato un *persistent identifier*.

La Commissione europea da anni sta finanziando progetti indirizzati alla raccolta, conservazione e riuso di dati che abbracciano il panorama degli istituti culturali, dalle biblioteche agli archivi e ai musei. L'aspetto evidente è che se, da un lato, i contenuti sono simili, dall'altro i metadati descrittivi, le politiche di gestione, di *digital preservation* e di diffusione rispondono a modelli, architetture e linguaggi differenti e spesso non comunicabili tra loro. L'interoperabilità è, quindi, diventato uno degli obiettivi strategici della politica culturale della CE.

Marie Veronique Leroi (Ministero della cultura e della comunicazione, Francia) ha illustrato le azioni della CE in questo ambito delineando il percorso dei progetti a supporto di Europeana, da Minerva e Michael ad Athena per arrivare a Linked Heritage, il progetto lanciato nell'aprile 2011 per supportare il recupero e il riuso dei contenuti di Europeana. Si tratta di progetti il cui ruolo principale è l'analisi di modelli e standard per l'elaborazione di linee guida e *best practices* che facciano da sfondo teorico e pratico al tema dei contenuti culturali.

Con Minerva⁹ vengono emanate le prime linee per il trattamento di oggetti digitali, ma è con Michael¹⁰ che viene introdotto il tema dei linked data finalizzato alla realizzazione di un catalogo di opere digitalizzate. In Athena¹¹ diventa centrale il problema del multilinguismo e dell'individuazione di terminologie e tassonomie capaci di comunicare tra loro e di definire standard descrittivi. È in questo ambito che il web semantico e i linked data prendono corpo come strumenti per la condivisione di vocabolari controllati attraverso l'utilizzo dello standard SKOS.

Il progetto Linked Heritage,¹² illustrato da Gordon McKenna (Collections Trust, UK), riunisce i temi già analizzati nei progetti precedenti con il fine di dare un quadro complessivo dello stato dei lavori in merito a linked data, *persistent identifiers*, metadati e standard, vocabolari controllati. McKenna si è soffermato in particolare sull'analisi del questionario teso a rilevare il panorama dell'utilizzo e della conoscenza di architetture di *linking* tra i partecipanti al progetto Europeana con particolare attenzione al tema delle licenze per i dati aperti e dei progetti

di *linking* dell'*open data cloud* tra cui Data Hub,¹³ il portale di condivisione dati promosso da CKAN (Comprehensive Knowledge Archive Network). L'analisi conferma e sottolinea i seguenti aspetti:

- prevedere l'assegnazione di una licenza per il riuso dei dataset;
- prevedere l'assegnazione di un *persistent identifier* per i dataset;
- sostenere la filosofia del *linked culture cloud* per collegare in rete i vari contenitori di oggetti digitali.

Di grande attualità è anche il tema dei dati aperti nelle pubbliche amministrazioni, fortemente sostenuto dal Ministero dell'istruzione, dell'università e della ricerca attraverso le politiche per l'e-government e l'accesso ai dati pubblici da parte dei cittadini. Su questo solco si muovono le due esperienze portate dalla Regione Piemonte e dal Comune di Firenze.

Gli obiettivi delle due esperienze si muovono lungo le due direttive di incremento di servizi per il mondo produttivo e per la società civile. La prima è quella di rendere le informazioni e i dati prodotti dalle PA fruibili e accessibili alle imprese per sviluppare nuove idee e creare business. La seconda è di consentire l'accesso ai singoli cittadini rispondendo a esigenze informative e di trasparenza.

Il progetto DATI. Piemonte.it¹⁴ della Regione Piemonte è il primo realizzato in Italia in questo ambito e ad esso molte altre PA stanno facendo riferimento. Il progetto è all'avanguardia anche sul piano internazionale come testimonia la partecipazione in veste di coordinatore della Regione Piemonte al progetto HOMER (Harmonising Open data in the MEditerranean through better access and Re-use of public sector information). Nella relazione

presentata dal responsabile del progetto, Roberto Moriondo, e illustrata in sua vece da Federico Morando sono state indicate le tre linee di azione lungo le quali si è mossa la Regione Piemonte:

- il piano normativo, attraverso un percorso di recepimento delle normative CE e nazionali in materia di accesso ai dati delle PA e loro riutilizzo, conclusosi recentemente con la promulgazione della prima legge regionale sull'argomento "Disposizioni in materia di pubblicazione e riutilizzo dei dati e delle informazioni dell'Amministrazione regionale" (LR 24 del 23 dicembre 2011);
- il piano organizzativo, attraverso il coinvolgimento partecipato e collaborativo con altre PA in particolare con la Regione Emilia Romagna, con i comuni piemontesi (accordo ANCI-Regione Piemonte) e con Unioncamere Piemonte. Il portale DATI.Piemonte.it, consente ai comuni e Unioncamere di depositare e aggiornare i propri dati utilizzando uno specifico spazio;
- il piano tecnologico, attraverso la realizzazione dell'architettura tecnica del portale. Gli obiettivi si muovono nella direzione dell'integrazione federata di portali di dati aperti secondo il modello CKAN e la loro messa in relazione attraverso l'architettura di linking (in fase di sperimentazione).

Gianluca Vannuccini, nel presentare il progetto del Comune di Firenze, ha posto l'accento sull'architettura tecnologica del portale OpenData,¹⁵ costruito con il software open source D2R e orientato ai linked data. In rapporto alla composizione dell'utenza sono state previste due tipologie di dati con conseguente modo differente di fruizione. Il pri-

mo gruppo di dati è orientato alle imprese: i dati sono caratterizzati da maggiore completezza e possono essere richiesti e confezionati *ad hoc* in base alle necessità espresse. Il comune di Firenze ha voluto così proporre la logica dell'accesso dei dati delle PA come servizio alle imprese, integrato con applicazioni che ne garantiscono la fruibilità anche da smartphone. Il secondo gruppo di dati è rivolto ai cittadini: si tratta di dati più semplici orientati all'acquisizione di informazioni e non specificatamente al riuso. Si è voluto qui porre l'accento sull'interazione dei cittadini con il portale integrandolo con strumenti di *social networking* come Twitter.

Conclusioni

Il convegno "Global Interoperability and linked data in libraries" ha offerto una panoramica degli open linked data come strumento tecnologico per la realizzazione del web semantico, ne ha esaminato l'architettura sottolineando come l'interoperabilità di standard e linguaggi sia il focus dal quale è impossibile prescindere e analizzando le prospettive nell'ambito degli istituti culturali, biblioteche, archivi, pubblica amministrazione. Nel programma, ricco di interventi, non è stato trattato il tema degli Open Data Repository e, più in generale, il tema del trattamento dei dati frutto della ricerca scientifica e della loro interazione con il mondo delle pubblicazioni scientifiche. Un tema ampio e di rilievo che sta guadagnando l'attenzione degli addetti ai lavori e del mondo accademico e potrebbe essere argomento di un futuro convegno sugli Open Data.

NOTE

¹ ILARIA PESCHINI - WALTER VOLPI, *Open data: i dati pubblici resi pubblici*, "Archivi e computer", 22 (2012), n. 1, p. 128-140.

² TIM BERNERS-LEE - JAMES HENDLER - ORA LASSILA, *The semantic web*, "Scientific American", May 2001, disponibile alla URL <<http://www.scientificamerican.com/article.cfm?id=the-semantic-web>>.

³ Si veda a riguardo il sito Linkeddata.org (<http://linkeddata.org/>) che fornisce un quadro dello stato corrente ed aggiornato della "Linked data cloud".

⁴ La relazione del CNR ITTIG è stata curata da Tommaso Agnoloni, Elisabetta Marinai, Ginevra Peruginelli, Maria Teresa Sagri, Daniela Tiscornia.

⁵ <<http://data.bnf.fr/>>.

⁶ <<http://libris.kb.se>>.

⁷ <<http://www.culturegraph.org>>.

⁸ <<http://www.hbz-nrw.de/dokumentencenter/veroeffentlichungen/open-data-leitfaden.pdf>>.

⁹ <<http://www.minervaeurope.org/>>.

¹⁰ <<http://www.michael-culture.org/>>.

¹¹ <<http://www.athenaeurope.org/>>.

¹² <<http://www.linkedheritage.org/>>.

¹³ <<http://thedatahub.org/>>.

¹⁴ <<http://dati.piemonte.it>>.

¹⁵ <<http://opendata.comune.fi.it/>>.

MARIA CASSELLA

Sistema bibliotecario di ateneo
Università di Torino
maria.cassella@unito.it

MADDALENA MORANDO

Sistema bibliotecario
Politecnico di Torino
maddalena.morando@polito.it

DOI: 10.3302/0392-8586-201208-065-1