

Iryna Solodovnik

## Comunicazione e ricerca semantica di contenuti informativi: tra Metadati, Linked open Data e Ontologie

*Ideas worth spreading*

Tim Berns-Lee

<http://www.linkedopendata.it/semantic-web>

### Abstract

*Metadata, Linked (open) Data and ontologies are a part of the complex galaxy of communication languages for web. Being supported by modern information technologies, these languages aimed at reassuring greater visibility, dissemination and sharing of content and, consequently, optimization and rationalization of services such as, for example, digital libraries and archives. Metadata, ontologies and Linked (open) Data, combined together, represent a fundamental starting point for determining the new search methods of the semantic web. Reviewing some existing experiences and initiatives of 'new performance' of data in the digital environment, the paper tries to lead the reader to discover a new way of knowledge creation that can be operated at different levels in every field of human action and, especially, in the scientific and socio-cultural fields.*

### 1. Metadati nel mondo dell'informazione

Nel mondo dell'informazione, i metadati rappresentano la base informativa di 'secondo livello', che descrive, struttura e gestisce i dati primari o le informazioni 'di primo livello' su cui vengono appoggiate le risorse informative. Per queste ultime, i metadati identificano alcune delle loro proprietà, assegnando ad esse specifici valori. L'esempio comune per avvicinare il lettore alla concezione dei metadati viene fornito dalla scheda del catalogo di una biblioteca, i cui dati - *autore, titolo, luogo, data, categoria di classificazione* ecc. - rappresentano i metadati veri e propri, necessari per organizzare, gestire e reperire gli oggetti informativi e/o le collezioni di essi. Attualmente, i metadati costituiscono dei tasselli fondamentali del puzzle del nuovo web, avendo un ruolo chiave nell'indicizzazione e nell'identificazione, nella classificazione e nella catalogazione, nell'autenticazione e nella conservazione, nell'integrità e nella gestione dei diritti, nonché nella distribuzione, nella ricerca e nel recupero delle risorse digitali per cui, scivolando dai metadati alla filosofia, possiamo anche concepirli come gli attori principali di tutta la comunicazione insita delle risorse informative, al fine di costruirne un modello semplificato, funzionale agli scopi informativi dell'utente e facilmente condivisibile con altri. In accordo di quanto sopra, i metadati, ciascuno dei quali potrebbe essere utile in certi contesti, costituiscono il valore aggiunto delle risorse informative, orientate all'erogazione di contenuti e non solo alla mera produzione di dati.

I metadati, siano essi descrittivi (MARC, Dublin Core, PURL, HANDLE, PICO AP ecc.), gestionali - amministrativi (MAG, DOI, CEDARS, METS ecc.) o strutturali (SGML, XML, EAD, MOA2 ecc.), alla fine, sono accomunati da un unico obiettivo multifunzionale, come quello di contribuire a una gestione più chiara e modulare di oggetti/collezioni digitali. Attualmente, uno tra i modelli di metadati più diffusi e condivisi a livello internazionale, è il set di metadati Dublin Core (DC) [26;50; 51], il "formato di metadati di

tipo descrittivo, che fino ad oggi ha supplito anche ad altre funzioni, ora riconducibili sotto differenti tipologie di metadati, più specifici e meglio adeguati a scopi non puramente descrittivi” [11]. Il continuo aggiornamento del modello DC e l’incessante attenzione alla sua “qualità, il confronto sia con il mondo dell’università (biblioteconomia, informatica, etc.), sia con quello della professione (biblioteche, archivi, musei, ecc.) sono garantiti dalla Dublin Core Metadata Initiative (DCMI), un’organizzazione non profit con sede a Singapore” [17].

Nel febbraio 2009, il DC è stato approvato formalmente come lo standard ISO 15836 [50], che ha un ruolo fondamentale nella descrizione di risorse digitali di diverso tipo e nella realizzazione di sistemi di reperimento delle informazioni più efficienti sul web. Il successo di questo standard è dovuto ad una serie di fattori, il primo dei quali consiste nel fatto che le stringhe di testo, strutturate secondo lo schema DC (quindici elementi nella loro forma non qualificata, tra l’altro facoltativi e ripetibili) e associate ai documenti digitali, garantiscono un livello minimo di auto-catalogazione e indicizzazione nel web. I metadati DC si applicano ai contenuti digitali come etichette <tag>, che descrivono le caratteristiche principali di dati strutturati (HTML, XML), consentendo una loro organizzazione più efficiente un recupero più agevole e anche la loro conservazione a lungo termine.

L’adozione del set di metadati DC da parte di diversi gruppi di operatori socio - culturali (come biblioteche, archivi, musei, agenzie governative ecc.) che lavorano in collaborazione per stabilire delle equivalenze tra i descrittori DC, consentirà una più puntuale identificazione dei contenuti digitali eterogenei in settori trasversali [63], faciliterà convergenza [7], comunicazione e interoperabilità (sintattica, semantica) [33;36] tra vari sistemi e permetterà un uso professionale del DC nell’ambito degli strumenti di mediazione propri delle diverse discipline (interoperabilità multidisciplinare). Un ulteriore motivo di successo è dovuto alla flessibilità del modello , esso infatti è capace di arricchire la propria semantica di base attraverso il meccanismo di qualificazione [17], per cui viene ampiamente utilizzato dalle risorse che richiedono descrizioni complesse (e in diverse lingue), ponendosi anche come una sorta di linguaggio descrittivo intermediario che interviene nella mappatura tra diversi modelli di metadati e, di conseguenza, nei raffinamenti dei meccanismi della ricerca.

La *mappatura* fra i metadati di più ampia diffusione, condotta in parallelo, sta permettendo di affrontare in parte la delicata questione delle equivalenze semantiche. L’analisi delle applicazioni sulle quali è stato condotto il confronto ha fatto emergere la questione della diversa granularità secondo cui vengono trattati i documenti nelle risorse informative, cioè la coesistenza all’interno degli archivi di più livelli di descrizione (singoli oggetti, serie, intere raccolte). È un fattore non trascurabile, se circoscritto ad un singolo sistema o a sistemi omogenei dal punto di vista disciplinare, ma che diventa di estrema criticità nel caso di sistemi eterogenei [...] richiede un costante monitoraggio degli standard e delle applicazioni [36].

Tra gli altri standard di metadati che assumono la responsabilità di certificare i contenuti informativi, di rappresentarli per la ricerca, di conservarli e di garantirne la preservazione a lungo termine nei diversi contesti, possiamo citare anche LOM (Learning Object Metadata), Moving Picture Experts Group (MPEG-21), Text Encoding Initiative (TEI), Music Encoding Initiative (MEI), Synchronized Multimedia Integration Language (SMIL), Open Digital Rights Language (ODRL), Open Language Archives Community

(OLAC) basato sul Dublin Core qualificato, PREMIS (metadati per la conservazione a lungo termine) e altri ancora, ognuno proponente di soluzioni specifiche per diverse esigenze di rappresentazione e supporto dell'informazione. L'uso degli standard permette di creare dai metadati gli accessi di qualità, aventi alto grado di precisione ed esattezza nel veicolare e custodire i contenuti delle risorse informative. La qualità di metadati [17] può essere migliorata attraverso un processo di normalizzazione che prevede l'utilizzo di terminologie, vocabolari controllati (thesauri), soggetti, liste di intestazioni per soggetto, classificazioni, informazioni sui diritti ecc. Comunque, in tutti i casi in cui è logico farlo, è più che ragionevole cercare di utilizzare sistemi di metadati standardizzati, siano essi comuni o specifici.

Ricordiamo, inoltre, che i metadati possono essere categorizzati come: *oggettivi*, ricavabili direttamente dall'analisi dell'informazione primaria, in maniera indipendente dalle valutazioni di chi classifica; *soggettivi*, in cui la valutazione soggettiva è preponderante; *primari*, forniti dal creatore della risorsa informativa; *secondari*, forniti dagli utenti; *terziari*, forniti da servizi o organizzazioni indipendenti; *embedded*, inclusi nella risorsa informativa ed *external*, esterni alla risorsa informativa. Secondo le necessità della descrizione, strutturazione o gestione di risorse informative si potrebbero scegliere metadati diversi, adottare prospettive diverse, e proporre diverse distinzioni e classificazioni [13].

La scelta dei formati di metadati è molto importante nell'implementazione e nel supporto degli archivi aperti (disciplinari e istituzionali), creati sull'architettura *Open Archives Initiative*, OAI (<<http://www.openarchives.org>>) che supporta insieme paralleli di metadati, disseminati dagli archivi distribuiti [10; 17]. Tali archivi sono i luoghi virtuali di 'archiviazione' e scambio dei metadati per promuovere e supportare la gestione degli elaborati (articoli, atti di convegni, libri e parti di essi, tesi di dottorato e progetti di ricerca ecc.) prodotti nell'ambito accademico, e rappresentano una realtà molto fertile in quanto fortemente innovativa di divulgazione del sapere scientifico attraverso i canali di comunicazione digitale ad accesso aperto (OA) [14;15]. In questo contesto, il supporto alla gestione dei contenuti viene spesso fornito dal già citato formato DC che, dentro l'architettura OAI, si abbina facilmente con altri schemi di metadati, aumentando granularità e raffinamento delle loro strutture. In questo scenario, possiamo citare i seguenti formati di metadati: "OAI DC Dublin Core codificato in XML, OAI RFC1807 RFC1807 codificato in XML, ArXiv (Old e Test) codificati in XLM, AMF Test-bed for Academic Metadata Format" [11], ReDIF [21].

Da questa e altre simili linee di sviluppo di nuovi strumenti di comunicazione delle risorse informative, nascono i nuovi modi di trasmissione dei contenuti eterogenei attraverso insieme paralleli di metadati. Questa trasmissione avviene dentro un'architettura client-server e, in particolare, dentro la struttura comunicativa di data e server provider (su cui si basano gli archivi aperti). Data provider rappresentano gli archivi distribuiti con i metadati e file sorgenti, mentre service provider eseguono la raccolta dei metadati da data provider, salvandoli in una collezione centralizzata, dove avverrà l'indicizzazione e quindi la ricerca di oggetti informativi attraverso i metadati, senza dover interrogare le collezioni periferiche e i file sorgenti. Una volta individuato il file di interesse, il service provider eseguirà una richiesta al data provider e si riceverà il

file ricercato. Le funzioni di service provider permettono la realizzazione dei servizi a valore aggiunto come la ricerca unificata su più archivi, sistemi più complessi come applicativi per l'analisi citazionale o per il reference linking, interfacce di ricerca evolute, tali da unificare il sistema di accesso a materiali eterogenei e multimediali, consentendo in una sola azione di ricerca il reperimento di documenti appartenenti a diverse tipologie di standard di registrazione e mezzi espressivi di creazione. L'interazione (che è alla base dell'interoperabilità) tra i vari provider per l'esposizione e la raccolta di metadati può avvenire tramite il protocollo *Open Archives Initiative Protocol for Metadata Harvesting* (OAI-PMH) [75] (Figura 1), l'OpenURL, i protocolli di interoperabilità come ISO-ILL, LDAP, SRU/SRW, NISO, Z39.50, gli standard tecnologici come HTTP, WebServices, XML, PKI.

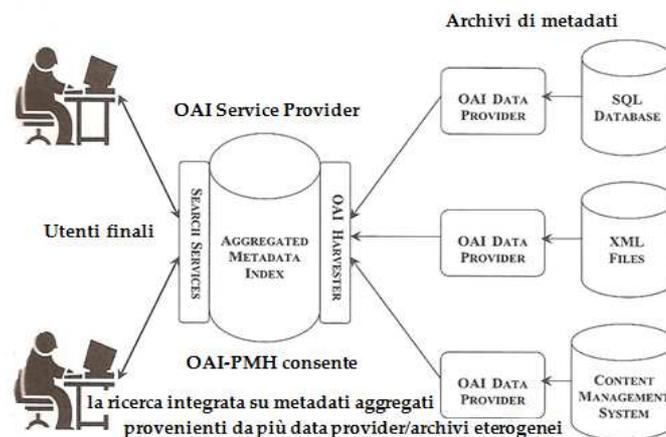


Figura 1. Ricerca integrata su metadati aggregati attraverso l'OAI-PMH

Il protocollo OAI-PMH, ad esempio, permette di implementare i servizi a valore aggiunto su aggregazioni di metadati raccolti da ambienti di dati eterogenei e architetture di storage multiple, anche se la natura e i dettagli di tali servizi non sono inclusi tra le transazioni di client server OAI-PMH definite dal medesimo protocollo. Obiettivo dell'OAI-PMH è, dunque, facilitare l'interazione fra sistemi differenti, nonché lo scambio e il riutilizzo delle informazioni anche fra sistemi informativi non omogenei (sia per software che per hardware).

Lo scenario di semantic web amplifica la concezione di metadati per la ricerca in rete, spaziando oltre l'interoperabilità del primo livello, quella sintattica, e rinviando a quella semantica che permette la cross-searching dei contenuti informativi. In tale contesto, i metadati, pur restando sempre gli attori principali della rappresentazione degli oggetti digitali, si offrono anche come gli accessi semantici alla rete di ontologia/e dell'informazione che permette di oscillare tra diversi piani di metadati e veicolare diverse 'espressioni' sui diversi piani di 'contenuto'.

## 2. Tra meta(dati), ontologie e Linked (open) Data

Il Workshop WWW2011, svoltosi di recente in India e dedicato al tema "*Linked Data on the Web*" [67] ha enfatizzato che il web, essendo uno spazio di informazione globale, deve consistere non solo di documenti linkati, ma anche di meta(dati) linkati. L'iniziativa *Linked*

Data, sostenuta dall'ideatore del *World Wide Web*, Tim Berners-Lee, ha l'obiettivo di promuovere la creazione di nuovi dataset e collegarli (linkarli) direttamente ai dataset esistenti, riducendo la duplicazione dei dati e, soprattutto, tenendo i dati sempre freschi e aggiornati [87]. *Linked Data* si basano sulle tecnologie standardizzate del web, come HTTP (che certificano automaticamente la provenienza di diverse versioni di record di metadati e di *referenced multimedia files*), Uniform Resource Identifier (URIs) e il modello Resource Description Framework (RDF) <[www.w3.org/RDF/](http://www.w3.org/RDF/)> (l'open web standard che può essere liberamente utilizzato da tutti); ovvero sui formalismi per la rappresentazione standardizzata dei dati e dei meccanismi condivisi per l'accesso e l'interrogazione di tali dati. *Linked Data* si sposano con la visione del movimento *Open Data* [76] (promosso sotto le licenze Creative Commons <<http://creativecommons.org/>> o Talis, <<http://www.talis.com/tdn/tcl>>) che ha l'obiettivo di abbattere le barriere sociali, culturali, legali ed economiche che ostacolano la libera condivisione dei dati tra persone e agenti software.

Vogliamo citare adesso il progetto di W3C Linking Open Data [68] (*Open Definition*, <<http://opendefinition.org>>) che fornisce tutte le informazioni necessarie per estendere il web con *Open Data Commons* (<<http://www.opendatacommons.org/>>). L'estensione di dataset avviene attraverso la pubblicazione di collegamenti (links) RDF, uno degli strumenti essenziali per esprimere, in maniera formalmente rigorosa, e per condividere metadati strutturati. Nell'ottobre 2007, i dataset sul web consistevano di oltre due miliardi di triple RDF (con i loro statement di: soggetto, predicato, oggetto), unite da oltre di due milioni di collegamenti RDF. Nel settembre 2010, questi dati sono cresciuti a 25 miliardi di triple RDF, interlinkate da circa 395 milioni di link RDF. I collegamenti tra diversi dataset vengono graficamente rappresentati nella forma di una grande 'nuvola' chiamata "LOD cloud diagram", in cui vi è una visualizzazione interattiva dei gruppi di dataset interoperabili (Figura 2).

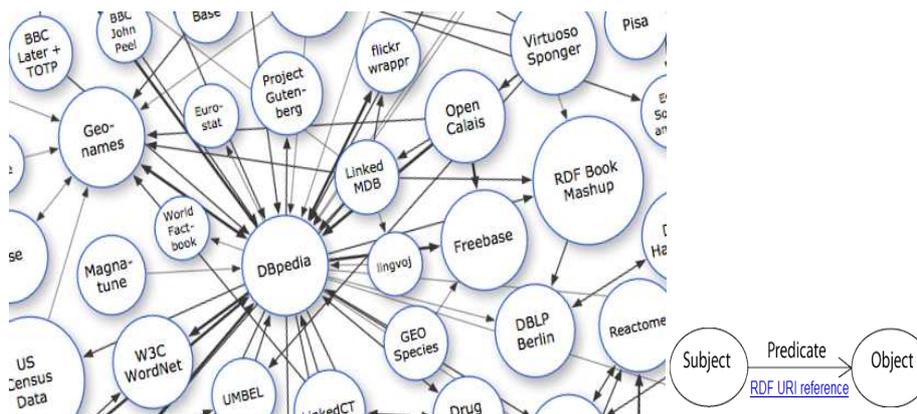


Figura 2. Linked Data. Parte di *Linking Open Data cloud diagram*

Fonte: Richard Cyganiak and Anja Jentzsch, <<http://lod-cloud.net/>>; <<http://linkeddata.org/>>

Il diagramma *LOD cloud* dimostra le categorie di dataset, che convergono nello CKAN <<http://ckan.net>>, un catalogo di dataset *Open Data* e *Linked Open Data* gestito dalla *Open Knowledge Foundation*. Nel progetto *Linking Open Data* sono state utilizzate 203 classi di

metadati basate su più di 25 miliardi di triple RDF, collegate fra loro da circa 395 milioni link RDF che consentono di navigare tra i dati utilizzando un browser web semantico. Infatti, cliccando sulle classi concettuali dei dataset rappresentate dai cerchi (nell'originale raffigurazione grafica della 'nuvola'), si arriva alla scoperta di nuovi contenuti pubblicati in formato *Linked Data* da parte della comunità che sviluppa il progetto Linking Open Data sulla base dei registri di sistemi aperti di dati e pacchetti di contenuti. L'idea di base del *LOD cloud diagram* è abbastanza ambiziosa, in quanto mira a creare un unico grande spazio di dati collegati tra loro e accessibili per diversi utenti e applicazioni software per possano scoprire nuove informazioni, creare nuove conoscenze e, nel caso di Open Data, collezionarle e ripubblicarle liberamente.

L'immagine di LOD cloud, presentata nella Figura 2, è solo uno dei possibili scenari in cui LOD possono favorire l'interoperabilità tra dataset. Le possibilità sono infinite, se pensiamo alla immensa quantità di LOD già presenti nel Web come, ad esempio, DBPedia.org, Wikipedia e WikiGuida [94], Geonames e MusicBrainz [73], WordNet [95], la bibliografia DBLP [47]. Essendo di particolare interesse, segnaliamo, inoltre, l'UMBEL Web Services (<<http://umbel.zitgist.com/>>), Virtuoso Universal Server (<<http://virtuoso.openlinksw.com/>>), Linked Open Data Around-The-Clock (<<http://latc-project.eu/>>), le piattaforme create su LOD che pubblicano e distribuiscono i dati sul web, usando il modello RDF, gli URIs e il protocollo Http. Simili iniziative pratiche permettono l'ampliamento e il riuso di domini di conoscenza, facilitando i cambiamenti sulle assunzioni dei domini nonché la comprensione e l'aggiornamento dei dati esistenti.

### 2.1. Concepire una rete di ontologie

La creazione di LOD richiama il concetto di rete di ontologie, che permette di muoversi da un dataset concettuale all'altro, scoprendo i nuovi universi di contenuti collegati a dataset di partenza. Supponiamo, per esempio, di avere, da una parte, un Archivio Istituzionale che pubblica i dati aperti relativi ai lavori scientifici del proprio staff accademico; dall'altra, la piattaforma con i curriculum dei medesimi autori e, in terzo luogo, l'anagrafe dei prodotti della ricerca di Ateneo che valuta l'efficienza e l'efficacia delle attività scientifica dell'istituzione. Collegare (linkare) questi dataset potrebbe essere di grande utilità, sia ai produttori dei lavori scientifici che al Nucleo di valutazione di Ateneo e agli utenti. Si potrebbe ottenere un dataset comune facilmente navigabile, da cui si potrebbero ottenere interessanti "viste". Ad esempio, tale dataset comune potrebbe offrire un servizio personalizzato per gli autori depositanti il materiale nell'archivio (permettendo l'aggiornamento automatico dei loro curriculum in base alle pubblicazioni e altre attività di ricerca inserite nell'archivio); contribuire ad un miglior censimento delle competenze di tali autori da parte del Nucleo, nonché il censimento dei nuovi risultati scientifici ottenuti in base all'aggiornamento dei curriculum professionali (partecipazione ai seminari, convegni, conferenze; specializzazioni professionali ecc.). Inoltre si potrebbe creare servizio di valore aggiunto nell'Archivio istituzionale stesso, in cui gli utenti potranno consultare in modo aperto non solo le pubblicazioni dei vari autori, ma anche visionare le schede complete dei loro curriculum, evidenziando i loro percorsi professionali, interessi di ricerca ecc.

Un altro esempio della creazione di una rete di ontologie con i (meta)dati linkati potrebbe essere quello dell'organizzazione dei libri di una libreria in base al contenuto informativo del testo, il titolo, l'autore, gli argomenti trattati ecc. In questo contesto potrebbe essere rilevante anche tener traccia delle persone alle quali prestiamo i libri, nonché, in caso la libreria sia di un grande scrittore, potrebbe essere importante sapere se e quali libri hanno delle sottolineature o delle note manoscritte, che potrebbero rappresentare un interesse per lo studio. In ciascuno di questi casi, all'interno dell'universo informativo di riferimento, dobbiamo selezionare le caratteristiche più rilevanti per la costruzione di un'ontologia di dati, e di utilizzarle per organizzare i dataset in base a uno fra i tanti modelli possibili.

La creazione di collegamenti tra diversi dataset non è un processo banale, perché si deve attentamente calcolare un riutilizzo organico dei dati, condivisi all'interno di comunità di utenti. La procedura di creazione di link può richiedere o un lavoro a mano o può essere svolta attraverso algoritmi ad hoc. Saranno necessari gli strumenti per esprimere in maniera esplicita i vincoli posti sui valori e sulle proprietà di dati (qui ritorniamo al concetto di metadati), per introdurre regole e magari, perfino, assiomatizzazioni per permettere di dedurre l'assegnazione di particolari valori a determinate proprietà di dati, conseguentemente alle scelte fatte in precedenza.

Il lavoro di costruzione di metadati di qualità 'solidi', anche se è complesso e faticoso, garantisce la compatibilità fra i nostri dataset e quelli di altri, consente la gestione automatica dei nostri dati e metadati, rispondendo, dunque, all'esigenza di interoperabilità dei sistemi di gestione dell'informazione. L'applicazione dei linguaggi descrittivi standardizzati aiutano a migliorare e a validare il nostro sistema con diversi dataset e, spesso, a capire meglio la natura dei dati primari del nostro dominio, anche se sono limitati alla specifica classe di metadati. D'altra parte, la descrizione rigorosa dell'informazione ha i suoi rischi, che è la conseguenza di scelte di descrizioni sbagliate, o di applicazioni sbagliate di sistemi di classificazione ecc., anche se possono sembrare validissime.

Rendere espliciti i link tra diversi dataset, e specialmente quelli a livello semantico, richiede l'analisi attenta e la definizione rigorosa di tutte le necessarie caratteristiche 'astratte' del sistema di dati e metadati. Questo processo corrisponde all'individuazione per diversi soggetti (siano essi persone o agenti software) di una specificazione formale, esplicita e condivisibile (o un'ontologia) di quali siano i tipi di dataset che compongono il dominio di riferimento, di quali siano le proprietà e relazioni che possono essere stabilite fra di essi, e di quali siano le regole che ne governano il funzionamento. Per esprimere in modo formale e rigoroso tutte le informazioni sul sistema di dati e di metadati nell'ambiente digitale, sono stati elaborati linguaggi specifici, tra cui il più noto è il linguaggio OWL, *Web Ontology Language*, che è l'estensione semantica del vocabolario RDF [83]. Definire i dati in RDF presuppone la scelta di uno o più vocabolari, o ontologie, cioè meccanismi di specificazione di diverse concettualizzazioni di oggetti, concetti e altre entità, e di relazioni che collegano tali entità. Una buona pratica è quella di scegliere Ontologie già esistenti e ampiamente utilizzate per altri dataset di domini simili, in modo da rendere intelligibili e interpretabili i dati RDF di diversi dataset relazionali. Questo approccio è l'elemento chiave per la creazione di valore aggiunto sui dati, cioè delle

applicazioni aperti e distribuiti per offrire contenuti e servizi altamente interoperabili. Per generare e fare le eventuali trasformazioni di link tra dataset in modo automatico, citiamo il famoso *tool Google-Refine* (<<http://code.google.com/p/google-refine/>>) che permette di importare i dati del dominio in database e, successivamente, di estrarre e di raffinare i metadati necessari per creare i link per dataset.

### 3. Da LOD a Content Management Systems

Come abbiamo già visto, per descrivere il modello di conoscenza di un certo dominio, occorre accordare sulla scelta del vocabolario che verrà utilizzato per rendere esplicita la concettualizzazione di tale conoscenza. Successivamente, bisogna definire il profilo di metadati e delle loro proprietà basandosi sui requisiti del modello astratto dei metadati (come esempio, segnaliamo *DCMI Abstract Model* [1]). In seguito, attraverso la sintassi e i vocabolari del modello di metadati scelto, bisogna rendere esplicite (popolare) le classi e le proprietà di tale modello, normalizzarle ed, in un secondo tempo, collegarle (linkarle) con dei modelli di domini di comunità già esistenti, rilevanti per l'arricchimento delle informazioni nel campo. Così, per potenziare l'espressività delle entità bibliografiche, è di beneficio collegare il loro modello di metadati descrittivi con delle classi di *Functional Requirements for Bibliographic Records* (FRBR) [56], importante per migliorare gli aspetti sotto i quali può essere concepita una produzione intellettuale (tali aspetti sono: opera, espressione, manifestazione, unità; persona, ente collettivo; concetti, oggetti, eventi, luoghi come soggetti dell'opera).

Nel passo seguente, bisogna definire e descrivere le cardinalità (obbligatorie o facoltative) tra le diversi classi di metadati (dataset) usando il modello RDF (Figura 3).

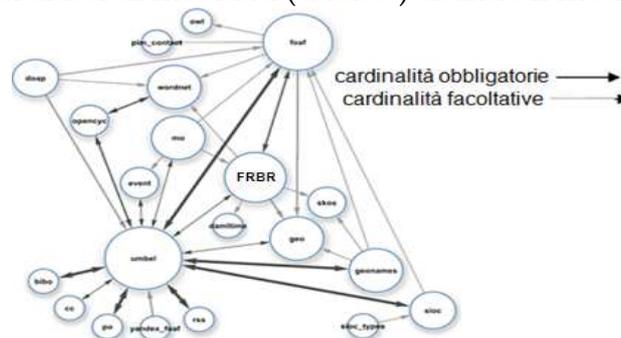


Figura 3. Linking-Open-Data-class-diagram

Fonte: Bergman, M.K., A New Constellation in the Linking Open Data (LOD) Sky, <<http://www.mkbergman.com/457/a-new-constellation-in-the-linking-open-data-lod-sky/>>

È fondamentale che l'espressione RDF conservi la semantica del modello astratto scelto. Inoltre, tutte le inferenze valide, che possono essere effettuate utilizzando la semantica RDF, devono essere efficaci (ragionevoli e relativamente semplici) quando interpretate in termini del modello astratto, per cui la formulazione delle interferenze avrebbe bisogno di definire una mappatura inversa da RDF al modello astratto dei metadati. La rappresentazione riassuntiva della creazione dei record di metadati "linked - data - compatible" riportiamo nella Figura 4.

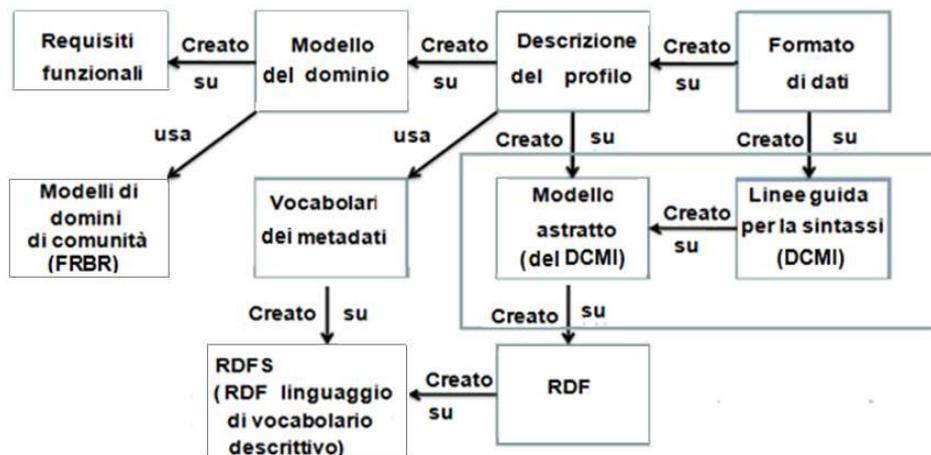


Figura 4. Progettazione di record dei metadati “linked – data - compatible”

Fonte: Nilsson, M. et al. The Singapore Framework for Dublin Core Application Profiles, <<http://dublincore.org/documents/singapore-framework/>>

La creazione dei metadati “linked – data - compatible” attraverso le triple RDF favorisce la mappatura tra gli schemi di metadati riferiti a diversi oggetti informativi, e potenzia le risposte del sistema alle query nella ricerca semantica delle informazioni distribuite ‘apertamente’ in tutta la rete. Tale concezione potrebbe, naturalmente, avvantaggiare i promotori del movimento Open Access (OA) nel contribuire con le informazioni aperte alla globale ‘LOD cloud’ della conoscenza, ma scoraggiare gli utenti delle reti sociali chiusi, che non vogliono rendere accessibili le informazioni contenenti i dati sensibili.

L’attualità della problematica di protezione dei dati è chiamata in causa anche per il fatto che nonostante lo specifico vocabolario di rete, chiamato *Web Access Control (WAC)*, limita l’accesso ai documenti RDF per determinati tipi di utenti, esso non fornisce le adeguate misure di ‘privacy’ che specificano le restrizioni complesse per l’accesso ai dati. Per colmare tali lacune, i recenti studi sul campo stanno proponendo le soluzioni adeguate, una tra le quali è il vocabolario *Privacy Preference Ontology (PPO)* [31], sviluppato dall’Istituto del Digital Enterprise Research dell’Università Nazionale di Irlanda. Tale vocabolario semantico, basato sul WAC, permette ai creatori delle informazioni di generare le adeguate raffinate misure di ‘privacy’ per i loro dati. Queste misure sono designate per restringere ogni risorsa informativa a un certo numero di attributi, i campi di cui devono essere pienamente soddisfatti dai requisiti forniti dagli utilizzatori delle risorse.

Il livello standard di dati o l’accesso e la sicurezza di web server possono essere rafforzati dalla scelta del sistema di hosting, nonché espressi attraverso le ontologie RDF, per cui, se un certo collegamento puntasse ad un data object (proveniente da una fonte con accesso limitato o controllato), le risposte del sistema non verranno visualizzate - per quegli utenti che hanno restrizioni d’accesso. In ogni caso, la protezione dei dati riservati è un tema ricorrente e irrisolvibile a livello generale. Tuttavia, l’argomento è in corso di studio da diverse comunità Open Data e presto ci saranno altre nuove soluzioni in merito.

Un’altra esperienza, volta a potenziare la distribuzione e conservazione a lungo termine del contenuto del patrimonio culturale attraverso LOD, è stata svolta in Belgio, nel quadro del progetto *Archipel* (<<http://www.archipelproject.be>>) dell’Università di Ghent

[9]. Qui è stata sviluppata una piattaforma digitale con tre linee di intervento, come: la raccolta (harvesting) di (meta)dati da diverse istituzioni (biblioteche, istituzioni archivistiche, musei, organizzazioni broadcast); la conservazione dei (meta)data a lungo termine e la distribuzione di quest'ultimi come record DC *Linked Open Data*. Tale piattaforma riesce a produrre le informazioni sulle diverse versioni (prodotte nel corso del tempo) dei dati raccolti, attraverso, ad esempio, la mappatura dei metadati o la transcodifica dei file multimediali, ma fornisce anche le informazioni relative alla provenienza di tutte le versioni disponibili di una risorsa informativa. Per pubblicare queste informazioni come LOD, si è proceduto con l'estensione del server *Linked Open Data* per mezzo di framework *Memento* (<<http://www.mementoweb.org/>>) [34], che, a sua volta, è stato anche rimodellato per poter pubblicare i dati su provenienza di diverse 'versioni negozianti' usando Http, per la scoperta automatica dei relativi dati sulla provenienza. Le informazioni sulla provenienza dei contenuti sono state modellate come LOD usando l'implementazione semantica dell'ontologia PREMIS OWL [81] che favorisce l'interoperabilità delle informazioni nel contesto di conservazione a lungo termine.

Nel campo di indagini su LOD, vengono, inoltre, studiate le questioni sul potenziamento dei meccanismi di ricerca delle informazioni attraverso LOD già esistenti, ricavati dal web [18]. Si propone di scoprire e recuperare i dati RDF provenienti dalla rete, ritenuti rilevanti per la ricerca federata di dati (*cross-searching*) e di sviluppare le strutture per la loro memorizzazione attraverso gli indici basati su hashing. Tali strutture possono essere proposte per l'implementazione dei sistemi con l'esecuzione delle query basate sulle triple RDF, potenzialmente efficaci per le diverse applicazioni locali.

L'importanza di LOD viene rivelata anche nel campo di annotazione automatica dei testi [30] attraverso le risorse come *WordNet*, *OpenCyc* and *DBpedia*, strutturati sui commenti RDFs per fornire un contesto aggiuntivo alle entità e alle unità informative rappresentate dai metadati, e per sviluppare i meccanismi di tagging semantico e di disambiguazione lessicale di elementi testuali in rete, in relazione al contesto in cui si trovano. I collegamenti (commenti) RDFs, sviluppati per le risorse di conoscenza, come dizionari, ontologie e thesauri, prendono in considerazione le relazioni che esistono tra gli elementi delle strutture di tali risorse in base alle loro caratteristiche comuni. Lo sviluppo dell'annotazione semantica attraverso LOD ha lo scopo di offrire una migliore performance di dataset semantici già esistenti (come ad esempio delle ontologie *OpenCyc* e *DBpedia*) e di trasferire tale performance ad altri sistemi semantici in rete che recentemente sono stati impiegati per la costruzione di linked - data browser, motori di ricerca semantica e altre applicazioni di domini specifici.

LOD contribuiscono, dunque, ad una migliore definizione condivisa di diverse strutture di *Knowledge Representation* (come metadata, ontologie, thesauri, terminologie, tassonomie), rendendo così più 'aperti' e 'liberi' i dati e gli archivi di dati. LOD promuovono la costituzione di sistemi informativi digitali indipendenti e, nel frattempo, altamente interoperabili con gli altri sistemi basati sul modello W3C, che attiva la diffusione di *Open Metadata* e degli Standard di interoperabilità dei contenuti con la conseguente creazione di *Content Management Systems* (CMS) (Figura 5).

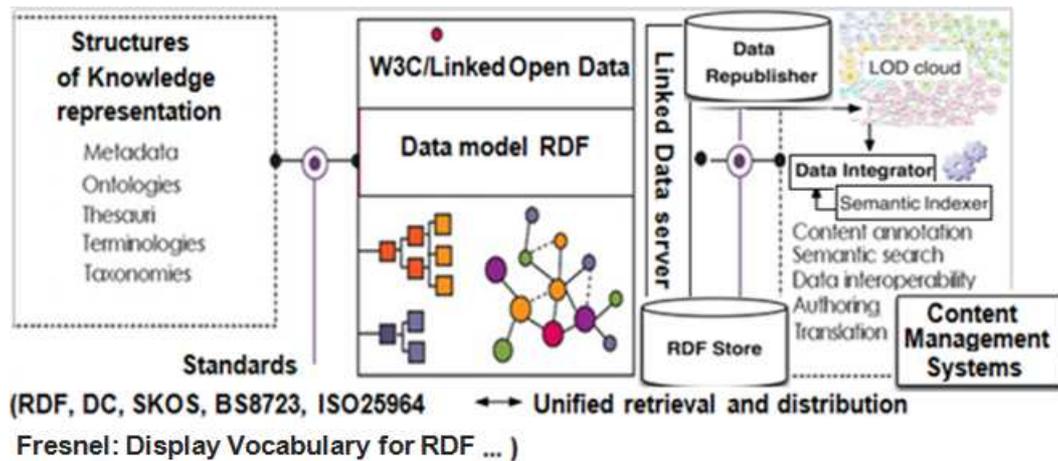


Figura 5. LOD per la creazione di CMS

Il modello W3C definisce le regole per la pubblicazione di dati sul Web in modo che essi possano essere facilmente individuati, incrociati e manipolati dal sistema di gestione dell'informazione in rete globale e distribuito, basato sugli standard di metadati e meccanismi condivisi per l'accesso, interrogazione, condivisione, annotazione e ricerca semantica dei contenuti. Le applicazioni di LOD devono essere presentate e promosse in pratica, con particolare attenzione al loro sfruttamento nelle relative interfacce di ricerca.

### 3.1. LOD: alcune applicazioni pratiche

Il *Semantic Computing Research Group SeCo* <<http://www.seco.tkk.fi/>> dell'Università di Helsinki, nel quadro del progetto FinnONTO sta lavorando sull'implementazione di una serie di applicazioni basate su LOD per adoperarle nella ricerca semantica dentro i sistemi informativi digitali come *MuseumFinland* [71], *CultureSampo* [44], *HealthFinland* [58] e altri ancora, basati sul *DataFinland* [46], un canale di distribuzione semantico dei contenuti e di set di metadati (dataset) dell'*ONKI Metadata Schema Library* <<http://schema.onki.fi/>>, che contiene la documentazione sui sistemi di metadati creati dal progetto FinnONTO [53]. I LOD, sviluppati per l'*ONKI Ontology Library Service* <<http://www.onki.fi/>>, possono essere riutilizzati per diverse viste di simili domini applicativi, essendo aperti come tools di orientamento per favorire la creazione di sistemi informativi semantici basati sugli schemi di metadati pubblicati in RDF.

Nel frame dello stesso progetto, citiamo l'ontologia *Finnish General Upper Ontology* (YSO) [52], basata su più di 20.000 dataset relazionati, ricavati dal *Finnish General thesaurus* ampiamente usato dalla Biblioteca nazionale della Finlandia. Lo scopo principale di tale ontologia è di rappresentare il punto di convergenza per le diverse ontologie di dominio e, in questo modo, di offrire la possibilità di collegare i contenuti di diversi campi di conoscenza, presenti nel web semantico finlandese. La seconda ontologia creata su LOD è *The Actor Ontology* (TOIMO) <<http://www.seco.tkk.fi/ontologies/toimo/>> che descrive e mette in relazione di ruolo diversi tipi di agenti (attori) come persone, istituzioni e organizzazioni, che rappresentano gli elementi centrali di molti sistemi informativi. In pratica, tale ontologia integra gli agenti di ULAN archives, il vocabolario strutturato di

nomi, biografie e altre informazioni [91] con gli agenti di *CultureSampo*, il portale basato sul semantic web 2.0 [2] che promuove il patrimonio culturale della Finlandia. L'ontologia TOIMO ha l'obiettivo di generare nuove informazioni sugli agenti che possono, a loro volta, rappresentare gli strumenti semantici per estendere le conoscenze in rete sugli agenti in questione.

Vogliamo riferire anche altre ontologie sviluppate dentro lo stesso progetto per le applicazioni in *eCulture*. Queste ontologie sono: la *History Ontology* (HISTO) [59], creata per modellare la struttura del dominio culturale attraverso il vocabolario con le equivalenze tra i termini della lingua finlandese arcaica e quella moderna; l'*ICONclass Ontology* [60] rappresentante la versione ontologica del sistema di classificazione *ICONclass*, sviluppato per categorizzare i domini dell'arte e dell'iconografia nei Paesi Bassi. Questa ontologia modella il sistema di classificazione seguendo i requisiti di *W3C SKOS Core Vocabulary Specification* [85], che supporta l'uso di sistemi di organizzazione della conoscenza (KOS) come thesauri, schemi di classificazione, liste di intestazione di soggetti e tassonomie entro il semantic web; l'*Ontology for museum domain* (MAO) [73], sviluppata per fornire e relazionare le dettagliate e uniformi descrizioni dei dataset degli oggetti museali; la *Finnish Ontology of Photography* VALO [88] che combina, attraverso LOD, i concetti del thesaurus (circa 1.700) per l'indicizzazione semantica delle fotografie prodotte in Finlandia con le categorie della YSO.

Le implicazioni pratiche di LOD citate offrono nuove prospettive di sviluppo dei sistemi informativi basati su applicazioni che ottimizzano la ricerca semantica. Ciò rappresenta un significativo passo in avanti nel semplificare la gestione della conoscenza in rete perché, permettendo un'interazione immediata con diversi dataset attraverso la determinazione del contesto, sia arriva a rappresentare - la "mappa cognitiva" da cui è possibile estrarre e interpretare i significati.

### 3.2. LOD e dati bibliografici: dati aperti, collegati e usabili

Che ruolo possono avere LOD nell'implementazione e nell'aggiornamento delle risorse delle biblioteche digitali? Esiste un modo per linkare diversi dati bibliografici - come nomi e identificatori di *autore/i* e *curatore/i*, *editore*, *data* e *luogo di pubblicazione*, *identificazione dell'opera madre*, *numeri di pagina*, *URI*, ecc. - delle diverse descrizioni bibliografiche? Quali sono le modalità per connettere le descrizioni bibliografiche con altre informazioni, che non rientrano nel concetto di dati bibliografici - come, per esempio, *formato dell'opera*, *identificatori esterni al web* (*ISBN*, *LCCN*, *OCLC* ecc.), *informazioni sul tipo di supporto*, *dichiarazioni sui diritti*, *dati amministrativi* (*ultima modifica*, ecc.), *link rilevanti* e altri dati, - che possono essere prodotti da editori, accademici, comunità on-line di bibliofili ecc.?

Simili ipotesi e domande nel campo sono ancora molte, anche se la comunità internazionale sta lavorando attivamente sullo sviluppo di una serie di iniziative pratiche per promuovere la conoscenza aperta sulle risorse bibliografiche nello spazio aperto del web semantico. Le biblioteche e le istituzioni collegate, per poter interconnettere le loro risorse in modo chiaro, condiviso ed esplicito, già da tempo stanno producendo vocabolari controllati per la descrizione bibliografica (*authority file*: l'archivio di autorità per il

controllo della forma, del nome degli autori personali e collettivi e dei soggetti; terminologie; thesauri; soggettari; classificazioni, ecc.). Inoltre, si sta attivamente discutendo sulle modalità pratiche della pubblicazione aperta dei dati bibliografici affinché essi possano essere linkati e condivisi con altri dataset. Ad esempio, l'*Open Knowledge Foundation Italia* fornisce le informazioni su come pubblicare apertamente i dataset bibliografici sul web e renderli, successivamente, parte del web semantico [87], invitando le istituzioni di adoperare le licenze aperte ed esplicite [42], anche se per i dati bibliografici molte licenze, pur essendo ampiamente riconosciute (come *Creative Commons*; escluso, CC0, GFDL, GPL, BSD) spesso non sono appropriate perché rendono impossibile un'integrazione adeguata dei dataset, nonché ostacolano la loro riproduzione e lo sviluppo dai servizi commerciali che possano aggiungere valore ai dati bibliografici o servire come supporto alla loro preservazione.

Si sta indagando anche sulle possibilità di rendere espliciti in RDF gli standard bibliografici, come *Metadata Object Description Schema* (MODS) (<<http://www.loc.gov/standards/mods/>>) e *Machine-Readable Cataloging Standards* (MARC) (<<http://www.loc.gov/marc/>>). Segnaliamo, a proposito, l'esperienza del progetto *MIT Libraries Cataloging OASIS* (<<http://libstaff.mit.edu/colserv/cat/>>), che ha sviluppato l'*Utility Tool* [92] (disponibile gratuitamente in rete) per convertire MARC e/o MODS in RDF. Un'altra importante iniziativa che fornisce i concetti generali e le proprietà utili per pubblicare le citazioni descrittive e i riferimenti bibliografici (libri, articoli, ecc.) nel web semantico è la *Bibliographic Ontology Specification* [45], sviluppata sotto la licenza CCL e appoggiata sulle tecnologie quali RDF. I concetti, forniti dalle specifiche, vengono racchiusi nella documentazione sull'Ontologia Bibliografica che offre gli esempi di pubblicazione dei documenti in RDF e fornisce le informazioni sull'evoluzione e l'estensione dell'Ontologia Bibliografica e dei relativi Standard a supporto. L'Ontologia Bibliografica può essere usata come un'ontologia citazionale, come un'ontologia di classificazione documentaria o, semplicemente, come una modalità per descrivere ogni tipologia del documento in RDF. Questa ontologia è stata 'accolta' da molti esistenti formati di metadati descrittivi, per cui può essere usata come una base comune per la conversione di molte fonti bibliografiche di dati.

Vogliamo riferire anche il progetto *Bibliografica* (<<http://bibliografica.org/>>) che rappresenta un catalogo aperto (basato su LOD) delle opere culturali, con, attualmente, di 3017569 lavori depositati in un apposito database. Proseguendo con la citazione delle iniziative di valorizzazione e promozione delle risorse bibliografiche nel web semantico, segnaliamo anche la *MarcOnt* (<<http://www.marcont.org/>>), che rappresenta un'ontologia di integrazione (attraverso RDF Translator) per i formati descrittivi bibliografici come MARC21, BibTeX and Dublin Core. Citiamo anche la *MODS Ontology* (Figura 5) [70], la motivazione primaria per lo sviluppo di cui è quella di identificare una strategia migliore per la migrazione dei metadati MARC in MODS e poi in RDF, nella forma più accettabile dagli utilizzatori e produttori di *Linked Data*.

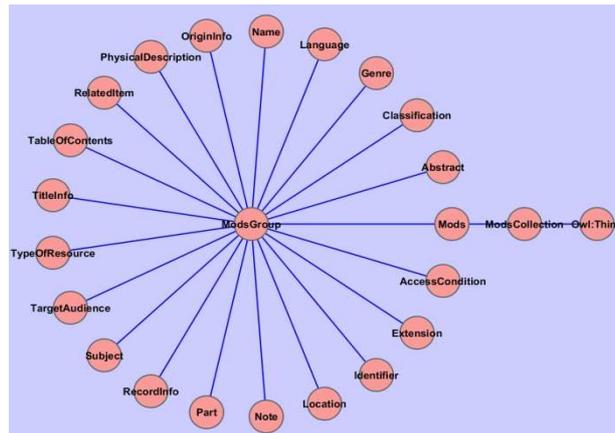


Figura 6. MODS Ontology - I livello

Fonte: "Another Step Toward Lifting Library Metadata into the Cloud", <http://www.chrisfrymann.com/2009/07/22/mods-ontology-2/comment-page-1/#comment-8>

La *MODS Ontology*, attraverso le espressioni ontologiche OWL, è in grado di far valere le relazioni di equivalenze (*owl: sameAs*) ad altre ontologie e altre affermazioni insiemistiche su elementi MODS. Ogni livello del grafo dell'Ontologia MODS, prodotto con l'open source Cytoscape, corrisponde ad un sottoinsieme sempre più grande dell'intero set di dichiarazioni RDF circa l'ontologia MODS.

L'*UCSD Libraries' Digital Library Program* [90] sta facendo diverse indagini sulle prestazioni di supporto di dati bibliografici in alcuni Open Source, per individuare le loro capacità di trasporto e integrazione dei medesimi dati nel web semantico. Particolarmente, indagando i software come DSpace (<http://www.dspace.org/>) e Fedora (<http://www.fedora-commons.org/>), ampiamente conosciuti per l'implementazione di archivi aperti, sono state scoperte dei limiti di stessi software nel supporto dei dati bibliografici per la descrizione bibliografica standardizzata come MARC o MODS e delle soluzioni XML. A questo punto si è deciso di codificare i dati MARC e MODS in RDF, servendosi del tool *ARK: Archival Resource Key* [23], con cui sono stati trasformati in RDF centinaia migliaia di record MARC → MODS, successivamente caricati, per mezzo di 15 milioni triple in *AllegroGraph* RDF [39]. Basandosi sui risultati ottenuti, sono state create anche le query SPARQL capaci di interrogare i grafi RDF [57]. Dentro l'*UCSD Libraries' Digital Library Program* segnaliamo anche un'altra esperienza volta a rendere adeguati i dati bibliografici per il web semantico attraverso linking dello schema MODS XML di *Library of Congress* (<http://www.loc.gov/index.html>) con l'Ontologia formale OWL. L'approccio scelto mira ad arricchire le entità bibliografiche attraverso le relazioni "*owl:sameAs*" in alternanza con i valori "*RDF:label*", rendendo, inoltre, tali dati aperti per eventuali interrogazioni nel web semantico.

Le iniziative dimostrate possono essere estese anche fuori della realtà di biblioteche digitali, cioè ad ogni settore dove vengono prodotti i dati bibliografici, come gli editori, le università, o le comunità che usano strumenti bibliografici condivisi (*social reference management system*). Per promuovere il 'trend' di dati bibliografici aperti, collegati e usabili e, soprattutto, adeguarsi alle leggi della globalizzazione sull'apertura e condivisione dei dati, è importante creare una community internazionale che possa contribuire attivamente alla messa in pratica dei Principi di dati bibliografici aperti

(<<http://openbiblio.net/principles/>>) e, conseguentemente, allo sviluppo di “*bibliographic linkable data*” in tutto il mondo, affinché l’intera società tragga pienamente beneficio dal lavoro bibliografico con i dati bibliografici liberamente disponibili per l’uso e il riuso da parte di chiunque e per qualsiasi scopo.

### **Considerazioni finali: liberando il sapere con LOD**

Come abbiamo visto, la comunicazione in rete basata sull’architettura semantica di *Linked Open Data* è un’espressione pratica del web semantico, utile e realizzabile per ‘liberare’ e rendere interoperabile il sapere nel web. Il punto forte di LOD è che essi creano l’accesso semantico integrato alle risorse informative attraverso un grosso quantitativo di dati interconnessi mediante i ‘predicati’ e appoggiati dalle ontologie (vocabolari semantici) che, definendo le loro classi, le istanze e le proprietà, consentono la trasformazione di schemi e dati in rete nei grafi RDF. In questo modo, LOD contribuiscono in modo determinante a un progresso significativo delle campagne in favore dell’accesso libero all’informazione, per il quale sono stati decisivi lo sviluppo e la diffusione dei nuovi sistemi di gestione dell’informazione e delle imprese digitali di reti semantiche.

Basandosi sul concetto di connessione logica, LOD hanno la capacità di completare, aggiornare e trasformare i dati di ogni schema di metadati che, naturalmente, non può che non avere delle positive ricadute nella modellazione di applicazioni innovative per gli archivi e le biblioteche digitali richiedenti diversi approcci di rappresentazione, di integrazione e di accesso ai dati. In questo contesto, ricordiamo di nuovo la MODS Ontology, volta a riformattare una vasta quantità di metadati descrittivi come MARC in MODS e RDF, per farli migrare nello spazio libero di LOD, dove verranno formattati in “*triplestores*” e sostenuti dalle query SPARQL. Questi ulteriori vantaggi dell’ontologia MODS potrebbero contribuire a stabilire il potenziale dei (meta)dati bibliografici più ricco da integrarli e interrogarli (via SPARQL) con altri set di LOD, indipendentemente dal formato con il quale i dati sono rappresentati, per renderli visibili, distribuiti ed accessibili da una qualsiasi applicazione.

In questo nuovo scenario di “*know how*” di modellazione della conoscenza, è importante che le istituzioni, in modo cooperante, si pongono il problema di come trasformare il loro “caos” di dati in un vero e proprio universo semantico di dataset interconnessi, cioè di come pubblicare e contestualizzare tali dati sul web, collegarli ad altri dati e renderli più utili e facilmente utilizzabili e, nel lungo periodo, molto più richiesti. La soluzione più incoraggiante (e largamente accettata dalla comunità internazionale) è, appunto, *Linked Data*.

Nonostante le iniziative italiane di avanguardia su *Open Data* come: OpenCamera, CKAN Italia, OKFN Italia, Open Polis, Open Street Map, Datagov.it, Spaghetti Open Data, Open Knowledge Foundation Italia (<<http://it.okfn.org/>>), GFOSS.it, Open Spending; i movimenti come Open Source, Software Libero e Open Access che mirano a “liberare” il sapere in ambito scientifico e culturale, l’Italia ‘denuncia’, comunque, un ritardo nelle esperienze di *Linked Data*, ed è proprio per questa ragione è fondamentale muoversi prontamente dalla semplice pubblicazione di *Open Data*, al loro inserimento nel globale *Semantic Web of Data*, rendendoli, cioè, dei *Linked Open Data* (vedi

<<http://www.linkedopendata.it/>>) che, pubblicati sotto le appropriate licenze, possano essere non solo apertamente consultabili e condivisibili ma anche direttamente interrogabili, manipolabili e riutilizzabili da qualsiasi applicazione e soggetto, indipendentemente da linguaggi di programmazione, tecnologie e ragione sociale, all'insegna di creazione di una conoscenza virtuosa e senza limiti.

## Bibliografia con siti di riferimento

- [1] Andy, P., Nilsson, M., Naeve, A., Johnston, P., Baker, T. "DCMI Abstract Model". DCMI Recommendation, 2007, <<http://dublincore.org/documents/2007/06/04/abstract-model/>>
- [2] Berners-Lee, T., Hendler, J., and Lassila, O. *The Semantic Web*. "Scientific American", 284 (2001) 5, p. 34-43.
- [3] Bizer, C., Heath, T., Ayers, D., Raimond, Y. *Interlinking Open Data on the Web*. Proceedings Poster Track, ESWC2007, Innsbruck, Austria, 2011, <<http://www4.wiwiw.fu-berlin.de/bizer/pub/LinkingOpenData.pdf>>
- [4] Bizer, C., Heath, T., Berners-Lee, T. *Linked Data - The Story So Far*. Preprint of a paper to appear in Heath, T., Hepp, M., and Bizer, C. (eds.). *Special Issue on Linked Data, International Journal on Semantic Web and Information Systems (IJSWIS)*, <<http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>>
- [5] Bizer, C., Heath, T., Idehen, K., Berners-Lee, T. *Linked Data on the Web*. Proceeding of the Seventeenth international conference on World Wide Web, Beijing, China 2008, <<http://www2008.org/papers/pdf/p1265-bizer.pdf>>
- [6] Buonazia, I. Masci, M. E. Il PICO Application Profile. Un Dublin Core Application Profile per il Portale della Cultura Italiana. Seminario nazionale di studi, Roma, 2007, <<http://www.otebac.it/getFile.php?id=128>>
- [7] Caffa, R., La convergenza tra archivi, biblioteche e musei e l'interoperabilità. Iniziative concrete: CulturalItalia, MICHAEL, MINERVAeC, Seminario nazionale di studi, Roma, 2007, <<http://www.otebac.it/getFile.php?id=132>>
- [8] Cho, A. *How RDF Can Use MARC in the Semantic Web World. Using Existing Library Cataloguing Methods in Organizing the Web*. "Internet by Suite 101", 2009, <<http://www.suite101.com/content/how-marc-realizes-the-rdf-in-the-semantic-web-a122881>>
- [9] Coppens, S., Mannens, E., Deursen, D.V. *Publishing Provenance Information on the Web using the Memento Datetime Content Negotiation*, LDOW2011, 2011, Hyderabad, India, <<http://www.google.it/url?sa=t&source=web&cd=1&ved=0CDAQFjAA&url=http%3A%2F%2Fevents.linkeddata.org%2Fldow2011%2Fpapers%2Fldow2011-paper02-coppens.pdf&ei=yCLvTcCoZDesgalhNWtCg&usq=AFQjCNHkNPAm1KKubZ1nOtxXBoaa2B-WDA>>
- [10] De Robbio, A. *Archivi aperti e comunicazione scientifica*, Napoli : ClioPress, 2007, <<http://www.storia.unina.it/cliopress/derobbio.htm>>
- [11] De Robbio, A. *Metadati per la comunicazione scientifica*, "Biblioteche Oggi", 9 (2001), p.20-22, <<http://www.bibliotecheoggi.it/2001/20011005401.pdf>>
- [12] Di Donato, F. *Lo stato trasparente. Linked Open Data e cittadinanza attiva*, Pisa, ETS, 2010, <<http://www.linkedopendata.it/wp-content/uploads/statotrasparente.pdf>>
- [13] Fini, A., Vanni, L. *Learning object e metadati. Quando, come e perché avvalersene*, Erickson, Trento 2004
- [14] Gargiulo, P. *Open Access: nuove prospettive nel campo degli strumenti di ausilio al mondo della ricerca scientifica*. "CASPUR Annual Report", 2008, p.58-60, <<http://eprints.rclis.org/handle/10760/13433>>
- [15] Giglia, E. *Institutional archives for research: experiences and projects in Open Access*. "Library Hi Tech News", 2(2007), p.6-8, <[http://eprints.rclis.org/bitstream/10760/12434/1/Roma\\_ISS\\_LHTN.pdf](http://eprints.rclis.org/bitstream/10760/12434/1/Roma_ISS_LHTN.pdf)>
- [16] Gruber, T. R. *Toward principles for the design of ontologies used for knowledge sharing*. "International Journal of Human-Computer Studies", 43(1995) 4-5, p.907-928, <<http://tomgruber.org/writing/onto-design.pdf>>
- [17] Guerrini M., *Gli archivi istituzionali. Open access, valutazione della ricerca e diritto d'autore*, a cura di Andrea Capaccioni, con saggi Antonella De Robbio, Roberto Delle Donne, Rosa Maiello e Andrea Marchitelli, Milano, Editrice Bibliografica, 2010.
- [18] Hartig, O., Huber, F. *A Main Memory Index Structure to Query Linked Data*. WWW2011 workshop: Linked Data on the Web (LDOW2011), <<http://events.linkeddata.org/ldow2011/papers/ldow2011-paper06-hartig.pdf>>
- [19] Heath, T., Bizer, C. *Linked Data: Evolving the Web into a Global Data Space*, 1st edition. *Synthesis Lectures on the Semantic Web: Theory and Technology*, Morgan & Claypool, 2011, <<http://linkeddatatoolkit.com/editions/1.0/>>
- [21] Krichel, T., Warner, S.M. *A metadata framework to support scholarly communication*. International Conference on Dublin Core and Metadata Applications, Tokyo, 2001, <<http://eprints.rclis.org/bitstream/10760/4245/1/kanda.a4.pdf>>
- [22] Kruk, S.R., Synak, M., Zimmermann, K. *MarcOnt - Integration Ontology for Bibliographic Description Formats*. DCMI '05 Proceedings of the 2005 international conference on Dublin Core and metadata applications: vocabularies in practice, 2005,

- [http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.155.201&rep=rep1&type=pdf:](http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.155.201&rep=rep1&type=pdf)  
 <<http://dc2005.uc3m.es/program/presentations/Thursday%2015.%2015.30%20h%20-%20s.kruk.pdf>>
- [23] Kunze, J.A. *Towards Electronic Persistence Using ARK Identifiers*. Draft. California Digital Library, 2003, <<http://pid.ndk.cz/dokumenty/zakladni-literatura/arkcdl.pdf>>
- [24] *Le ontologie*. Volume monografico, a c. di Maria Teresa Biagetti. "AIDAinformazioni", 1-2 (2010), <<http://www.aidainformazioni.it/2010/122010.html>>
- [25] *Manifesto per le biblioteche digitali. Gruppo di studio sulle biblioteche digitali*, a c. di Angela Di Iorio e della Redazione AIB-WEB, 2005, <[http://it.wikisource.org/wiki/Manifesto\\_per\\_le\\_biblioteche\\_digitali](http://it.wikisource.org/wiki/Manifesto_per_le_biblioteche_digitali)>
- [26] Nilsson, M. *DCMI Description Set Profile Model*. Working Draft, 2007, <<http://dublincore.org/architecture/wiki/DescriptionSetProfile>>
- [27] Nilsson, M., Powell, A., Pete Johnston, Ambjörn Naeve, 2008, *Expressing Dublin Core metadata using the Resource Description Framework (RDF)*. DCMI Recommendation, <<http://dublincore.org/documents/2008/01/14/dc-rdf/>>
- [28] Oh, S. *MARC, FRBR and RDA: Topic Maps Perspective*, JTC1 SC34 Chairman, Sungkyun University, <<http://www.topicmaps.com/tm2008/oh.ppt>>
- [29] Pretorius, A. J. *Ontologies - Introduction and Overview*. MSc Thesis, Brussels, Vrije Universiteit Brussel, 2004, <[http://www.starlab.vub.ac.be/teaching/Ontologies\\_Intr\\_Overv.pdf](http://www.starlab.vub.ac.be/teaching/Ontologies_Intr_Overv.pdf)>
- [30] Rusu, D., Fortuna, B., Mladenici, D. *Automatically Annotating Text with Linked Open Data*. WWW2011 workshop: Linked Data on the Web (LDOW2011), 2011, <<http://events.linkedata.org/ldow2011/papers/ldow2011-paper09-rusu.pdf>>
- [31] Sacco, O., Passant, A. *A Privacy Preference Ontology (PPO) for Linked Data*. WWW2011 workshop: Linked Data on the Web (LDOW2011), 2011, <<http://events.linkedata.org/ldow2011/papers/ldow2011-paper01-sacco.pdf>>
- [32] Tennant, R. *A Bibliographic Metadata Infrastructure for the 21st Century*. "Library Hi Tech", 22 (2004) 2, <<http://roytennant.com/metadata.pdf>>
- [33] Trigari, M. *L'interoperabilità semantica nei metadati*. Seminario nazionale di studi, Roma, 2007, <<http://www.otebac.it/getFile.php?id=134>>
- [34] Van de Sompel, H., Sanderson, R., Nelson, M.L., Balakireva, L., Shankar, H., Ainsworth, S. *Memento: Time Travel for the Web*. Submitted to arXiv.org, Cornell University Library, 2009, <[http://arxiv.org/PS\\_cache/arxiv/pdf/0911/0911.1112v2.pdf](http://arxiv.org/PS_cache/arxiv/pdf/0911/0911.1112v2.pdf)>
- [35] Weinstein, P.C. *Ontology-Based Metadata: Transforming the MARC Legacy*. Proceedings of the Third ACM Digital Library conference, Pittsburgh, PA, USA, <<http://www-personal.umich.edu/~peterw/Ontology/beethoven.paper.rtf>>
- [36] Weston, P.G. *I metadati e il catalogo elettronico. Interoperabilità di contenuti e servizi digitali: metadati, standard e linee guida*. Seminario nazionale di studi, Roma, 2007, <<http://www.otebac.it/getFile.php?id=131>>
- [37] Yee, M. M. *Can Bibliographic Data be put Directly onto the Semantic Web?* "Information technology and libraries", 28 (2009) 2, p. 55 - 80, <<http://escholarship.org/uc/item/91b1830k;jsessionid=8FF812FAE44110D9AA179065C6B8A27F#page-1>>

## Note con siti di riferimento

- [38] AllegroGraph® RDFStore 4.2.1, <http://www.franz.com/agraph/allegrograph/>, <[http://www.chrisfrymann.com/image/mods/rdf\\_graph.png](http://www.chrisfrymann.com/image/mods/rdf_graph.png)>
- [39] Bizer, C., Cyganiak, R., Heath, T. *How to Publish Linked Data on the Web*. Tutorial 2007, <<http://sites.wiwiwiss.fu-berlin.de/suhl/bizer/pub/LinkedDataTutorial/>>
- [40] Bizer, C., Gauß, T., Cyganiak, R., Hartig, O. *Semantic Web Client Library, Querying the complete Semantic Web with SPARQL*, Freie Universität Berlin, 2009, <<http://www4.wiwiwiss.fu-berlin.de/bizer/ng4j/semwebclient/>>
- [41] CKAN - the Data Hub, <<http://ckan.net/group/bibliographic>>  
 CKAN Italia, <<http://blog.okfn.org/2010/06/14/launch-of-itckanet-for-open-data-in-italy/>>, <<http://ckan.net/package/italy/italyregionalaccounts>>
- [42] Conformance Licenses, <<http://www.opendefinition.org/licenses/#Data>>
- [43] *Cool URIs for the Semantic Web*. W3C Working Draft, 2008, <<http://www.w3.org/TR/2008/WD-cooluris-20080321/>>
- [44] CultureSampo - Finnish Culture on the Semantic Web 2.0, <<http://www.seco.tkk.fi/applications/kulttuurisampo/>>
- [45] D'Arcus, B., Giasson, F. *Bibliographic Ontology Specification*. Specification document 2009, <<http://bibliontology.com/specification>>
- [46] DataFinland, <<http://www.seco.tkk.fi/linkedata/datasuomi/>>
- [47] DBLP Computer Science Bibliography, <<http://www.informatik.uni-trier.de/~ley/db/>>
- [48] Dekkers, M. *Core Metadata Initiative*. Status report, 2011, <<http://dublincore.org/news/communications/statusreports/2011/03/>>
- [49] Digital Library Program, <<http://libraries.ucsd.edu/about/digital-library/index.html>>

- [50] Dublin Core Metadata Element Set, v.1.0, 2010, <http://dublincore.org/documents/dces/>; ISO 15836:2009, <http://dublincore.org/documents/dces/#ISO15836>
- [51] Dublin Core Metadata Initiative (DCMI) Specifications, 2011, <http://dublincore.org/specifications/>
- [52] Finnish General Upper Ontology (YSO), <http://www.seco.tkk.fi/ontologies/yso/>
- [53] FinnONTO: National Semantic Web Ontology Project in Finland, 2003-2012, <http://www.seco.tkk.fi/projects/finnonto/>
- [54] Frymann, C. MARC/MODS and Automating Migration to Linked-Data Standards, 2009, University of California San Diego Libraries, <http://www.chrisfrymann.com/category/standard/>
- [55] Frymann, C. *Another Step Toward Lifting Library Metadata into the Cloud*. Blog Musings, Digital Library Architect, University of California San Diego Libraries, 2009, <http://www.chrisfrymann.com/2009/07/22/mods-ontology-2/>
- [56] Functional Requirements for Bibliographic Records (FRBR), 2011, <http://www.ifla.org/en/publications/functional-requirements-for-bibliographic-records>
- [57] Graph Visualization Tool (RDF Gravity). User Documentation. Knowledge Information Systems Group, Austria, <http://semweb.salzburgresearch.at/apps/rdf-gravity/>
- [58] HealthFinland, <http://www.seco.tkk.fi/applications/terveysuomi/>
- [59] History Ontology (HISTO), <http://www.seco.tkk.fi/ontologies/histo/>
- [60] ICONClass Ontology, <http://www.seco.tkk.fi/ontologies/iconclass/>; <http://www.iconclass.nl/>
- [61] Jakob, M. The DBpedia Data Set, 2011, <http://wiki.dbpedia.org/Datasets>
- [62] Krichel, T. ReDIF v.1, revision of 2007, [http://openlib.org/acmes/root/docu/redif\\_1.html](http://openlib.org/acmes/root/docu/redif_1.html)
- [63] La biblioteca di ARCHIVI, Sistema Archivistico Nazionale, <http://www.archivi.beniculturali.it/Biblioteca/DCbiblioteca.html>
- [64] Library Data Resources. W3C, 2011, [http://www.w3.org/2005/Incubator/ld/wiki/Library\\_Data\\_Resources](http://www.w3.org/2005/Incubator/ld/wiki/Library_Data_Resources)
- [65] Library of Congress MODS XML schema, <http://www.loc.gov/standards/mods/v3/mods-3-3.xsd>
- [66] Linked Data FAQ, Structured Dynamics LLC, [http://structuredynamics.com/linked\\_data.html#question\\_8](http://structuredynamics.com/linked_data.html#question_8)
- [67] Linked Data on the Web (LDOW2011), WWW2011 workshop, <http://events.linkedata.org/ldow2011/>
- [68] LinkingOpenData, W3C SWEO Community Project, <http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>
- [69] LinkedOpenData.it, <http://www.linkedopendata.it/>, nasce dalla collaborazione tra un gruppo di "appassionati" di OpenData e due aziende pisane (Net7 e Hyperborea) che da tempo lavorano in settori come Digital Libraries, Web 2.0 e Semantic Web.
- [70] MODS ontology, <http://www.chrisfrymann.com/2009/05/21/mods-ontology/>
- [71] MuseumFinland: Finnish Museums on the Semantic Web, <http://www.museosuomi.fi/>
- [72] MusicBrainz, <http://musicbrainz.org/>; Geonames, <http://www.geonames.org/>
- [73] Ontology for museum domain (MAO), <http://www.seco.tkk.fi/ontologies/mao/>
- [74] Ontology: Ontologies on semanticweb.org, <http://semanticweb.org/wiki/Ontology>
- [75] Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), <http://www.openarchives.org/pmh/>
- [76] Open Knowledge Definition: Defining the Open in Open Data, Open Content and Open Services, <http://www.opendefinition.org/okd/>; Open Knowledge Foundation, <http://okfn.org/>
- [77] Open Knowledge Foundation Italia, <http://it.okfn.org/2011/02/19/linkedopendata-it-una-piattaforma-italiana-per-i-dati-%e2%80%9caperti%e2%80%9d-e-%e2%80%9ccollegati%e2%80%9d/>. Openbiblio Italia, <http://it.okfn.org/>
- [78] Open Data in science - technical and cultural aspects. CERN workshop on Innovations in Scholarly Communication (OAI4), 2005, <http://indico.cern.ch/contributionDisplay.py?contribId=17&sessionId=9&confId=0514>
- [79] OpenCyc 2.0: open source version of the Cyc technology, <http://www.opencyc.org/>
- [80] Partially complete MODS ontology candidate, [http://www.chrisfrymann.com/mods/mods\\_TB.owl](http://www.chrisfrymann.com/mods/mods_TB.owl)
- [81] PREMIS OWL, <http://multimedialab.elis.ugent.be/users/samcoppe/ontologies/Premis/index.html>
- [82] RDF Schema: Vocabulary Description Language 1.0: RDF Schema, W3C Recommendation, 2004, <http://www.w3.org/TR/rdf-schema/>
- [83] RDF Semantics. W3C Recommendation, 2004, <http://www.w3.org/TR/2004/REC-rdf-mt-20040210/>
- [84] *Semantic Web. Linked Data*. Talk of Tim Berners-Lee on Linked Data Open Italia, <http://www.linkedopendata.it/semantic-web>
- [85] SKOS Simple Knowledge Organization System. W3C SKOS Core Vocabulary Specification, <http://www.w3.org/2004/02/skos/>
- [86] SPARQL Query Language for RDF. W3C Recommendation, 2008, <http://www.w3.org/TR/rdf-sparql-query/>. Esempio di query utili per capire come interrogare i dati con SPARQL, <http://www.linkedopendata.it/datasets/musei>
- [87] TED: ideas worth spreading, <http://www.ted.com/pages/about/>; <http://www.linkedopendata.it/semantic-web>
- [88] The Finnish Ontology of Photography VALO, <http://www.seco.tkk.fi/ontologies/valo/>
- [89] *Tim Berners-Lee on the next web*. Talk on TED 2009,

<[http://www.ted.com/talks/tim\\_berners\\_lee\\_on\\_the\\_next\\_web.html](http://www.ted.com/talks/tim_berners_lee_on_the_next_web.html)>

[90] UCSD Libraries' Digital Library Program, <<http://libraries.ucsd.edu/about/digital-library/index.html>>

[91] Union List of Artist Names® Online, ULAN, <<http://www.getty.edu/research/tools/vocabularies/ulan/index.html>>

[92] Utility Tool. Revision 9541: /RDFizers/marcmods2rdf, <<http://simile.mit.edu/repository/RDFizers/marcmods2rdf/>>

[93] Web Ontology Language (OWL). Linguaggio Ontologico per il Web, Raccomandazione W3C, 2004,

<<http://digilander.libero.it/giovanideangelis/OWLWebOntologyLanguageGuida.htm>>; <<http://www.w3.org/TR/owl-features/>>

[94] Wikipedia, <http://www.wikipedia.org/>; WikiGuida, <[http://www.wikimedia.it/index.php/Pagina\\_principale](http://www.wikimedia.it/index.php/Pagina_principale)>

[95] WordNet: a lexical database for English. The Trustees of Princeton University, 2011, <<http://wordnet.princeton.edu/>;  
<<http://wordnet.princeton.edu/>; <http://multiwordnet.itc.it/english/home.php>>

*Siti consultati in data 2011-06-15*