

# Uno sguardo sul futuro semantico dell'universo bibliografico

di Iryna Solodovnik

«It's critical for Libraries to begin preparations to become full participants in the world of Linked Data»<sup>1</sup>.

## **Dal Web semantico all'elaborazione di nuove strategie comunicative per l'universo bibliografico**

Il termine *Semantic Web* è stato coniato nel 2001 da Tim Berners-Lee<sup>2</sup> per definire l'estensione del Web attuale, attribuendone la capacità di comprendere la semantica dei documenti digitali in rete e, in fase di ricerca, di interagire con l'utente nelle operazioni sofisticate di scoperta ed estrazione dei contenuti, nonché di elaborazione di nuovi *set* di risorse di conoscenza<sup>3</sup>.

L'idea del Web semantico nasce semplicemente dall'ampliamento dell'idea di utilizzare schemi di metadati (le informazioni relative ai dati, attraverso le quali è possibile ricavare informazioni sulla risorsa a cui sono associate) per descrivere domini di informazione, mappandone i dati rispetto a classi o concetti. «Un punto di forza principale del Web semantico è sempre stata l'espressione, sul Web, della grande quantità di informazioni del database relazionale formulate in una modalità processabile da una macchina. Il formato di serializzazione RDF [...] è un formato fun-

IRYNA SOLODOVNIK, Scuola dottorale internazionale degli studi umanistici (SDISU), Università della Calabria, via P. Bucci, Cubo 28/B, 87036 Arcavacata di Rende (CS), e-mail iryna.solodovnik@gmail.com. Ultima consultazione siti web: 12 novembre 2012.

**1** Gillian Byrne – Lisa Goddard, *The Strongest Link: Libraries and Linked Data*, «D-Lib Magazine», 16 (2010), n. 11-12, <<http://dlib.org/dlib/november10/byrne/11byrne.html>>.

**2** Tim Berners-Lee – James Hendler – Ora Lassila, *The Semantic Web*, «Scientific American Magazine», May 2001, <<http://www.scientificamerican.com/article.cfm?id=the-semantic-web>>.

**3** Zareen Syed – Tim Finin, *Creating and Exploiting a Hybrid Knowledge Base for Linked Data*, in: *Agents and Artificial Intelligence*, Joaquim Filipe, Ana Fred, Bernadette Sharp (eds.) «Revised Selected Papers Series: Communications in Computer and Information Science», Springer, 129 (2011), p. 3-24, <<http://ebiquity.umbc.edu/paper/html/id/535>>.

zionale ad esprimere le informazioni di database relazionale»<sup>4</sup>. In questo modo si può disporre di strutture in grado di descrivere e automatizzare i collegamenti esistenti fra i dati, che nel Web semantico sono composti di tre livelli fondamentali. Al livello più basso abbiamo i dati, i metadati riportano questi dati ai concetti di uno schema, nello schema (spesso chiamato ontologia) si esprimono le relazioni fra concetti, che diventano classi di dati coinvolte nella strutturazione della conoscenza di una certa realtà (processi, domini di informazione)<sup>5</sup>. Il genere più tipico di ontologia per il Web è rappresentato da una tassonomia che definisce le classi di oggetti e le relazioni tra loro, e da una serie di regole di inferenza.

Le singole applicazioni del Web semantico possono operare processi di inferenza utilizzando specifici vocabolari definiti con i linguaggi formali e semi-formali comprensibili dai computer, interrogando così la semantica dei dati pubblicati sul Web.

In questo contesto, si aprono preziose opportunità per lo sviluppo e l'utilizzo di una nuova generazione di formati di metadati, di tecnologie per le biblioteche digitali e per la generazione di dati bibliografici di alta qualità sfruttando le tecnologie come Linked Data – «a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF»<sup>6</sup> – atte a connettere i dati web tra di loro e con diversi schemi di dati di autorità pubblicati, arricchiti e validati sul Web in modo partecipato e condivisibile da diverse comunità di utenti. «Nell'universo bibliografico c'è un chiaro cambiamento di paradigma da formati record fissi alle ricombinabili dichiarazioni sui metadati. Le istituzioni e i sistemi bibliotecari che implementano archivi di dati bibliografici e relativi servizi dovranno necessariamente prevedere processi e strategie di fornitura dati compatibili con il framework Linked Data e di licenze adeguate»<sup>7</sup>.

### Tecnologie e soluzioni del Web semantico

Le tecnologie più rilevanti del Web semantico, definiti dal World Wide Web Consortium (W3C) sono:

- 1) XML (eXtensible Markup Language), una sintassi elementare per strutturare il contenuto dei documenti digitali. Attualmente, nelle tecnologie del Web semantico, XML non è una componente necessaria, considerando l'esistenza di una sintassi alternativa come Turtle (Terse RDF Triple Language), uno standard *de facto* per la serializzazione dei grafi RDF.
- 2) XML Schema, un linguaggio di descrizione del contenuto XML, il cui scopo è delineare quali elementi sono permessi, quali tipi di dati sono associati ad essi e quale relazione gerarchica hanno fra loro gli elementi contenuti in un file XML.
- 3) URIs (Uniform Resource Identifiers), indirizzi/identificatori univoci di risorse sul

<sup>4</sup> Mauro Guerrini – Tiziana Possemato, *Linked data: un nuovo alfabeto del web semantico*, «Biblioteche oggi», 30 (2012), n. 3, p. 10, <<http://www.bibliotecheoggi.it/content/201200300701.pdf>>.

<sup>5</sup> Claudio Gnoli – Carlo Scognamiglio, *Ontologia e organizzazione della conoscenza*, Lecce: Pensa Multimedia, 2008. V. anche il volume monografico *Le ontologie*, a cura di Maria Teresa Biagetti, «AIDA Informazioni», 28 (2010), n. 1-2, <<http://www.aidainformazioni.it/2010/122010.html>>.

<sup>6</sup> Linked Data, *Connect Distributed data across the web (the site of the Linked Data community)*, <<http://linkeddata.org/>>.

<sup>7</sup> Antonella De Robbio, *Forme e gradi di apertura dei dati. I nuovi alfabeti dell' Open Biblio tra scienza e società*, «Biblioteche oggi», 30 (2012), n. 6, pp. 16-17, <<http://www.bibliotecheoggi.it/content/201200601101.pdf>>.

Web. URIs possono esprimersi tramite URN, DOI, URL accompagnati dal prefisso HTTP in modo che gli oggetti possano essere individuati da persone e da *user agent* sul Web. 4) RDF (Resource Description Framework), una sintassi formale di base per la codifica, lo scambio e il riutilizzo di metadati strutturati che possono descrivere la semantica di una risorsa, sia di quella reperibile sul Web (una pagina HTML, un documento XML o parte di esso), sia di quella che non si trova direttamente (un libro, un quadro). L'utilizzo di RDF per la descrizione semantica delle risorse offre, ovviamente, numerosi benefici in molti settori del Web: si potrebbero sviluppare diverse connessioni semantiche per aggregare i dati correlati, si potrebbero lanciare motori di ricerca più efficienti in grado di basare la ricerca non solo sulle occorrenze di parole contenute nei documenti ma anche in base alla caratterizzazione semantica dei documenti stessi, si potrebbero realizzare agenti software per il filtraggio dei contenuti di una risorsa in funzione di determinati criteri semantici impostati dall'utente.

RDF si esprime attraverso due componenti: RDF Model and Syntax<sup>8</sup> e RDF Schema (RDFS o RDF Vocabulary Description Language)<sup>9</sup>. Mentre la prima componente definisce il modello dei dati specificato attraverso la sintassi (RDF/XML, N3, Turtle, RDFa) per descrivere le risorse, la seconda permette di definire il significato e le caratteristiche delle proprietà e delle relazioni che esistono tra le risorse descritte dalla prima componente. RDF Data Model è basato su uno *statement* costituito da una tripla del tipo soggetto (una risorsa descritta mediante RDF), predicato (una proprietà, un attributo, una relazione definiti tra risorsa e valore tramite RDF Schema) e oggetto (un valore della proprietà) accompagnati, nella maggior parte di casi, da URIs che puntano ad altre risorse, consentendo così l'interoperabilità tra applicazioni che scambiano informazioni sul Web. La triplicazione automatica dei dati secondo l'articolazione soggetto-predicato-oggetto – ad esempio, la pagina <<http://www.w3.org/DesignIssues/LinkedData.html>> è scritta dall'autore Tim Berns Lee – è strettamente correlata al nostro modo di pensare e di costruire concetti (figura 1):

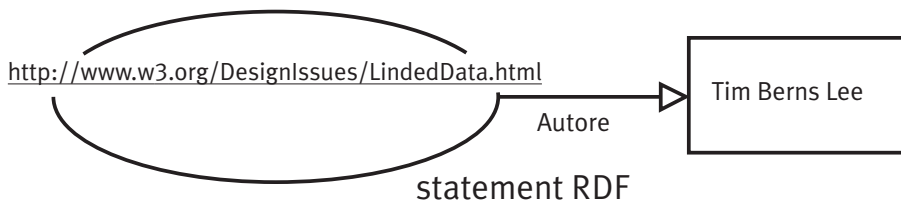


Figura 1. L'esempio di componenti di uno statement RDF

<sup>8</sup> World Wide Web Consortium (W3C), *RDF/XML Syntax Specification (revised)*, <<http://www.w3.org/tr/REC-rdf-syntax/>>.

<sup>9</sup> World Wide Web Consortium (W3C), *RDF Vocabulary Description Language 1.0: RDF Schema*, <<http://www.w3.org/tr/rdf-schema/>>.

L'oggetto (Tim Berns Lee) dello statement rappresentato in figura 1, a sua volta, può essere un soggetto (risorsa) per aprire possibili percorsi di approfondimento sulla proprietà della risorsa. In questa prospettiva, dallo *statement* emerge un grafo RDF di conoscenza, rappresentato da un insieme di nodi interconnessi da URIs (ove è possibile) che identificano le loro proprietà (figura 2).

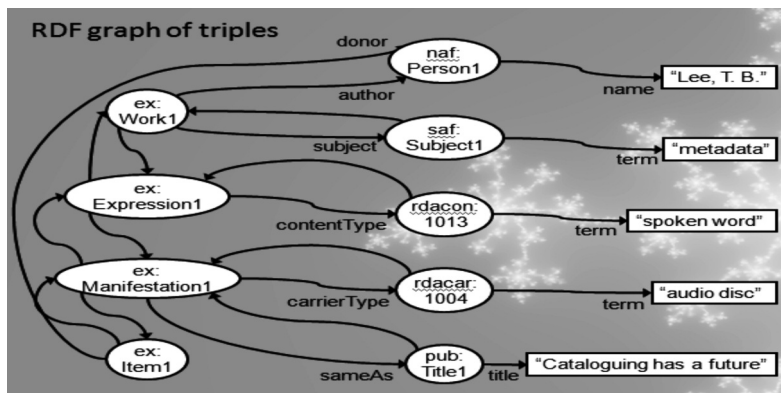


Figura 2. L'esempio del grafo formato da triple RDF basato sulla scheda del catalogo<sup>10</sup>

La tecnologia RDFS rappresenta un linguaggio ontologico semplice<sup>11</sup>, che consente di creare una sintassi per definire classi e proprietà per interpretare la semantica delle risorse. In un contesto RDF, le ontologie sono i vocabolari (definizioni di classi, istanze, proprietà) che catturano le strutture di dati di grafi RDF collegati. Nonostante l'esistenza di molti vocabolari RDF di dominio pubblico, che possono essere facilmente riutilizzati se coincidono con la semantica del dominio *target*, molti domini applicativi richiedono ontologie specifiche che possono riflettere meglio la semantica di dati e relazioni. Riconoscendo, però, i limiti di espressività di RDF e RDFS – in quanto RDF consente unicamente di indicare predicati binari (valore-attributo), e RDFS consente solo di stabilire gerarchie di classi e proprietà e di imporre vincoli per dominio e codominio – il W3C ha definito il Web Ontology Language (OWL).

5) Il linguaggio OWL<sup>12</sup> permette di rappresentare esplicitamente il significato e la semantica di dati mediante vocabolari (ontologie) web formali. OWL esprime non solo proprietà (*properties*), classi (*classes*), istanze delle classi (*individuals*) e relazioni (di disgiunzione, di cardinalità, di uguaglianza) tra le classi, ma anche esporta tale conoscenza a diverse applicazioni, contribuendo maggiormente alla loro interoperabilità. L'espressività di OWL è costituita da tre sottolinguaggi: OWL Lite, OWL DL (Description Logic),

<sup>10</sup> Gordon Dunsire, *Linked Data, Libraries and the Semantic Web*, in: *Library science talk*, 12-13 March 2012, <[http://www.nb.admin.ch/aktuelles/ausstellungen\\_und\\_veranstaltungen/00726/01611/03953/03958/index.html?lang=en](http://www.nb.admin.ch/aktuelles/ausstellungen_und_veranstaltungen/00726/01611/03953/03958/index.html?lang=en)>.

<sup>11</sup> Per approfondire vedere il *Manuale per l' interazione con gli utenti del Web culturale*, a cura di Pierluigi Feliciati e Maria Teresa Natale, 2009, <[http://www.minervaeurope.org/publications/handbookwebusers\\_it/chapter4\\_5.html](http://www.minervaeurope.org/publications/handbookwebusers_it/chapter4_5.html)>.

<sup>12</sup> World Wide Web Consortium (W3C), *OWL Web Ontology Language. Overview*, <<http://www.w3.org/TR/owl-features/>>.

OWL Full, ciascuno dei quali riproduce un'estensione del suo modello precursore più semplice (non viceversa). OWL Lite rappresenta un percorso di migrazione più rapida per thesauri e altre tassonomie gerarchiche e sistemi inferenziali, supportando, però, solo i valori di cardinalità a zero o uno, ed escludendo alcuni costrutti che esprimono relazioni tra classi (ad esempio disgiunzione, unione, istanziazione). OWL DL è abbastanza espressivo per trattare tutti i costrutti OWL, assegnando, però, solo alcuni vincoli sul loro uso (ad esempio una classe può essere sottoclasse di un'altra classe, ma non un'istanza di una classe; non si possono esprimere restrizioni di cardinalità per le proprietà transitive). OWL Full rende la massima espressività e la libertà sintattica dei costrutti RDFS, ma è difficilmente implementabile.

6) La tecnologia Simple Knowledge Organization System (SKOS)<sup>13</sup>, basata su RDF e RDFS, che supporta la rappresentazione e agevola la pubblicazione dei sistemi per l'organizzazione della conoscenza (KOS, Knowledge Organisation Systems) quali thesauri, schemi di classificazione, soggetti, schemi di instestazione per soggetto, e altri vocabolari controllati strutturati nel contesto del Web semantico.

7) SPARQL (Simple Protocol and RDF Query Language)<sup>14</sup>, un linguaggio di interrogazione dei grafi e sottografi RDF, il cui output sarà l'estrazione delle informazioni dalle basi di conoscenza distribuite sul Web.

Le tecnologie appena descritte rappresentano strati sovrapposti di linguaggi, ognuno dei quali usa o estende gli strati precedenti. Il collocamento di queste tecnologie può essere osservato nella figura 3.

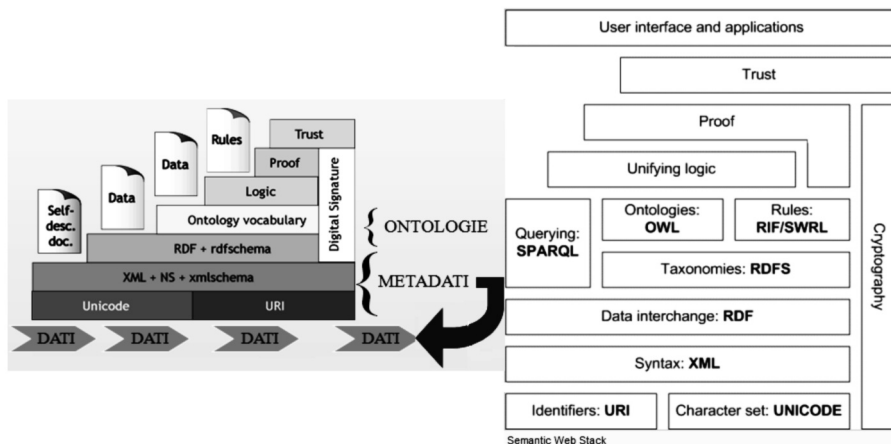


Figura 3. Le tecnologie del Web semantico (Semantic Web Layer-Cake Diagram e Semantic Web Stack<sup>15</sup>)

13 World Wide Web Consortium (W3C), *SKOS Simple Knowledge Organization System Primer*, <<http://www.w3.org/TR/skos-primer>>; *SKOS Use Cases and Requirements*, <<http://www.w3.org/TR/2009/NOTE-skos-ucr-20090818/>>.

14 World Wide Web Consortium (W3C), *SPARQL Query Language for RDF*, <<http://www.w3.org/TR/rdf-sparql-query/>>.

15 *Semantic Web Architecture*, <<http://orbitko.com/tutorials/ontologies-semantic-web/semantic-web-architecture.html>>; *Semantic Web*, <[http://en.wikipedia.org/wiki/Semantic\\_Web](http://en.wikipedia.org/wiki/Semantic_Web)>.

Fino al livello delle ontologie (metadati strutturati secondo un ordine logico e relazionati attraverso i concetti di un dominio di conoscenza) non vi è inferenza, ma solo rappresentazione della conoscenza. Sopra questo livello si colloca il livello logico che, tramite certi linguaggi (come SWRL, RuleML) e ragionamenti inferenziali automatici, estrae dalle asserzioni (*statements*) web una nuova conoscenza (*machine understandable*), che a sua volta sarà automaticamente integrata e riutilizzata dalle applicazioni. Inoltre, per raggiungere la piena potenzialità, le informazioni del Web semantico devono essere approvate (*proof*) e, di conseguenza, affidabili (*trusted*).

Uno dei problemi fondamentali del Web semantico è quello di rendere disponibili, integrabili e interoperabili le varie tipologie di dati sul Web. Tecnicamente, questo obiettivo può essere raggiunto attraverso opportune tecnologie capaci di operare la conversione da diversi formati di dati e database in RDF<sup>16</sup>, o di eseguire un *markup* RDF automatizzato dei siti (ad esempio attraverso le tecnologie POWDER, RDFa GRDDL, R2RML, RIF, Drupal7).

Tra i maggiori servizi che si prestano alla trasformazione dei dati strutturati in formati RDF/OWL possono essere elencati: Web services links & resources, SemWeb, Beckett, SIMILE (RDF crosswalks), Semantic Bank, D2R Server (per l'esposizione di database relazionali in RDF). Infine, i servizi di *tagging* semantico API (Application Programming Interface: «un modo per i programmi di comunicare tra loro. Si può comprendere nei termini di come un programmatore invia istruzioni tra programmi diversi»<sup>17</sup>) come, ad esempio, OpenCalais e Zemanta sono utili per trasformare blocchi di testi non strutturati in entità principali, *topics*, relazioni, e per eseguire l'estrazione di *keywords*, *tagging* automatico e la disambiguazione di entità e concetti, che possono fungere da output in RDF.

Al fine di sviluppare servizi e applicazioni web semanticamente interoperabili a livello globale, occorre lavorare non solo in direzione dell'armonizzazione di strutture, ma anche di apertura verso i formati dei dati<sup>18</sup> e di uso di una terminologia precisa, specifica e condivisa (coinvolgendo attivamente i meccanismi dell'*authority control*<sup>19</sup> che curano la qualità degli indici delle registrazioni) impiegando i *dataset* dei dati di autorità pubblicati in LOD Cloud, nonché le tecnologie SKOS<sup>20</sup>, sia per evitare ambiguità di significati dei metadati sia per contribuire all'arricchimento semantico dei loro valori. Questo, ovviamente, influirà in modo positivo anche in fase di definizione finale degli opportuni punti di accesso attraverso SPARQL Query Language.

16 W3C RDB2RDF Incubator Group, *A survey of current approaches for mapping of relational databases to RDF*, 2009, <[http://www.w3.org/2005/Incubator/rdb2rdf/RDB2RDF\\_SurveyReport.pdf](http://www.w3.org/2005/Incubator/rdb2rdf/RDB2RDF_SurveyReport.pdf)>.

17 Open Knowledge Foundation, *Open Data Handbook. Glossario*, <<http://opendatahandbook.org/it/glossary.html#term-application-programming-interface>>.

18 Antonella De Robbio, *Forme e gradi di apertura dei dati. I nuovi alfabeti dell' Open Biblio tra scienza e società* cit.

19 *Authority control. Definizioni ed esperienze internazionali. Atti del convegno internazionale, Firenze, 10-12 febbraio 2003*, a cura di Mauro Guerrini e Barbara, Tillet, con la collaborazione di Lucia Sardo. Firenze University Press, Associazione italiana biblioteche, 2003, <<http://www.fupress.com/Archivio/pdf/4383.pdf>>.

20 World Wide Web Consortium (W3C), *SKOS Simple Knowledge Organization System*, <<http://www.w3.org/2004/02/skos/>>.

### Linked Data e altri meccanismi del Web semantico con la loro proiezione sull'universo bibliografico

Il termine Linked Data (LD) è stato coniato da Tim Berners-Lee nel 2006<sup>21</sup> ed è considerato il primo passo verso il vero e proprio Web semantico. LD ha l'obiettivo di strutturare e di connettere i dati sul Web utilizzando URIs e RDF. In particolare, la tecnologia LD è formata dai già citati meccanismi del Web semantico come RDF Data Model, HTTP, URIs per costruire espliciti link RDF tra le entità correlate nel loro significato provenienti dai diversi *dataset* (di domini pubblico e privato)<sup>22</sup> e SPARQL Query per interrogare i costrutti di LD (figura 4).

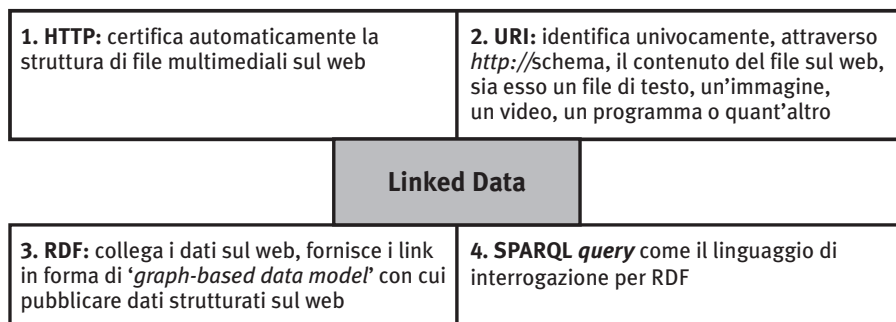


Figura 4. Componenti della tecnologia Linked Data.

I dati provenienti sia dal dominio pubblico (dati aperti) che da quello privato (dati chiusi) possono essere ugualmente espressi come LD. In quanto al rapporto tra Linked Data, Open Data e Web semantico è interessante la definizione data da Paul Walk<sup>23</sup>:

- 1) i dati possono essere aperti, ma non linkati,
- 2) i dati possono essere linkati, ma non essere aperti,
- 3) i dati che siano sia aperti sia linkati sono assolutamente realizzabili e riutilizzabili,
- 4) il Web semantico può funzionare solo con dati che siano sia aperti sia linkati (più dati aperti sono linkati sul Web con un significato ben definito, più le applicazioni web saranno performanti).

L'obiettivo principale della tecnologia LD è quello di trasformare il Web attuale in un Web dei dati (*Web of data*<sup>24</sup>), concepito come un *database* globale contenente

<sup>21</sup> Lisa Goddard – Gillian Byrne, *Linked data tools: semantic web for the masses*, «First Monday», 15 (2010), n. 11, <<http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/3120/2633>>.

<sup>22</sup> «While, to date, it is the case that Linked Data has been demonstrated using public Web data and many desire to expose more through the Open Data movement, there is nothing preventing private, proprietary or subscription data from being Linked Data», in: Al3 (Adaptive Information Innovation Infrastructure), *What is Linked Data?*, <<http://www.mkbergman.com/447/what-is-linked-data/>>.

<sup>23</sup> Paul Walk's weblog, *Linked, open, semantic?*, 2009, <<http://blog.paulwalk.net/2009/11/11/linked-open-semantic/>>.

<sup>24</sup> Tom Heath – Christian Bizer, *Linked Data: evolving the web into a global data space*, «Synthesis lectures on the semantic web: theory and technology», Morgan & Claypool, 2011, <<http://linked-databook.com/editions/1.0/>>.



i dati strutturati linkabili e in RDF e, quindi, interconnessi tra loro, come se fossero pagine web: «si costruisce così un reticolo di dati collegati (Linked Data, appunto) appartenenti a un dominio (che costituisce il contesto di partenza), collegato a sua volta ad altri set di dati esterni, ovvero fuori dal dominio, in un contesto di relazioni sempre più estese»<sup>25</sup>. Il Web dei dati renderebbe i dati web – forniti da diverse comunità di utenti (figura 5) – il più possibile integrabili, condivisibili e riusabili da diverse comunità di utenti, facilitando così l'accesso<sup>26</sup> ai contenuti da parte di molteplici applicazioni che, a loro volta, avrebbero un nuovo livello di integrazione e di interoperabilità<sup>27</sup>, arricchendo ulteriormente la praticità della rete di collegamenti.

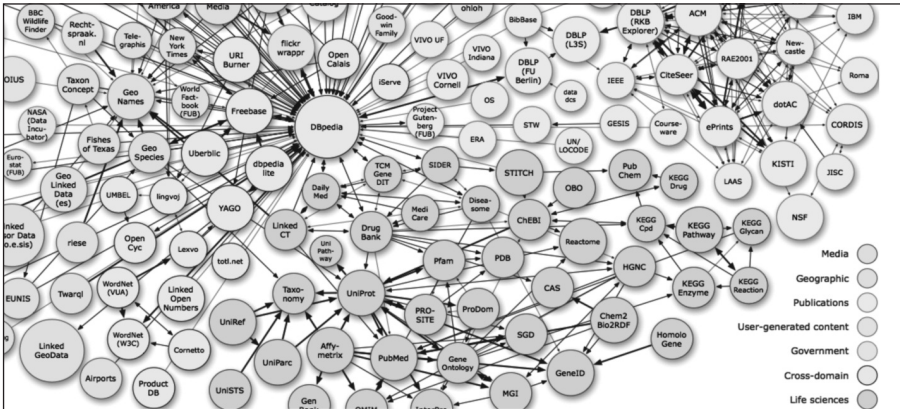


Figura 5. Connessione di diversi dataset nel Linked Open Data (LOD) Cloud Diagram<sup>28</sup>

Il diagramma LOD<sup>29</sup> Cloud è stato disegnato nel 2007 entro il progetto Linking Open Data della SWEO community<sup>30</sup> per visualizzare graficamente la raccolta di dati Open (dati pubblicati sotto le licenze aperte come, ad esempio, la CCo, PDDL, ODC-BY e non condizionati da licenze commerciali o restrizioni d'uso) messi a disposizione sulla rete di Linked Data. Grazie al modello LOD, i dati (in continua crescita) interoperabili entrano a far parte di un

25 Mauro Guerrini – Tiziana Possemato, *Linked data: un nuovo alfabeto del web semantico* cit., p. 7.

26 Richard Cyganiak [et al.], *Semantic Sitemaps: Efficient and Flexible Access to Datasets on the Semantic Web*, in: *Proceedings of the 5th European Semantic Web Conference, ESWC, 2008*, <<http://richard.cyganiak.de/2008/papers/sitemaps-eswc2008.pdf>>.

27 Dublin Core. Building blocks for interoperability, *Designing Interoperable Metadata on Linked Data Principles*, 2009, <[http://dublincore.org/resources/training/NISO\\_Webinar\\_20100825/dcmi-webinar-03.pdf](http://dublincore.org/resources/training/NISO_Webinar_20100825/dcmi-webinar-03.pdf)>.

28 Richard Cyganiak – Anja Jentzsch, *Linking Open Data Cloud Diagram*, <<http://richard.cyganiak.de/2007/10/lod>>.

29 Europeana ha pubblicato un video per spiegare cosa sono i LOD e perché apportano benefici sia agli utenti che ai fornitori di dati: *Linked Open Data. What is it?*, <<http://vimeo.com/36752317>>.

30 World Wide Web Consortium (W3C), *SweoIG/TaskForces/CommunityProjects/LinkingOpenData*, <<http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>>.



immenso database aperto nel quale vengono pubblicati set di dati grezzi resi disponibili da istituzioni diverse che, però, possono essere liberamente incrociati da terze parti, con la possibilità di generare valore aggiunto per i servizi personalizzati, magari in base alle esigenze di una tipologia specifica di utenti. Mentre nel 2007 il diagramma LOD contava più di 2 miliardi di terne RDF provenienti da diversi database web (le prime e più aggiornate sono DBPedia e Freebase<sup>31</sup>), nel 2012 il LOD Cloud ha raggiunto oltre di 52 miliardi di terne RDF.

La crescente richiesta da parte degli utenti di servizi che consentono un elevato grado di interconnessione tra i dati provenienti da diverse fonti ha portato allo sviluppo collaborativo e alla promozione di soluzioni pratiche per LOD in diversi settori, puntando ai fattori di maggiore interesse come l'interoperabilità semantica, la migliore fruibilità, l'ampio riutilizzo dei dati e la loro libera, aperta<sup>32</sup> disponibilità per l'utente. L'importanza di LOD viene ampiamente riconosciuta anche per l'annotazione semantica<sup>33</sup>, per l'arricchimento dei *dataset* web attraverso modelli concettuali e *authority data* resi disponibili attraverso RDF<sup>34</sup>, per l'implementazione di *high-level web-scale services* per la ricerca.

In questa prospettiva e con il taglio tematico per i dati bibliografici, è interessante prendere visione delle iniziative di Archives & Museums (LODLAM) Group, IFLA'S Semantic Web Special Interest Group, Library Linked Data Group, dei progetti come Learning Linked Data, Linked Open Vocabularies, nonché delle relazioni di alcuni eventi recenti come Global Interoperability and Linked Data in Libraries, 2<sup>nd</sup> Linked Open Data Conference<sup>35</sup>, e altri ancora.

Nonostante il costante sviluppo di numerose guide pratiche, di una diversità di progetti e di un ricco *framework* tecnologico per la pubblicazione di LD, mancano ancora i

**31** Evan Sandhaus – Rob Larson, *First 5000 tags released to the Linked Data Cloud*, «New York Times», 29 ottobre 2009, <<http://open.blogs.nytimes.com/2009/10/29/first-5000-tags-released-to-the-linked-data-cloud/>>.

**32** Antonella De Robbio, *Forme e gradi di apertura dei dati. I nuovi alfabeti dell' Open Biblio tra scienza e società* cit. A supporto di LOD vedi anche Foundation Conference of European National Librarians, *Europe's national librarians support Open Data licensing*, 2011, <<http://app.e2ma.net/app2/campaigns/archived/1403149/f14691f55d5483aff43360a9b4aa7d35/>>; Harvard Library, *Open Metadata*, 2012, <<http://openmetadata.lib.harvard.edu/>>.

**33** Delia Rusu – Blaž Fortuna – Dunja Mladenčić, *Automatically Annotating Text with Linked Open Data*, in: *WWW2011 workshop: Linked Data on the Web (LDOW2011)*, Hyderabad (India), 2011, <<http://events.linkedata.org/ldow2011/papers/ldow2011-paper09-rusu.pdf>>.

**34** Gordon Dunsire, *Interoperability and semantics in RDF representations of FRBR, FRAD and FRSAD*, 2010, <<http://www.gordondunsire.com/pubs/pres/InteropSemanticsRDFFRBR.ppt>>; World Wide Web Consortium (W3C), *Use Case Authority Data Enrichment*, <[http://www.w3.org/2005/Incubator/ld/wiki/Use\\_Case\\_Authority\\_Data\\_Enrichment](http://www.w3.org/2005/Incubator/ld/wiki/Use_Case_Authority_Data_Enrichment)>.

**35** LODLAM, *Linked Open Data in Libraries, Archives, and Museums*, <<http://lodlam.net/>>; IFLA, *Semantic Web Special Interest Group*, <<http://www.ifla.org/swsig>>; World Wide Web Consortium (W3C), *W3C Library Linked Data Incubator Group*, <<http://www.w3.org/2005/Incubator/ld/>>; *Learning Linked Data project*, <<http://lld.ischool.uw.edu/wp/About/>>; *Linked Open Vocabularies*, <<http://lov.okfn.org/dataset/lov/index.html>>. Presentazioni del seminario internazionale *Global Interoperability and Linked Data in Libraries*, Firenze, June 2012, «Jlis.it», 3 (2012), <<http://leo.cilea.it/index.php/jlis/issue/view/368/showToc>>; 2<sup>nd</sup> *Linked Open Data Conference from the Cataloguing and Indexing Group in Scotland (CIGS). Opening Library Linked Data to National Heritage: Perspectives on International Practice*, September 2012, <<http://aims.fao.org/events/2nd-linked-open-data-conference-cataloguing-and-indexing-group-scotland-cigs-opening-library>>.

meccanismi di *automated reasoning* su quest'ultimi. Inoltre, una delle critiche più comuni della visione del Web semantico è che gli standard come RDF, OWL e SPARQL sono difficili da comprendere concettualmente ed estremamente complessi da implementare.

Tuttavia, fino ad oggi la propagazione e l'implementazione di tecnologie Linked Data e Open Data allo scopo di mettere sul Web insieme di dati aperti e connessi in modo interoperabile ha coinvolto diverse istituzioni, tra cui:

– il Digital Curation Centre (DDC), che ha pubblicato la guida *Cite Datasets and Link to Publications*<sup>36</sup> che spiega come creare collegamenti tra le pubblicazioni di ricerca e i dati su cui si basano, così da rendere possibile la localizzazione dei *dataset* per chi legge un articolo e viceversa;

– la FAO (Food and Agriculture Organization of the United Nations) attraverso la piattaforma AIMS, ha reso disponibile la seconda versione delle raccomandazioni Linked Open Data-enabled Bibliographical Data (LODE-BD)<sup>37</sup> per supportare la selezione di appropriate strategie di codifica per l'implementazione di significativi LOD-enabled con cui descrivere risorse bibliografiche come articoli, monografie, tesi, relazioni di conferenze, *learning objects*;

– l'Università di Münster che coordina il progetto The Linked Open Data University of Münster (LODUM) entro cui è stata sviluppata un'infrastruttura per la pubblicazione di dati della medesima università in modalità LOD (figura 6), che favorisce lo scambio di dati scientifici e didattici in modalità LD;

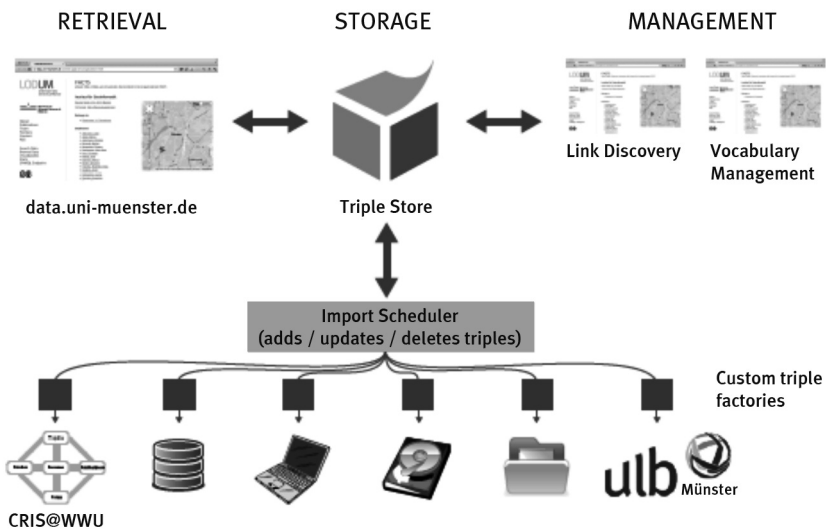


Figura 6. Panoramica dell'architettura LODUM<sup>38</sup>

<sup>36</sup> Digital Curation Centre, *Cite Datasets and Link to Publications*, <<http://www.dcc.ac.uk/resources/how-guides/cite-datasets>>.

<sup>37</sup> AIMS, *LODE-BD Recommendations 2.0 available on AIMS now!*, <<http://aims.fao.org/community/blogs/lode-bd-recommendations-20-available-aims-now>>.

<sup>38</sup> Carsten Keßler – Tomi Kauppinen, *Linked Open Data University of Münster. Infrastructure and Applications*, <<http://kauppinen.net/tomi/lodum-eswc-2012.pdf>>.

– la Library of Congress, il cui servizio Authorities and Vocabularies<sup>39</sup> ha reso disponibile in modalità LOD il sistema terminologico Semantic Web Tools for Libraries sviluppato presso la stessa biblioteca. L'attenzione verso LOD è ben espressa nell'iniziativa della biblioteca *A Bibliographic Framework for the Digital Age*<sup>40</sup>;

– la German National Library (DNB) che, in cooperazione con la German Serials Database (Zeitschriftendatenbank – ZDB), ha creato Linked Data Service<sup>41</sup>. Tale servizio esegue l'*authority data linking* dei record bibliografici disponibili<sup>42</sup> (inclusi *authority data* proposti dalle grandi agenzie come la Library of Congress e la British Library) con LD forniti da DBpedia e VIAF (Virtual International Authority File)<sup>43</sup> sotto la licenza aperta CCo<sup>44</sup>;

– la Hungarian National Library, che pubblica i propri dati bibliografici e quelli di autorità attraverso LD utilizzando gli schemi semantici come RDFDC, FOAF e SKOS<sup>45</sup>;

– la British National Library (BNB) che, oltre a fornire i propri dati con RDF download, li pubblica come LD collegandoli ai dataset esposti in LOD Cloud come VIAF, LCSH, Lexvo, GeoNames, MARC country, Dewey.info, RDF Book Mashup<sup>46</sup>;

– la Bibliothèque Nationale de France che, tramite il progetto Data.bnf.fr, mira a rendere i dati bibliografici relativi a opere e autori maggiormente integrati e fruibili sul Web attraverso le terne RDF, seguendo i principi di LOD<sup>47</sup>;

– la Biblioteca Nacional de España (BNE), che in collaborazione con l'Ontology Engineering Group (OEG) dell'Università Politecnica di Madrid (UPM) ha lanciato il progetto aperto Linked Data at the BNE<sup>48</sup>. Attraverso il tool MARiMba<sup>49</sup> i dati MARC21

**39** Library of Congress, *Library of Congress Authorities*, <<http://authorities.loc.gov/>>; *LC Linked Data Service Authorities and Vocabularies*, <<http://id.loc.gov/>>.

**40** Library of Congress, *A Bibliographic Framework for the Digital Age (October 31, 2011)*, <<http://www.loc.gov/marc/transition/news/framework-103111.html>>.

**41** Open bibliography and Open Bibliographic Data, *German National Library goes LOD & publishes National Bibliography*, <<http://openbiblio.net/2012/01/26/german-national-library-goes-lod-publishes-national-bibliography/>>.

**42** Attualmente ci sono più di 1.6 milioni di record bibliografici con l'identificatore per titoli seriali ZDB-ID che, a differenza di ISSN, distingue le versioni a stampa da quelle elettroniche, <<http://dispatch.opac.d-nb.de/DB=1.1/>>.

**43** Mirjam Keßler, *Linked Open Data of the German National Library*, in: *ECO4r Workshop LOD of DNB*; Berlin: Deutsche National Bibliothek, 2010, <[http://www.eco4r.org/workshop2010/eco4r\\_workshop2010\\_mirjam\\_kessler.pdf](http://www.eco4r.org/workshop2010/eco4r_workshop2010_mirjam_kessler.pdf)>.

**44** Secondo l'Open definition. V. Open Knowledge Foundation Italia, <<http://it.okfn.org/>>; CCo 1.0 Universal (CCo 1.0), <<http://creativecommons.org/publicdomain/zero/1.0/>>.

**45** ISKO UK, *Hungarian National Library OPAC and Digital Library Published as Linked Data*, 2010, <<http://iskouk.blogspot.com/2010/05/hungarian-national-library-opac-and.html>>.

**46** Neil Wilson, *Linked Data Prototyping at the British Library*, 2010, <<http://talis-linkeddata-libraries.s3.amazonaws.com/Linked%20Data%20Prototyping.pdf>>; British National Bibliography (BNB), *Linked Open Data*, <<http://thedatahub.org/dataset/bluk-bnb>>.

**47** Bibliothèque nationale de France, <<http://data.bnf.fr/>>; *Data.bnf.fr: an overall presentation*, <<http://data.bnf.fr/docs/databnf-presentation-en.pdf>>.

**48** Biblioteca Nacional de España, *Datos enlazados en la BNE*, <<http://www.bne.es/es/Inicio/Perfiles/Bibliotecarios/DatosEnlazados/index.html>>.

**49** MARiMba, <<http://mayor2.dia.fi.upm.es/oeg-upm/index.php/en/downloads/228-marimba>>.

e gli *authority data*<sup>50</sup> della BNE saranno trasformati in metadati RDF, pubblicati come LD sotto le licenze aperte CCo e interrogati con le *query SPARQL*;

– la Biblioteca di North Rhine-Westphalia, che ha lanciato il servizio LOD [lobid.org](http://lobid.org)<sup>51</sup> dove *lobid* sta per Linking Open Bibliographic Data. Il servizio è a supporto del processo della creazione di «Linked Open Bibliographic Data out of existing libraries and other associated data»;

– l'Online Computer Library Center (OCLC), che ha reso disponibile oltre un milione di file delle più importanti opere contenute nel catalogo digitale globale WorldCat, in modalità di download LD (LD è reso disponibile sotto la licenza ODC-BY e la serializzazione RDF è basata sull'ontologia Schema.org e sulle estensioni a tale ontologia dei dati bibliografici): «this is an important step for Libraries and Linked Data [...] Organizations wishing to develop Linked Data Services can experiment with this data set before going into full development. They'll also be able to stress-test new services using a very large and important set of up-to-date, Linked Library Data. We are really interested to see what people will do with this data [...] OCLC expects that the file will be useful as a source of raw data. Information about works, authors and publishers can be dissected and recombined in this format much more easily»<sup>52</sup>;

– la Bayerische Staatsbibliothek che, assieme ad altre biblioteche<sup>53</sup>, sta provvedendo all'allineamento dei propri metadati bibliografici con l'Europeana Data Model (EDM)<sup>54</sup>, costituito attraverso il riutilizzo di *namespaces* come RDF, RDFS, OAI-ORE, SKOS e DCMI Terms (figura 7) e basato sui meccanismi di LOD per rendere i medesimi dati apertamente consultabili e maggiormente interoperabili sul Web.



Figura 7. Il modello EDM

**50** Circa 2,5 milioni di record bibliografici (monografie antiche e moderne, registrazioni sonore) e 4 milioni di *authority record* sono stati trasformati in RDF. Il processo di trasformazione ha generato circa 58 milioni di terne RDF e 600 collegamenti (*owl:sameAs*) ad altri set di dati come DBpedia o VIAF, <<http://thedatahub.org/dataset/datos-bne-es>>.

**51** [lobid.org](http://lobid.org), <<http://lobid.org/about>>.

**52** OCLC, *OCLC provides downloadable linked data file for the 1 million most widely held works in WorldCat*, <<http://www.oclc.org/us/en/news/releases/2012/201252.htm>>. I contenuti in un unico file LD sono scaricabili da: <<http://purl.oclc.org/dataset/WorldCat/datadumps/WorldCatMostHighlyHeld-2012-05-15.nt.gz>>.

**53** Europeana Libraries, *Aggregating digital content from Europe's libraries. D5.1. Report on the alignment of library metadata with the European Data Model (EDM)*, <<http://www.europeana-libraries.eu/documents/868553/1eade085-34ac-487f-82af-d5cd2545e619>>; Bibliographic Data, <<http://thedatahub.org/group/bibliographic>>.

**54** Martin Doerr [et al.], *The Europeana Data Model (EDM)*, in: *World Library and Information Congress: 76th IFLA General Conference and Assembly. Open access to knowledge - promoting sustainable progress*, Gothenburg (Sweden), 2010, Meeting 149, *Information Technology, Cataloguing, Classification and Indexing with Knowledge Management*, <<http://conference.ifla.org/past/ifla76/149-doerr-en.pdf>>.

La qualità dei dati nel modello EDM dipende direttamente dalla presenza nel medesimo modello dei formalismi SKOS che rendono non solo disponibili ma anche interoperabili diversi dati dei sistemi di organizzazione della conoscenza (KOS) pubblicati sul Web, connettendoli ai dati pubblicati in LD, portando così benefici per entrambe le parti<sup>55</sup>.

Tra i servizi web potenziati dai meccanismi di LD e SKOS segnaliamo:

- il portale Europeana che raccoglie, converte in terminologie SKOS-based e pubblica in LOD<sup>56</sup> diversi KOS a supporto di oggetti culturali provenienti da un'ampia *range* di istituzioni (biblioteche, musei e archivi) che hanno sottoscritto il Data Exchange Agreement (2012), il quale prevede che tutti i metadati presenti in Europeana siano ad accesso aperto e corredati della licenza per il pubblico dominio. Nel settembre 2012 Europeana ha rilasciato i dati aperti (sotto la licenza CCo) di più di 20 milioni di oggetti culturali<sup>57</sup>;
- l'AGROVOC Thesaurus messo in linea di recente nell'ambito di un'iniziativa della FAO che incoraggia gli sviluppatori dei sistemi di gestione dell'informazione agricola ad incorporare AGROVOC nelle loro applicazioni attraverso servizi web, piuttosto che utilizzare copie locali del database<sup>58</sup>,
- il Thesaurus del Nuovo soggetto della Biblioteca nazionale centrale di Firenze<sup>59</sup>;
- il Thesaurus Earth (della Sezione di Genova dell'Istituto di matematica applicata e tecnologie informatiche)<sup>60</sup>.

### Library Linked Data: l'universo bibliografico verso il Web semantico

Affinché il materiale dell'universo bibliografico diventi onnipresente nell'attuale contesto informativo dominato dal Web, è sempre più importante la sfida di rielaborare le rappresentazioni di dati bibliografici attraverso metodi innovativi e il lavoro collaborativo di diverse comunità che propongono nuove soluzioni di pubblicazione, integrazione, *browsing* e *discovery* dei contenuti informativi nel contesto globale del Web dei dati.

Seguendo i lavori del W3C, in particolare del progetto LOD della SWEO community, nel 2011 il gruppo di lavoro W3C Library Linked Data (LLDXG)<sup>61</sup> - con lo scopo di sviluppare una standardizzazione condivisa del Web semantico all'inter-

55 «Libraries, museums, newspapers, government portals, enterprises, social networking applications, and other communities that manage large collections of books, historical artifacts, news reports, business glossaries, blog entries, and other items can now use Simple Knowledge Organization System (SKOS) to leverage the power of linked data», in: World Wide Web Consortium (W3C), SKOS Simple Knowledge Organization System, cit.

56 Europeana Linked Open Data (LOD), <<http://pro.europeana.eu/linked-open-data>>. Europeana LOD Pilot datasets comprende circa 2,5 milioni di record (Europeana conta più di 21 milioni record) sotto la licenza CCo.

57 Europeana, *Business Plan 2012*, <<http://pro.europeana.eu/documents/858566/coc6e31f-5174-4898-9771-f9b9a8d1d4d7>>.

58 AGROVOC, <<http://aims.fao.org/standards/agrovoc/about>>.

59 Marta Motta - Dario Rodighiero, *Il Thesaurus del Nuovo soggetto interpreta SKOS*, 2010, <<http://www.iasummit.it/2010/papers/motta-rodighiero.pdf>>.

60 EARTH, <<http://uta.iaa.cnr.it/earth.htm>>.

61 Presentazione dei lavori di LLDXG su *Library science talk*, 12-13 marzo 2012, <[http://www.nb.admin.ch/aktuelles/ausstellungen\\_und\\_veranstaltungen/00726/01611/03953/03958/index.html?lang=en](http://www.nb.admin.ch/aktuelles/ausstellungen_und_veranstaltungen/00726/01611/03953/03958/index.html?lang=en)>.

no della comunità bibliotecaria – ha pubblicato la relazione finale sui Library Linked Data (LLD), e i documenti di approfondimento come *Datasets*, *Value Vocabularies*, *Metadata Element Sets* proiettati su alcuni *Use cases*<sup>62</sup> (figura 8).

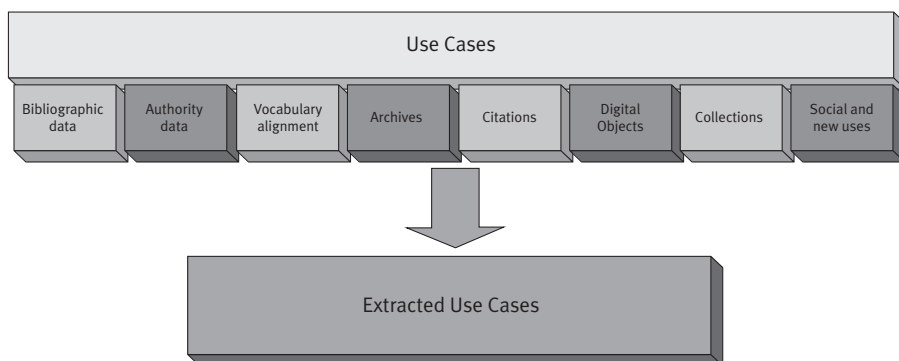


Figura 8. Casi d'uso dell'applicazione di LD<sup>63</sup>

Lo scopo di LLD, secondo la definizione del gruppo LLDXG, è quello di aumentare l'interoperabilità globale dei *dataset* bibliografici sul Web (incluso quelli dei vocabolari controllati), coinvolgendo esperienze e *know-how* di diverse comunità (biblioteche nazionali e di università, centri di ricerca, musei, archivi<sup>64</sup>) che promuovono le esperienze di integrazione dei dati bibliografici nel Web semantico, pubblicandoli sotto CKAN Registry (Comprehensive Knowledge Archive Network, un catalogo di *dataset* aperto) *MetaData Conventions*<sup>65</sup>. La versione grafica di diversi pacchetti di dati collegati entro lo spazio LLD (figura 9) – elaborata dal gruppo CKAN LLD<sup>66</sup> – anche se ancora non è stata effettivamente tradotta in un Linked Data service, rappresenta la programmazione di un vero e proprio percorso di LLD verso LOD Cloud<sup>67</sup>.

<sup>62</sup> Library Linked Data Incubator Group, *Final Report*, <<http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/>>; *Library Linked Data Incubator Group: Use Cases*, <<http://www.w3.org/2005/Incubator/lld/XGR-lld-usecase-20111025/>>.

<sup>63</sup> World Wide Web Consortium (W3C), *Use Cases*, <[http://www.w3.org/2005/Incubator/lld/wiki/Use\\_Cases](http://www.w3.org/2005/Incubator/lld/wiki/Use_Cases)>.

<sup>64</sup> Library Linked Data Incubator Group, *Charter*, <<http://www.w3.org/2005/Incubator/lld/charter>>.

<sup>65</sup> World Wide Web Consortium (W3C), *Linked Library Data - CKAN MetaData Conventions*, <<http://www.w3.org/wiki/TaskForces/CommunityProjects/LinkedLibraryData/Datasets/CKAN-metainformation>>.

<sup>66</sup> Library Linked Data, <<http://thedatahub.org/group/lld>>.

<sup>67</sup> World Wide Web Consortium (W3C), *Guidelines for Collecting Metadata on Linked Datasets in CKAN*, <<http://www.w3.org/wiki/TaskForces/CommunityProjects/LinkingOpenData/DataSets/CKAN-metainformation>>.



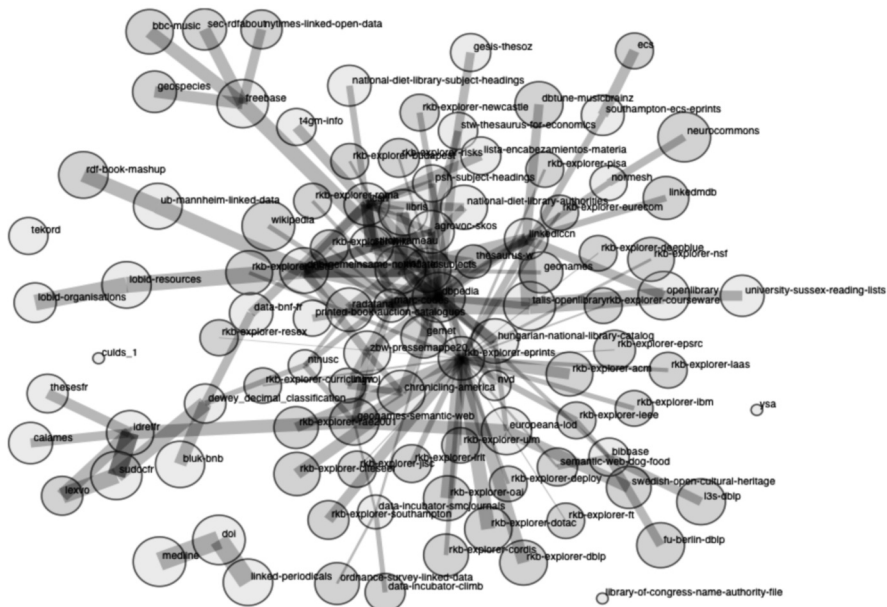


Figura 9. Rappresentazione grafica di LLD (in stato di aggiornamento)<sup>68</sup>

Mentre i cerchi colorati rappresentano i pacchetti di dati sviluppati entro il dominio LLD, i cerchi grigi rappresentano i pacchetti di dati collegati a LLD dall'esterno (ad esempio, da DBpedia, GeoNames di LOD Cloud). Le dimensioni dei cerchi e lo spessore delle linee riflettono le dimensioni dei pacchetti e il numero di collegamenti esterni.

I *dataset* LLD potranno diventare i maggiori fornitori di *authoritative datasets* (persone, soggetti)<sup>69</sup> nello spazio del Web semantico e potranno essere facilmente arricchiti (*authority data enrichment*)<sup>70</sup> da record di autorità, pubblicati come LOD da altre comunità operanti sullo stesso Web<sup>71</sup>. Con riferimento all'ambiente digitale, è evidente l'importanza di controllare con gli strumenti di *authority control* la varietà di raccolte (coesistenti in tanti formati documentari) alle quali le biblioteche forniranno l'accesso via Web e che non gestiranno fisicamente, in particolare per quanto riguarda i collegamenti delle descrizioni a persone fisiche ed enti collettivi, e l'individuazione del ruolo che questi soggetti hanno avuto nella creazione, produzione, diffusione e conservazione dei singoli documenti. Inoltre, integrandosi nel Web semantico, i *dataset* bibliografici potranno interagire con diverse ontolo-

<sup>68</sup> Library Linked Data Incubator Group, *Datasets, Value Vocabularies, and Metadata Element Sets*, <<http://www.w3.org/2005/Incubator/lld/XGR-lld-vocabdataset-20111025/>>.

<sup>69</sup> NKOS, *The 10th European Networked Knowledge Organisation Systems (NKOS) Workshop*, <<https://www.comp.glam.ac.uk/pages/research/hypermedia/nkos/nkos2011/>>; Gordon Dunsire, *Linked Data, Libraries and the Semantic Web* cit.

<sup>70</sup> World Wide Web Consortium (W3C), *Use Case Authority Data Enrichment*, cit.

<sup>71</sup> Library Linked Data Incubator Group, *Final Report. Appendix C: Semantic alignment*, <[http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/#Semantic\\_alignment](http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/#Semantic_alignment)>.



gie web (ad esempio SKOS, Library of Congress Name Authority File, BIBO, Lexvo.org ontology, CC REL, OPM, UMBEL vocabulary, VIAF, FOAF, DCMI Collection Description Community, FRBR WEMI ontology, NSTL bibliography ontology), attingendo automaticamente da esse le descrizioni per proprietà e relazioni che ulteriormente arricchiranno il loro profilo semantico. Generando i ricchi *cluster* di entità, di proprietà descrittive e normalizzatrici, nonché di collegamenti attraverso i meccanismi di LD, sarà anche possibile elaborare varie *query* sofisticate per interrogare le fonti bibliografiche integrate nel Web.

Nonostante gli indiscutibili vantaggi di LLD<sup>72</sup>, per i numerosi progetti che promuovono i metodi innovativi di gestione dei dati con le tecnologie del Web semantico sarà necessario un coordinamento specifico, accurato e deciso, al fine di contrastare il fenomeno dell'implementazione irregolare di tecnologie e principi LD (nei contesti di LOD e LLD) da parte di diverse comunità di utenti web.

Tra i principali ostacoli alla pubblicazione dei dati con i meccanismi LD vi è ancora la riluttanza, da parte di diverse comunità, nell'adottare nuove piattaforme software, così come la mancanza del personale con adeguate conoscenze e competenze per utilizzare queste piattaforme, prima e dopo la loro implementazione<sup>73</sup>. Se i bibliotecari, esperti in catalogazione, nell'elaborazione e nella gestione di metadati si potrebbero sentire sopraffatti dalla prospettiva di produzione di metadati semantici e di ontologie, allora lo sarebbero ancora di più altre comunità operanti sul Web. Inoltre, ci sono tensioni tra *open standards*, *open publishing* richiesti per il Web semantico e i metodi esistenti per il supporto e il mantenimento delle infrastrutture di gestione dati. Da un punto di vista positivo, si potrebbe imitare le esperienze di alcune istituzioni (delle quali abbiamo dato un accenno nel paragrafo precedente) che hanno iniziato a pubblicare i loro dati bibliografici con tecnologie LD.

In quanto alle conclusioni del rapporto LLD, esse sono le seguenti:

- i dati bibliografici non sono integrati con le risorse web;
- gli standard bibliotecari sono progettati solo per la comunità bibliotecaria;
- i dati bibliografici sono espressi principalmente nel linguaggio naturale testuale;
- la comunità bibliotecaria e quella del Web semantico hanno una terminologia diversa per concetti di metadati simili;
- i cambiamenti tecnologici della biblioteca dipendono dallo sviluppo dei principali venditori dei sistemi informativi digitali.

Per contrastare le difficoltà di integrazione di dati bibliografici nel Web semantico, la relazione finale LLDXG ha fornito quattro principali raccomandazioni, con cui: 1) invita le comunità bibliotecarie a individuare pacchetti di dati (inclusi thesauri e altri vocabolari controllati) da pubblicare in modalità LOD e in linguaggi standard del Web semantico (ad esempio SKOS), così da aprire le porte alla creazione di nuovi e potenzialmente illimitati servizi; invita, dunque, a diffondere quanto più possibile le proprie esperienze riguardo al *free, open data*<sup>74</sup> attraverso licenze aperte e formati non proprietari. Nella relazione vengono individuate tre potenziali categorie di dati bibliografici da pubblicare in LD e gestire con gli appositi *namespace*<sup>75</sup>:

72 Gillian Byrne – Lisa Goddard, *The Strongest Link: Libraries and Linked Data* cit.

73 Joab Jackson, *Tim Berners-Lee: Machine-readable Web still a ways off*, in: CGN, October 2009, <<http://gcn.com/articles/2009/10/30/berners-lee-semantic-web.aspx>>.

74 Linked Open Data Italia, *Linked Open Data: dati aperti collegati e usabili*, <<http://www.linkedopendata.it/>>.

75 Entro la British National Bibliography (BNB) pubblicata come LD dalla British Library vengono gestiti diversi *namespace* (MARC, VIAF, ISBD, DC, LCSH, Lexvo, GeoNames ecc.), <<http://thedatahub.org/it/dataset/bluk-bnb>>.

- a) *dataset* formanti le terne su risorse bibliografiche, persone, organizzazioni e soggetti connessi,
  - b) vocabolari controllati da cui possono derivare le terne formate da valori, proprietà e attributi per *authority file*,
  - c) formati e vocabolari (schemi) di metadati che generano le terne descrittivi attributi e relazioni usati entro *dataset* bibliografici e vocabolari controllati;
- 2) si appella agli enti<sup>76</sup> che producono gli standard per incrementare la partecipazione delle biblioteche nella standardizzazione del Web semantico e nello sviluppo di standard per i dati bibliografici compatibili con LD, e per disseminare *best practices* sulla modellazione di LLD; si appella, poi, per influenzare i *vendor* (le aziende commerciali fornitrici di risorse elettroniche e database online) affinché vengano considerati e adottati questi standard;
- 3) si rivolge agli sviluppatori dei sistemi informativi digitali affinché essi:
- a) progettino *high-level services* di gestione dei dati basati sulle capacità del Web semantico e di LD,
  - b) creino URIs (anche in modo sperimentale) per poter puntare agli elementi di *dataset* bibliografici e a dati prodotti da altre comunità in modo da definire un contesto,
  - c) sviluppino politiche per la gestione dei vocabolari RDF e di relativi URI,
  - d) esprimano i *dataset* bibliografici in termini di loro riuso o mappatura in collegamento a esistenti vocabolari pubblicati in LD e adeguati al contesto di riferimento;
- 4) invita le comunità di bibliotecari e archivisti a preservare i *dataset* e i vocabolari controllati in set di elementi LD e ad applicare i loro *know-how*, riferiti alla gestione dati e alla conservazione a lungo termine, ai *dataset* pubblicati in LD così che le terne di LD diventino immutabili, gli identificatori univoci URIs rimangano permanenti e nomi, titoli, intestazioni per soggetto permangano stabili nel tempo. Con il progredire delle tecnologie del Web semantico, le biblioteche e le istituzioni archivistiche avranno una buona opportunità per arricchire i loro attuali ruoli di gestori dell'universo bibliografico con la qualifica di autorità garanti<sup>77</sup> della conservazione a lungo termine dei *dataset* in LD relativi al patrimonio culturale.

## Conclusioni

Il primo passo verso un Web che contenga dati bibliografici interpretabili nel loro significato anche dalle macchine è sicuramente quello di adottare un formato standard aperto come RDF, utilizzabile per raccogliere e categorizzare le informazioni sparse nel mondo della rete. RDF renderà disponibili i dati nella loro forma più pura (*raw data*<sup>78</sup>, dati puri senza formati e formattazioni), che consentirà di strutturarli e connetterli attraverso la tecnologia Linked Data in modo omogeneo entro lo spazio del *Web of data*. «Upcom-

<sup>76</sup> IFLA (Namespace Technical Group reposting to Committee on Standards); JSC for Development of RDA (DCMI/RDA Task Group, Dublin Core Metadata Initiative); DCMI Bibliographic Metadata Task Group (DCMI Vocabulary Management Community).

<sup>77</sup> «Libraries and librarians have a very high “trust” factor with the general public, and we are in a position to raise the general quality of the Semantic Web by ensuring Library data competing with data from the widest range of sources», Gordon Dunsire, *Linked Data, Libraries and the Semantic Web* cit.

<sup>78</sup> Vedi *Tim Berners-Lee on the next Web* in: TED (Technology, Entertainment, Design) conference 2009, <[http://www.ted.com/talks/tim\\_berniers\\_lee\\_on\\_the\\_next\\_web.html](http://www.ted.com/talks/tim_berniers_lee_on_the_next_web.html)>.

ing challenges will involve the strategic and technical alignment of catalogues and legacy systems in libraries and the authoring environments for scholarly communication - both in a data and service infrastructure based on the principles of the Semantic Web»<sup>79</sup>.

Un investimento finanziario e in *know-how* degli esperti di diversi settori che possa accrescere la qualità di architetture e tecnologie semantiche di LD, genererà di sicuro un ritorno positivo a medio e lungo termine nell'armonizzazione di diversi formati di *Web-based data* (provenienti da vari database relazionali), nel loro riutilizzo da parte diversi sistemi informativi e comunità web e nella creazione di un nuovo e potente strumento di ricerca e reperimento di informazioni attraverso le *query* mirate.

La sfida per i gestori dell'universo bibliografico, quindi, è quella di approfondire la conoscenza su LD, conoscere tutti i suoi vantaggi, leggere con sguardo laico i suoi strumenti, e promuovere la conoscenza e il consapevole utilizzo di Library Linked Data, tema che «non può non interessare la comunità bibliotecaria nel suo complesso»<sup>80</sup>, il che significa che ci saranno motivi maggiormente utili a far crescere il livello di riutilizzo e di interoperabilità dei dati, secondo la ormai famosa scala delle stelline promossa da Sir Tim Berners-Lee<sup>81</sup>.

**ABSTRACT** AIB studi, DOI 10.2426/aibstudi-8624, vol. 52 n. 3 (settembre/dicembre 2012), p. 325-343.

IRYNA SOLODOVNIK, Scuola dottorale internazionale degli studi umanistici (SDISU), Università della Calabria, via P. Bucci, Cubo 28/B, 87036 Arcavacata di Rende (CS), e-mail iryna.solodovnik@gmail.com.

#### **Uno sguardo sul futuro semantico dell'universo bibliografico**

Il Web semantico è progettato in modo da consentire non solo ai programmi informatici di implementare da un punto di vista semantico raccolte di dati già esistenti, ma anche alle diverse comunità di utenti di connettere, condividere e comprendere in modo semplice e veloce informazioni strutturate e semi-strutturate. Il movimento *Linked Data* risponde allo sforzo del Web semantico di sviluppare buone pratiche per la pubblicazione e la connessione di dati strutturati sul Web. Come possono le biblioteche sfruttare quest'opportunità e allo stesso tempo affrontare le sfide che si preannunciano per il futuro semantico dell'universo bibliografico? Come fare in modo che *Linked Data* e *Open Data* non si riducano a uno slogan effimero ma guadagnino terreno nello scenario delle future infrastrutture digitali della biblioteca? Obiettivo principale dell'articolo è analizzare alcune questioni teoriche e pratiche diffuse nell'ambiente del web semantico, mettendone in luce vantaggi e svantaggi in relazione all'universo bibliografico.

**79** Presentazione della conferenza *SWIB12 Semantic Web in Libraries*, 26-28 novembre 2012, <<http://swib.org/swib12/>>.

**80** Carlo Bianchini, *Il Library Linked Data Incubator Group e il futuro delle biblioteche*, 2012, <<http://sites.google.com/site/homepagecarlobianchini/Ricerca/library-linked-data-e-il-futuro-delle-biblioteche>>.

**81** Tim Berners-Lee, *The 5 stars of open linked data*, <<http://inkdroid.org/journal/2010/06/04/the-5-stars-of-open-linked-data/>>.

**A look at the semantic future of the bibliographic universe**

Semantic Web is designed to enable computer programs to enrich semantically big data solutions, as well as to connect, share and understand structured and semi-structured information easily by different user agents. The Linked Data movement is an effort of Semantic Web to develop best practices for publishing and connecting structured data on the Web. How could libraries take advantage of these opportunities and how can they handle the challenges of the semantic future of the bibliographic universe? How to guarantee that Linked Data and Open Data won't be a temporary hype but that they will gain ground in the scenery of future digital library infrastructures? The main objective of this paper is to survey some theory and practical issues used in Semantic Web environment and summarize some of their advantages and drawbacks in dealing with the bibliographic universe.