

The Semantic Web and Libraries in the United States: Experimentation and Achievements

Sharon Yang, Associate Professor-Librarian, Rider University, USA
Yan Yi Lee, Systems Librarian, Wagner College, USA
Amanda Xu, Assistant Professor-Librarian, St. John's University, USA

(Special Thanks to Dorothy Davison, Dean of the Library, Wagner College
Cathy Perkins, Web/Instruction Librarian, Wagner College)

Abstract: This paper starts by introducing the concept of Semantic Web, related standards and technologies, and the Semantic Stack of Sir Tim Berners-Lee. Next it explains the possible roles that libraries can play in semantic Web development. Over 40 Semantic Web projects have been reviewed and listed. The projects from the U.S. libraries are described in detail, including those by LC, OCLC, and academic libraries. In comparison, there are some visible differences in library-related semantic development and orientation between the U.S. and European libraries, and those differences are fully discussed. Increased globalization is taking place for some semantic projects. One of the conclusions in this paper is that library-related Web of Data is ready to be exposed to Semantic Web application development.

Keywords: Semantic Web; semantic projects in libraries; digital library projects;

INTRODUCTION

The Web has progressively become a global information space where both linked documents and data become deeply interwoven and navigable. Semantic Web, as defined by Sir Tim Berners-Lee and others, is an extension of the current Web (Semantic Web, 2009). It involves machine processing and machine actionable data. These two are interoperable for publishing, entity resolution, linking, interchange, explanatory searching, and connecting people, communities, things, and knowledge on the Web with meta-languages: Resource Description Framework (RDF)¹, RDF Vocabulary Description Language (RDFS)², Web Ontology Language (OWL)³, Simple Knowledge Organization Systems (SKOS)⁴, RDFa⁵, and their query language and data access protocol (Dodds, 2005): SPARQL⁶.

¹ <http://www.w3.org/RDF/>

² <http://www.w3.org/TR/rdf-schema/>

³ <http://www.w3.org/2004/OWL/>

⁴ <http://www.w3.org/TR/skos-primer/>

⁵ <http://www.w3.org/TR/rdfa-syntax/>

⁶ <http://www.w3.org/TR/rdf-sparql-query/>

HTML publishes documents and provides hypertext links between them. RDF, RDFS, OWL, SKOS, and RDFa, by contrast, can describe arbitrary things such as people, concepts, events or objects on the Web. As a result, Semantic Web is known as a Web of Data, where data can not only be uniquely identified as each resource having its own URI distributed on the Web, but also securely shared and reused across the boundaries of applications, enterprises and communities over HTTP-based protocols and other Internet technologies using explicit metadata derived from Semantic Web standards. In addition, the Web of Data can be related for both humans and machines to make intelligent inferences based on connections represented in the relations and underlying logic and rules. Sir Tim Berners-Lee called the resulting network of Linked Data (Berners-Lee, 2007), the Giant Global Graph (Giant Global Graph, 2009), in contrast to the HTML-based World Wide Web⁷.

SEMANTIC STACK

The W3C architecture vision for Semantic Web Technologies is commonly known as ‘Semantic Stack’ or layer cake (Berners-Lee, 2000), which loosely consists of the following seven layers:

1. The URI (Uniform Resource Identifier including URL and URN) layer is for naming things and information resources, and providing means to access information about them over the Web. The assumption is that each web resource, including concepts from the real world – people, organizations, topics, and things, can have their own URIs. When someone sends a request to look up a URI or its related URIs, the representation of the identified resources can be returned in response to that HTTP request. According to Dave Beckett, URIs benefit to the Web with “extensibility, unbounded identifier space, easy data merging and aggregation, and semantics can change but the tools still work (Beckett, 2001).”
2. The XML/XML Schema layer is for content markup and syntax validation using user-defined vocabulary and data types. It enables the serialization of RDF statements.
3. The RDF/RDF schema layer uses the basic data model⁸ to describe things, classes, properties, sub-classes, sub-properties, domain and range restrictions, etc. by representing them in two ways: *i*) RDF triples, which is also known as RDF statement which consists of subject, predicate (or property) and object (value), or *ii*) RDF graphs; and then serializing them in N-Triples, RDF/XML, Turtle, RDF/JSON, and others⁹.

⁷ http://en.wikipedia.org/wiki/World_Wide_Web

⁸ http://en.wikipedia.org/wiki/Data_model

⁹ <http://arc.semsol.org/features>

Each component of an RDF statement is a resource. According to Chris Bizer, “the subject of a triple is the URI identifying the described resource. The object can either be a simple literal value, like a string, number, or date, or the URI of another resource that is somehow related to the subject. The predicate, in the middle, indicates what kind of relation exists between subject and object. These predicate URIs come from vocabularies, collections of URIs that can be used to represent information about a certain domain (Bizer, Cyganiak & Heath, 2007).”

A set of RDF triples is also known as RDF graph because the URIs occurring as subject and object URIs are the nodes in the graph, which are connected directly by an arc. For full machine-processing, an RDF statement has to be a set of three URLs (Dunsire, 2008). RDF Schema is a vocabulary for describing properties and classes of RDF-based resources, with semantics for generalized-hierarchies of such properties and classes.

SPARQL is the query language for “diverse data sources stored natively as RDF or viewed as RDF in the middleware (W3C, 2007).” There are three specifications dealing with SPARQL query language, results format, and data access protocol. For SPARQL query language, there are four basic forms of query construction, e.g. ‘Select’, ‘Construct’, ‘Ask’, and ‘Describe’. Some SPARQL query tools offer extension to the query specification, e.g. SPARQL⁺ from ARC2¹⁰. In the SPARQL results format, XML format is used to bind variables and Boolean results.¹¹ “Result sets can not only be accessed by a local API but also can be serialized into XML, an RDF graph (W3C, 2006-2007)” and others. SPARQL protocol¹² uses WSDL 2.0 to describe a means for conveying SPARQL queries to a SPARQL query processing service and returning the query results to the entity that requested them.

4. The ontology layer is for explicit meta-information, complex-relationship representation, and formal specification of a conceptualization, e.g. transitive properties, value restrictions, uniqueness, cardinality (e.g. “exactly one”), etc. According to Ivan Herman from W3C, the value of semantifying the Web is in the “combination of data (Herman, 2009)”. Vocabularies encoded in OWL and SKOS promote linked data with rich typing of properties, property characteristics (e.g. symmetry), and enumerated classes.

Like SKOS, other common vocabularies that use RDF and OWL for expression are: *i*) FOAF¹³ (Friend of a Friend) - a machine-readable

¹⁰ <http://arc.semsol.org/docs/v2/sparql+>

¹¹ <http://www.w3.org/TR/rdf-sparql-XMLres/>

¹² <http://www.w3.org/TR/rdf-sparql-protocol/>

¹³ <http://xmlns.com/foaf/spec/>

ontology describing persons, their activities and their relations to other people and objects; *ii*) SIOC¹⁴ (Semantically-Interlinked Online Communities) - a machine-readable ontology expressing the information contained both explicitly and implicitly in discussion forums on the Web, e.g. blogs, forums, and mailing lists; *iii*) Dublin Core¹⁵ - a set of metadata elements for cross-domain information description.

Unlike common vocabularies, Semantic Wiki is “a wiki that has an underlying model of the knowledge described in its pages, available in a formal language so that machines can process it into entity-relationship or relational database. Semantic wiki systems differ in the level of ontology support they offer. While most systems can export their data as RDF, some even support various levels of ontology reasoning (Semantic wiki, 2009).” Semantic MediaWiki is “an extension to MediaWiki that allows for annotating semantic data within wiki pages, thus turning a wiki that incorporates the extension into a semantic wiki (Semantic MediaWiki, 2009).”

Data encoded using the vocabularies can be used in *i*) semantic searches and alerts¹⁶ with feature topics or discussions selected from Twitter¹⁷, Internet Relay Chat (e.g. IRC-Galleria¹⁸), data feed (e.g. Moreover.com¹⁹) or other social networking sources; *ii*) aggregation of pages (e.g. Morph Semantic Converter²⁰); *iii*) display in formats like maps, calendars and graphs, address books, and exported to formats like RDF and CSV for data applications.

5. Logic layer is for reasoning, etc.;
6. Proof layer is for validation; and
7. Trust via digital signatures and other knowledge layer is for recommendation, rating, certification, etc.

The key technologies of the semantic stack have been widely adopted by most semantic Web applications, and mature enough to find a large amount of RDF data which can be accessed over the Web as Linked Data²¹. If we regard cool URIs (Sauermann, Cyganiak, Ayers, & Völkel, 2008) as the premises for linked data, then RDF/RDFS, OWL/SKOS, RDFa, and SPARQL are the W3C standards and vocabularies representing the data for machine-processing in the form of

¹⁴ <http://sioc-project.org/>

¹⁵ <http://dublincore.org/>

¹⁶ <http://bnode.org/blog/2009/04/22/semantic-web-apps-to-simplify-my-life>

¹⁷ <http://en.wikipedia.org/wiki/Twitter>

¹⁸ <http://en.wikipedia.org/wiki/IRC-Galleria>

¹⁹ <http://w.moreover.com/>

²⁰ <http://convert.test.talis.com/>

²¹ <http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData/>

data model, ontology, a set of XHTML attributes, and query language respectively.

Open Ontology Registry²² and National Science Digital Library Metadata Registry²³ are examples of the best practices for managing vocabularies in a database, outputting metadata as RDFS/SKOS and OWL, and automatically creating a URI for each entry. Authority and vocabulary services from Library of Congress, National Library of Medicine, National Library of Agriculture, OCLC WorldCat Identities, OCLC xISBN Web Services, and others serve as resolvers for library-related semantic applications. They can enable applications to make the appropriate choice of the form of entries within the information space that they are linking, weaving or navigating.

Semantic-aware tools such as Google's RDFa support²⁴, Yahoo's Pipes²⁵, Firefox's extensions for Piggy Bank²⁶, FOAFfox²⁷, and Zotero²⁸ in Web browser applications; Twinkle²⁹ in Jena and ARC³⁰ in XAMP for SPARQL query processing in desktop applications; Protégé³¹ and TopBraid³² for ontology editors; Oracle 11g's RDF database³³ for managing RDF triples in RDF store; Vulcan's Semantic Wiki for Project Halo³⁴, Knoodl's OWL-based ontologies and RDF knowledgebases³⁵; Hakia³⁶, Freebase Parallax³⁷, and Semantic Engines' SenseBot³⁸ for semantic searching, and open Semantic Web framework like Drupal's Content Management Systems, Semol's ARC and Trice³⁹, Twine⁴⁰, Jena⁴¹, Sesame⁴², and Ontotext's OWLIM⁴³ are examples of the enablers for Semantic Web technologies for individuals, communities, and RESTful Web services (Fielding, 2002).

The top layers of the semantic stack, e.g. logic, proof and trust are still under development even though the framework of the technologies has already been

²² <http://ontolog.cim3.net/cgi-bin/wiki.pl?OpenOntologyRepository>

²³ <http://meta-dataregistry.org/>

²⁴ <http://google.com/support/webmasters/bin/answer.py?answer=146898>

²⁵ <http://pipes.yahoo.com/pipes/>

²⁶ http://simile.mit.edu/wiki/Piggy_Bank

²⁷ <http://www.aidanf.net/semantic-web-extensions-for-firefox>

²⁸ <http://www.zotero.org/support/>

²⁹ <http://www.ldodds.com/projects/twinkle/>

³⁰ <http://arc.semsol.org/home>

³¹ <http://protege.stanford.edu/>

³² http://www.topquadrant.com/products/TB_Composer.html

³³ http://www.oracle.com/technology/tech/semantic_technologies/index.html

³⁴ <http://www.ontoprise.de/en/home/news/news-en/vulcan-smw/>

³⁵ <http://knoodl.com/ui/home.html>

³⁶ <http://www.hakia.com>

³⁷ <http://mqlx.com/~david/parallax/>

³⁸ <http://www.sensebot.net/>

³⁹ <http://semsol.com/>

⁴⁰ <http://www.twine.com/>

⁴¹ <http://jena.sourceforge.net/>

⁴² <http://www.openrdf.org/>

⁴³ <http://www.ontotext.com/owlim/>

defined for nearly 10 years. The level of support for the top layers of the semantic stack is also tool-dependent. For instance, NG4J is the named graphs API for Jena, which claims to have an extension to support digital signature (Bizer, Cyganiak, & Hartig, 2009).

THE LIBRARY'S ROLE IN SEMANTIC WEB

Even though the building blocks of Semantic Web standards and technologies are similar to the library catalog services for organizing and sharing the Web of data, they will provide new opportunities for libraries. Eric Miller, W3C Semantic Web Activity Lead, defined the library's role as follows:

- Exposing collections – use Semantic Web technologies to make contents available
- Web'ifying Thesaurus/Mappings/Services
- Sharing lessons learned
- Persistence (Harper & Tillett, 2007)

By “Exposing collections”, Miller means enabling semantic searching of library collections and scholarly contents on the Internet. “Web'ifying thesaurus/mappings/services” involves converting controlled vocabularies and thesaurus into Semantic Web standards. Librarians have been engaged in developing terms and concepts to describe contents for so long that “sharing lessons learned” with others is important. “Persistence” is what libraries and the semantic researchers need in order to make semantic linking a reality on the Web.

Semantic Web development for libraries can be summarized as three developmental stages. The first stage is to create semantic tools for the following:

- i)* Weaving Semantic Wiki and Semantic Media Wiki;
- ii)* Translating LC controlled vocabularies, authority control for named entities, thesauri from domain specific societies and institutions (e.g. IEEE, ACM, Getty, NLM, Data Harmony), and others into Web Ontology Language (OWL) or Simple Knowledge Organization System (SKOS) with URIs assigned according to Sir Tim Berners-Lee's Linked Data Design Principles (Berners-Lee, 2007);
- iii)* Converting semantic-aware data sets in MARCXML, RDFa, and DBPedia into RDF triples in conformance with content models defined in FRBR, RDA, and Dublin Core, and registering them with global metadata registries such as National Science Digital Library Registry⁴⁴, Dublin Core

⁴⁴ <http://meta-dataregistry.org/>

- Metadata Registry⁴⁵, and others with extensive use of URIs for individual data elements in bibliographic records;
- iv)** Increasing explicit correlation and referencing between LCSH terms, and LCC and DDC numbers in lists of new LCSH headings and LCC numbers (Marcum, 2008);
 - v)** Developing built-in links to thesauri with Microsoft Office tools (Marcum, 2008);
 - vi)** Supporting SPARQL endpoint for both querying and analyzing the data sets; and
 - vii)** Building business rules and knowledge bases for web-based services for information extraction, identity resolution, filtering, semantic annotation and search.

The second stage is to make contents available as linked data with the tools created. This will support semantic technology infrastructure for:

- i)** Bibliographic data discovery, sharing and reuse, e.g. OCLC WorldCat Identities⁴⁶ and Talis Connected Commons⁴⁷;
- ii)** Authority data discovery, sharing and reuse, e.g. LC Authorities & Vocabularies⁴⁸, OCLC's Faceted Application of Subject Terminology (FAST); Multilingual Access to Subjects (MACS)⁴⁹ sponsored by the Conference of European National Librarians; LC ILS support of subject terms from 16 thesauri in addition to LCSH for CIP records (Marcum, 2008, p. 63)⁵⁰;
- iii)** Explanatory search, e.g. OCLC WorldCat Local⁵¹; Scriblio⁵² (formerly known as WPopac) at Plymouth State University; Endeaca implementation at NCSU Libraries⁵³, and Solr-based search tools, e.g. Vufind⁵⁴, Primo⁵⁵, and Blacklight⁵⁶;
- iv)** Community-based collaboration and social networking, e.g. FRBR Blog⁵⁷, PennTags⁵⁸, LibGuides' Widget, Facebook Apps⁵⁹, Slideshare in

⁴⁵ <http://demi.kc.tsukuba.ac.jp/dcregistry/>

⁴⁶ <http://www.worldcat.org/identities/>

⁴⁷ <http://blogs.talis.com/n2/cc>

⁴⁸ <http://id.loc.gov/authorities/>

⁴⁹ <https://macs.hoppie.nl/pub/>

⁵⁰ <http://www.loc.gov/bibliographic-future/news/LCWGResponse-Marcum-Final-061008.pdf>

⁵¹ <http://www.oclc.org/worldcatlocal/>

⁵² <http://about.scriblio.net/>

⁵³ <http://www.lib.ncsu.edu/endeaca/>

⁵⁴ <http://vufind.org/>

⁵⁵ <http://www.exlibrisgroup.com/category/PrimoOverview>

⁵⁶ <http://blacklight.rubyforge.org/>

⁵⁷ <http://www.frbr.org/>

⁵⁸ <http://larissah.wordpress.com/2007/12/06/penntaggs/>

⁵⁹ <http://www.springshare.com/libguides/widgets.html>

- Twitter⁶⁰, Academia.edu⁶¹, GetGlue⁶², and Scientific Collaboration Framework⁶³;
- v) Preservation and archiving, e.g. Dspace⁶⁴, Fedora Commons⁶⁵, Greenstone⁶⁶, arXiv⁶⁷, OAI-PMH⁶⁸, Cheshire III⁶⁹, ePrints.org⁷⁰, and OCLC's ArchiveGrid⁷¹;
 - vi) Distributed content management systems for contents of multiple arrays of formats, especially those that are media rich in nature, e.g. LC's MIC (Moving Images Collections), Drupal Core and Site Vocabulary⁷², YouTube⁷³, Flickr⁷⁴, Vimeo⁷⁵, and Blinx⁷⁶;
 - vii) Information dissemination, e.g. e-mail alerts from LinkedIn Groups and Meetup groups⁷⁷;
 - viii) Information visualization, e.g. Grokker⁷⁸; and
 - ix) Intelligent information analysis and decision support, e.g. automatic creation of some report like OpenDOAR and ROAR RSP Service Day⁷⁹.

The third stage is to support any library-related functions that future Semantic Web development may call for, for instance, **linking** authority data to DBPedia⁸⁰, GeoNames⁸¹, public data in Amazon Web Services⁸², Internet Archives⁸³, LibraryThing⁸⁴, and **working** in the trust layer of the semantic stack, including semantifying Web services in review and rating things for library-related

⁶⁰ <http://twitter.com/slideshare>

⁶¹ <http://www.academia.edu/>

⁶² <http://getglue.com/>

⁶³ <http://sciencecollaboration.org>

⁶⁴ <http://www.dspace.org/>

⁶⁵ <http://www.fedora-commons.org/>

⁶⁶ <http://www.greenstone.org/>

⁶⁷ <http://arxiv.org/>

⁶⁸ <http://www.openarchives.org/OAI/openarchivesprotocol.html>

⁶⁹ http://cheshire.berkeley.edu/Cheshire_Sardinia.pdf

⁷⁰ <http://www.eprints.org/>

⁷¹ <http://archivegrid.org/web/index.jsp>

⁷² <http://drupal.org/node/23405>

⁷³ <http://en.wikipedia.org/wiki/YouTube>

⁷⁴ <http://www.flickr.com/>

⁷⁵ <http://www.vimeo.com/>

⁷⁶ <http://www.blinkx.com/>

⁷⁷ <http://www.meetup.com/>

⁷⁸ <http://www.grokker.com/>

⁷⁹ www.sherpa.ac.uk/documents/OpenDOAR_and_ROAR.ppt

⁸⁰ <http://en.wikipedia.org/wiki/DBpedia>

⁸¹ <http://en.wikipedia.org/wiki/GeoNames>

⁸² <http://developer.amazonwebservices.com/connect/kbcategory.jspa?categoryID=243>

⁸³ <http://www.archive.org/index.php>

⁸⁴ <http://www.librarything.com/>

applications, e.g. citeulike⁸⁵, ISI Web of Science⁸⁶, and journal rankings and acceptance rates⁸⁷.

THE SEMANTIC WEB DEVELOPMENT IN THE U.S. LIBRARIES

The development of Semantic Web has been slow and incremental. One reason for this identified by Sir Tim Berners-Lee, points to unfinished ontology work (Harper & Tillett, 2007). As each layer of the semantic stack is built upon the layer below, work on rules and logics cannot go on without the completion of ontology. The same can be said about the U.S. library Semantic Web development. Most of the library-related projects are carried on by big organizations or in a specific discipline, concentrating on creation of semantic tools, especially in the area of converting MARC and cataloging rules into XML and RDF. A literature review reveals an on-going discussion on the significance of converting controlled vocabularies and thesauri into ontologies. There are many journal articles on visions and theoretical models for library Semantic Web development. So far, most work completed or in progress in United States libraries falls into the category of the developmental stage one, namely, creating semantic tools. The work for “exposing collections” has been carried out only by big organizations like LC and OCLC. Slowly but steadily Semantic Web technologies are being incorporated into existing digital content management systems. Library applications are maturing in the URI, XML and RDF layers of the Semantic Stack.

Most library projects on Semantic Web in the United States have been conducted on several fronts: the national libraries, including the Library of Congress (LC) and the National Library of Medicine, the Dublin Core Metadata Initiative (DCMI), and OCLC. There are other individual library projects on a smaller scale scattered around the country.

Library of Congress Projects

The U.S. Library of Congress has been one of the leading forces in Semantic Web development. One of its achievements involves collaboration with W3C Semantic Web Development Group in creating the format of Simple Knowledge Organization System (SKOS) and SKOS Primer.

The actual LCSH/SKOS project included several small projects, one of which was the LCCN Permalink⁸⁸ project. It created persistent URLs from LC bibliographic records in LC’s online catalog using the existing LCCN in the records as the base identifiers in the URLs. This practice was expected to expand to all authority records – name/title, subject, and classification. As of

⁸⁵ <http://www.citeulike.org/>

⁸⁶ http://thomsonreuters.com/products_services/scientific/Web_of_Science

⁸⁷ <http://www.stjohns.edu/academics/libraries/resources/journals.stj>

⁸⁸ <http://lccn.loc.gov/>

June 2008, LC planed to provide Web-based identifiers for all other LC-maintained vocabularies, including MARC, MODS, and other controlled vocabularies it maintains (Marcum, 2008, p. 44)⁸⁹

To optimize data reuse, LC developed content and format guidelines for submission of ONIX data to the CIP program. The process for re-tooling ECIP to accept ONIX data directly from publisher is under development. LC's Virtual International Authority File (VIAF) enabled retrieval via OAI-PMH (Open Archive Initiative – Protocol for Metadata Harvesting). However, the approach is limited to large-scale harvesting of entire resources. In addition, LC finding aids created according to Encoded Archival Description (EAD) standard are available online at LC on the Web and with linking entries through ILS.

To optimize data interoperability, LC contributed “resources to support the work of coordinating the definitions and linkages of data elements in nationally and internationally accepted bibliographic standards (Marcum, 2008, p.40)⁹⁰.” In collaboration with W3C Consortium, the Dublin Core Metadata Initiative community, and the W3C Semantic Web Deployment Working Group, LC “provided SKOS representations for vocabularies and data elements used across standards such as RDA, MARC, PREMIS, and METS (Marcum, 2008, p. 41)⁹¹.”

As an experiment, LC made LCSH in RDF using SKOS. The LC Name Authority File will also be available in that format. In collaboration with Library and Archives Canada, and the British Library, LC planned to enable RDA use of MARC and its derivatives (MARCXML, MODS, MADS) (Marcum, 2008, p. 42).⁹²

LC maintains registries of standards and of data elements pertinent to individual standards, particularly related to RDA/DCMI⁹³, which has been led by Diane Hillmann and Gordon Dunsire (Marcum, 2008, p.43)⁹⁴.

To increase explicit correlation and referencing between LCSH terms, and LCC and DDC numbers, LC is using OCLC Product ‘Dewey Correlations’ and LCSH/LCC Correlations in Classification Web (Marcum, 2008, p.67)⁹⁵. All of these are part of the Web as infrastructure development and necessary steps in making progress towards semantic digital libraries.

With the completion of SKOS, the next step is to convert the controlled vocabularies and thesaurus into RDF. “As Harper (2007) has illustrated, there is great potential value in making controlled vocabularies like LCSH available and reference-able on the Web using Semantic Web technologies (Summers, 2008).” In 2006, LC started the project converting LCSH MARC to SKOS, and delivering

⁸⁹ <http://www.loc.gov/bibliographic-future/news/LCWGResponse-Marcum-Final-061008.pdf> & <http://lcn.loc.gov/00714874>

⁹⁰ <http://www.loc.gov/bibliographic-future/news/LCWGResponse-Marcum-Final-061008.pdf>

⁹¹ <http://www.loc.gov/bibliographic-future/news/LCWGResponse-Marcum-Final-061008.pdf>

⁹² <http://www.loc.gov/bibliographic-future/news/LCWGResponse-Marcum-Final-061008.pdf>

⁹³ <http://dublincore.org/dcmirdataskgroup/>

⁹⁴ <http://www.loc.gov/bibliographic-future/news/LCWGResponse-Marcum-Final-061008.pdf>

⁹⁵ <http://www.loc.gov/bibliographic-future/news/LCWGResponse-Marcum-Final-061008.pdf>

LCSH/SKOS as linked data (Summers, 2008). In late April of 2009, the LC Authorities and Vocabularies Service went live and LCSH/SKOS was released. The Library of Congress had prepared this subject terminology system for several years, and now has made it available as a public domain data set. An estimated 342,684 subject authority records can be searched, or downloaded in RDF/XML or N-Triples formats at <http://id.loc.gov/authorities>. LC will also release additional vocabularies including the Thesaurus of Graphic Materials, MARC Geographic Area Codes, MARC Language Codes, and MARC Relator Codes in the near future.

Library of Congress is also involved in the development of RDA (Resource Description and Access). RDA is "A new standard for resource description and access, design for the digital world (Tillett, 2009)." Currently it is being developed by the Joint Steering Committee for Development of RDA with contributions from the Canadian National Library, the British National Library, and the National Library of Australia. The first release will be in late 2009 or early 2010.

Upon completion, RDA will be a Web-based tool for cataloging all types of contents and media. The records created by RDA will be readily adaptable to Semantic Web database structures. The United States' implementation plans include the following (Tillett, 2009):

- 1) LC will prepare training materials of RDA
- 2) The three U.S. national libraries (Library of Congress, National Agricultural Library, and National Library of Medicine) will test RDA jointly
- 3) Around 20 additional organizations in the U.S. will participate in RDA testing
- 4) The three U.S. national libraries will make a joint decision on implementation of RDA, based on the results of a test of both RDA and the Web product.

OCLC Projects

While the Library of Congress is busy with converting controlled vocabularies into RDF and experimenting with RDA, OCLC has been conducting a series of FRBR-izing projects. FRBR (Functional Requirements for Bibliographic Record) is a conceptual model created by IFLA in 1998 for cataloging purpose. "IFLA's FRBR is a semantic expression of the relationships between items in the library catalog (Wallis, 2004)." Among the tools created for FRBR-izing bibliographic data are FRBR Work-Set Algorithm and xISBN Web Service. The former is a computer algorithm for converting MARC21 bibliographic databases to the "FRBR" model and the latter is an ISBN retrieval device for a list of related ISBNs and selected metadata. In the future, OCLC plans to "FRBR-ize" WorldCat as it implements its new database technology.

FictionFinder⁹⁶ is another OCLC achievement. It is a FRBR-based prototype for fiction which supports searching & browsing of fiction materials cataloged in WorldCat. It employs FRBR to build a “work” view & cluster related records. It provides access to over 2.8 million records for books, eBooks, and audio materials tagged as fiction in OCLC WorldCat.

The OCLC project WorldCat Identities⁹⁷ is based on the concept of creating a summary page for every name in WorldCat. It provides pages for 20 million 'identities' (predominantly authors and persons who are the subjects of published titles). It uses an entity-relationship model of metadata for an information object. It has a fuzzy searching module. In this project FRBR grouping is at the “expression” level (Hickey, 2007).

Under the auspices of LC, OCLC, and RLG (Research Library Group), PREMIS Data Dictionary is in its second release since 2008. A data dictionary, as defined in the *IBM Dictionary of Computing*, is a "centralized repository of information about data such as meaning, relationships to other data, origin, usage, and format." The PREMIS Data Dictionary is a specification that emphasizes metadata that may be implemented in a wide range of repositories, supported by guidelines for creation, management and use, and oriented toward automated workflows. The working group developed a simple data model to organize the semantic units defined in the Data Dictionary, rather than metadata elements.

Closely related to the PREMIS Data Dictionary project is ECHODEP (ECHO DEpository). It is a 5-year (2004-2009) digital preservation research and development project at the University of Illinois at Urbana-Champaign, funded by the Library of Congress under its National Digital Information Infrastructure and Preservation Program (NDIIPP). In this project PREMIS is integrated with METS (Metadata Encoding and Transmission Standards). Phase I has been completed; it includes digital archiving tool development, repository evaluation, interoperability tool development, and long-term semantic preservation research. Phase II is still under development. It includes semantic archiving for preservation of meaning and structure⁹⁸.

Other Projects

Many projects aim to add semantic technologies to existing digital management systems. For example, DSpace is a well-known digital depository system developed by MIT Libraries and HP. There are three projects aiming to incorporate semantic features into DSpace. Those projects are SIMILE, Semantic Search, and SKOS (Dspace Projects, 2009) SIMILE⁹⁹ is a project jointly sponsored by MIT libraries, MIT CSAIL, W3C, and Hewlett-Packard Research Labs. It aims at extending support for metadata schema in DSpace through Semantic Web technologies. As DSpace is one of the major open source digital

⁹⁶ <http://fictionfinder.oclc.org/>

⁹⁷ <http://orlabs.oclc.org/Identities/>

⁹⁸ <http://ndiipp.uiuc.edu/>

⁹⁹ <http://simile.mit.edu/>

management systems in the United States, it has been used by many libraries and institutions in the U.S. and other countries all over the world. However, its metadata support is currently limited to the general but relatively small Dublin Core descriptive metadata schema. The goal of SIMILE is to add DSpace support for additional metadata schemas for a variety of purposes: finding digital research material described in various, domain-specific ways, and managing the digital content over time in order to preserve it. In addition, SIMILE has developed some robust, open source tools for the following functions:

- 1) Converting structured data to RDF: a tool called RDFizer is being created for converting structured data to RDF, for instance, from MARC to RDF. It is still under development.
- 2) Visualized searching: Gadget and Welkin are completed. They are data graph viewers for XML.
- 3) Browsing capability: Longwell and Piggy Bank are completed. Longwell is a Web-based, RDF-powered, highly-configurable faceted browser. Piggy Bank is a Firefox plug-in that can turn a browser into a mashup platform by extracting data from different Web sites and mixing it up.
- 4) Lightweight UI-SIMILE has finished building web widgets, such as Timeline and Timeplot.

Another semantic achievement in the United States is Fedora, an acronym for Flexible, Extensible, and Digital Object Repository Architecture. It is a joint project of the Fedora Commons, Cornell University, and the University of Virginia. What is unique about Fedora is its digital asset management (DAM) architecture and that libraries can take advantage of it to build many applications including digital libraries or archives and institutional repositories. Fedora integrates semantic technologies (such as Mulgara, RDF Triple-store) into its services for describing and inter-relating digital objects, providing a simple, practical way to begin using these exciting new capabilities. The architecture includes a generic Resource Description Framework (RDF)-based relationship model that represents relationships among objects and their components. Queries against these relationships are supported by an RDF triple store. The architecture is implemented as a web service, with all aspects of the complex object architecture and related management functions exposed through REST and SOAP interfaces (Lagoze, 2006).

Among the successful implementations of Fedora is the National Science Digital Library (NSDL), a project completed by the National Science Foundation. It is a free online library for education and research in science, technology, and mathematics. In January 2007, the project team released NSDL 2.0. Built on Fedora, this new version is a fully functional semantic digital library.

Recently Fedora Commons and the DSpace Foundation, two of the largest providers of open source software for managing and providing access to digital content, announced that they would merge their organizations to pursue a common mission. The new organization is named DuraSpace. Its mission is to support both DSpace and Fedora by working closely with both communities and

when possible, developing synergistic technologies, services, and programs that increase interoperability of the two platforms. DuraSpace will also support other open source software projects including the Mulgara semantic store, a scalable RDF database¹⁰⁰.

Appendix A is a list of academic library-related projects in the United States that are aimed at creating tools and converting library controlled vocabularies and thesauri into Semantic Web standards. Appendix B is a list of Semantic Web projects undertaken by libraries in Europe.

A COMPARISON

Library Semantic Web development is stronger and more aggressive in Europe than in the United States. European academic and research libraries showed more enthusiasm and passion towards adopting Semantic Web technologies, with Italy being the center of library semantic activities. This is partly evidenced by the number and frequency of semantic conferences held in the U.S. and Europe. In 2009 there are four conferences in Italy on Semantic Web and digital libraries. A search on the Internet shows many presentations and articles on the subject from European-held conferences and meetings. European librarians welcome semantic technologies with greater enthusiasm.

In addition, the nature and emphasis on Semantic Web development is different between European and the United States libraries. European libraries emphasize delivering Semantic Web and digital library applications, while American libraries work hard at developing Semantic Web tools and infrastructure, e.g. converting controlled vocabularies, and laying ground for further semantic development. Many American Semantic Web projects have been carried out for domain-specific applications, e.g. medicine, law, and occurred in the backend of the digital library systems. They were less visible to the public with the exception of big organizations like the Library of Congress, OCLC, and DCMI. For American librarians, the Semantic Web is still a very abstract concept that has not yet taken concrete form. In contrast, many European projects, once finished, became useful applications attracting publicity and in turn serving as advertisements for this new technology. The advantages of Semantic Web technologies are being demonstrated in those applications. For instance, the National Library of Sweden has created LIBRIS, their national union catalog, and made it part of the Semantic Web applications. Another example may be JeromeDL. In both cases semantic technologies are being seen and put to use in the public domain.

LIBRIS is a bold endeavor and the “the first union catalogue or national library catalogue to be published in its entirety as linked data (Söderbäck & Malmsten, 2009).” The project team stated that “the best way to get to know a technology—

¹⁰⁰ <http://www.duraspace.org/pressrelease.html>

whether MARC records or the Semantic Web—is to get our hands dirty and work with it (Söderbäck & Malmsten, 2009).” In comparison, American librarians are more cautious as our cataloging practice is deeply rooted in AACR cataloging rules and traditions. For instance, the British National Library, the National Library of Canada, and the Australia National Library will thrust themselves forward with RDA development and implementation, while the American national libraries have decided to conduct further tests before making any decisions or commitment for implementation.

Despite the differences listed above in Semantic Web application implementations between the U.S. and European libraries, some projects for semantifying digital libraries are international in scale, e.g. Drupal, Talis, Dspace, and Fedora. These projects cover a broad spectrum of data applications from building institutional repositories to supporting browsing, search and retrieval, linking, and interchanging the Web of data on the Internet through Semantic Web standards and technologies. The projects have also been developed with the close collaboration of metadata standards institutions and communities, facilitating easy discovery, interoperability, and reuse.

CONCLUSION

In this paper, we shared our findings about Semantic Web development in the U.S. libraries in comparison with that of the European libraries. Information has been obtained by seeking collaboration with thought-leaders in Semantic Web standards, technologies development from academia, and IT industries, as well as library information technologies vendors. In order to get the feel of the semantic wave, we went to many gatherings at NYC Semantic Web Meetups and Enterprise Search Summit, watched demos, compared products, and reached out to leading IT solution providers in Semantic Web development. In addition, we have researched extensively the literature on recent Semantic Web development in both U.S. and European libraries.

Our findings can be summarized as the following: 1) Semantic-aware tools have been increasingly embedded in current Web applications with different levels of semantic support from browsers to desktops, from clients to servers, and servers to servers; 2) IT vendors started to view the capability of combining the Web of Data as an opportunity to move from current HTML-based Web to the Web of Linked Data, especially big vendors such as Google, Yahoo, Oracle and others; 3) the vast amount of library data, e.g. bibliographic data, authority data, controlled vocabularies, non-MARC oriented meta-data and classification schemes have been standardized for representations, controlled for quality and entity resolutions, and cross-mapped for interoperability and reuse; and 4) the library-related Web of Data is ready to be exposed to Semantic Web application development.

If we regard the glue of the traditional Web as hypertext links between HTML pages, the glue of the Web of Data can be considered as the RDF triples. With more and more RDF triples found in semantic applications, and the building blocks of Semantic Stack being similar to the library catalog services, libraries should be at the front of the Semantic Web wave. Libraries, particularly European ones, have already developed plausible infrastructure and tools for the Semantic Wave. Their achievements have proved that the skills of librarians in information organization and resource sharing can be leveraged to build the universal digital library on the Semantic Web.

In the meantime, librarians should take advantage of the Semantic Wave either through semantic-aware tools or through fully functional semantifying digital library packages. In addition, they should actively participate in the development of Web standards and technologies, share best practices in weaving the Semantic Web, and update their services into the context of the users' problem-solving domains in whatever the appropriate Web infrastructure calls for, such as Web in HTML, Web in social networking, Web in semantics, and Web in intelligence.

REFERENCES

- Beckett, D. (2001). Practical RDF tools towards a semantic web. In *ILRT* [PowerPoint presentation]. Retrieved May 30, 2009, from Institute for Learning and Research Technology University of Bristol Web site: <http://www.dajobe.org/talks/kt2001/slides.html>
- Berners-Lee, T. (2000). Architecture. In *W3C*. Retrieved May 30, 2009, from W3C Web site: <http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html>
- Berners-Lee, T. (2007, May 2). Linked data. In *W3C*. Retrieved May 30, 2009, from W3C Web site: <http://www.w3.org/DesignIssues/LinkedData.html>
- Bizer, C., Cyganiak, R., & Hartig, O. (2009). The Named Graphs API for Jena (NG4J) is an extension to the Jena Semantic Web framework for parsing, manipulating and serializing sets of Named Graphs. In *NG4J - Named Graphs API for Jena*. Retrieved May 31, 2009, from <http://www4.wiwiw.fu-berlin.de/bizer/ng4j/>
- Bizer, C., Cyganiak, R., & Heath, T. (2007, July 27). How to Publish Linked Data on the Web. In *WWW4*. Retrieved May 31, 2009, from <http://www4.wiwiw.fu-berlin.de/bizer/pub/LinkedDataTutorial>
- Cacao Project. (2009, May 25). Overview. In *Cacao Project*. Retrieved May 30, 2009, from <http://www.cacaoproject.eu/project-overview/>

- Dodds, L. (2005, November 16). Introducing SPARQL: Querying the semantic web. In *O'reilly's XML.com*. Retrieved May 30, 2009, from <http://www.xml.com/pub/a/2005/11/16/introducing-sparql-querying-semantic-web-tutorial.htm>
- DSpace Projects. (2009, May 21). *DSpacewiki*. Retrieved May 28, 2009, from <http://wiki.dspace.org/index.php/DspaceProjects>
- Dunsire, G. (2008, August 8). RDA vocabularies and concepts. In *The IFLA Satellite Meeting on RDA August 8, 2008, QuébecCity, Canada* [PowerPoint presentation]. Retrieved May 30, 2009, from Digital Library Research Center, University of Strathclyde Web site: <http://www.collectionscanada.gc.ca/jsc/docs/iflasatellite-20080808-dunsire.pdf>
- Fielding, R. T. & Taylor, R. N. (2002). Principled Design of the Modern Web Architecture. *ACM Transactions on Internet Technology*, 2(2), 1-36. Retrieved June 21, 2009, from <http://portal.acm.org/citation.cfm?doid=514183.514185>
- Giant Global Graph. (2009, May 21). *Wikipedia: The free encyclopedia*. Retrieved May 30, 2009, from http://en.wikipedia.org/wiki/Giant_Global_Graph
- Hakia aims to woo librarians. (2008, August 1). *Library Journal*, 133(18), 22. Retrieved April 4, 2009, from Academic Search Premier database.
- Harper, C. A., & Tillett, B. B. (2007). Library of Congress Controlled Vocabularies and Their Application to the Semantic Web. *Cataloging & Classification Quarterly*, 43(3/4), 47-68.
- Herman, I. (2009, May). Introduction and applications of the semantic web. In *W3C Semantic Web* [PowerPoint Presentation]. Retrieved May 30, 2009, from W3C Australian Office Web site: <http://www.w3.org/2009/Talks/05-Oz-IntroSW-IH/Slides.pdf>
- Hickey, T.B. (2007). WorldCat Identities: Another view of the catalog. *NextSpace: the OCLC Newsletter*, no.6, 18-19.
- Hoaro, D. (2006). SPDL: Semantic personal digital library. In *Semantic Personal Digital Library* [an open source Java project for managing PDF documents. This project allows user to express information's on documents, to classify and retrieve documents in different ways]. Retrieved April 23, 2009, from <http://spdl.sourceforge.net/index.html>
- JeromeDL - e-Library with semantics home. (2009). *JeromeDL - e-Library with Semantics*. Retrieved April 25, 2009, from Main Library of Gdansk University of Technology, DERI.International Web site: http://www.jeromedl.org/index.php?option=com_content&task=blogcategory&id=1&Itemid=103
- Lagoze, C., Payette, S., Shin, E., & Wilper, C. (2006, April). Fedora: an architecture for complex objects and their relationships. *International Journal on Digital Libraries*, 6(2), 124-138. Abstract retrieved May 28, 2009, from Academic Search Premier of EBSCO

- Malmsten, M. (2008). Making a Library Catalogue Part of the Semantic Web. In Berlin Proc. Int'l Conf on Dublin Core and Metadata Applications. Retrieved June 21, 2009, from the conference Web site: <http://dcpapers.dublincore.org/ojs/pubs/article/viewArticle/927>
- Marcum, D. B. (2008, June 1). *Response to On the Record: Report of the Library of Congress Working Group on the Future of Bibliographic Control*. Retrieved May 31, 2009, from <http://www.loc.gov/bibliographic-future/news/LCWGResponse-Marcum-Final-061008.pdf>
- National Cancer Institute. (n.d.). National Cancer Institute Thesaurus. In *Maryland Information and Network Dynamics Lab Semantic Web Agents Project*. Retrieved April 23, 2009, from Maryland Information and Network Dynamics Lab Web site: <http://www.mindswap.org/2003/CancerOntology/>
- RDA vocabularies project. (2008, April). *ALCTS Newsletter Online*, 19(2), 41. Retrieved from American Library Association Web site: <http://lita.org/ala/mgrps/divs/alcts/resources/ano/v19/n2/nws/prof.cfm#vocab>
- Sandars, J. (2009, March/April). Reflective Learning for the Net Generation. *D-Lib Magazine*, 14(3/4). Retrieved May 30, 2009, from University of Leeds, United Kingdom Web site: <http://www.dlib.org/dlib/march09/03inbrief.html>
- Semantic MediaWiki. (2009, April). *Wikipedia: The free encyclopedia*. Retrieved May 30, 2009, from http://en.wikipedia.org/wiki/Semantic_MediaWiki
- Semantic Web. (2009, May 29). *Wikipedia: The free encyclopedia*. Retrieved May 30, 2009, from Wikipedia Project Web site: http://en.wikipedia.org/wiki/Semantic_Web
- Semantic wiki. (2009, May 30). *Wikipedia: The free encyclopedia*. Retrieved May 30, 2009, from http://en.wikipedia.org/wiki/Semantic_wiki
- Sauermann, L., Cyganiak, R., Ayers, D., & Völkel, M. (2008). W3C Interest Group note 03 December 2008. In *Cool URIs for the semantic web*. Retrieved May 31, 2009, from W3C Web site: <http://www.w3.org/TR/cooluris>
- Söderbäck, A., & Malmsten, M. (2009, January 16). LIBRIS - Linked library data. In *Nodalities blog: From semantic web to web of data*. Retrieved May 30, 2009, from <http://blogs.talis.com/nodalities/2009/01/libris-linked-library-data.php>
- Summers, E., Isaac, A., Redding, C., & Krech, D. (2008). LCSH, SKOS and Linked Data. *Proceedings of the International Conference on Dublin Core and Meta-data Applications*. Retrieved from <http://dcpapers.dublincore.org/ojs/pubs/article/viewArticle/916>
- Tillett, B. B. (2009, May 7). Getting Ready for RDA: What You Need to Know. *Massachusetts Library Association Conference, MLA Technical Services*

Section. Retrieved May 26, 2009, from <http://www.collectionscanada.gc.ca/jsc/rdapresentations.html>

Tunkdang, D. (2009, March 11). *Exploring semantic means*. PowerPoint presented at New York Semantic Web Meetup: Exploring Semantic Means, New York.

Wallis, J. C. (2004, June 14). An argument for a semantic web based FRBR union catalog. In *Information retrieval systems: User centered design*. Retrieved May 12, 2004, from UCLA Department of Information Studies IS 277: User centered design Web site: <http://www.moebiusstrip.org/277/final.html>

W3C. (2006-2007). SPARQL Query Language for RDF. In W3C. Retrieved May 30, 2009, from W3C Web site: <http://www.w3.org/TR/rdf-sparql-query/>

W3C. (2007). SPARQL query results XML format. In W3C. Retrieved May 30, 2009, from <http://www.w3.org/TR/rdf-sparql-XMLres/>

APPENDIX A

Library-Related Semantic Projects in the U.S.

Project 1

- Name: MARC XML
- Organization: LC, OCLC, and RLG
- Description: Created an XML schema to convert MARC 21 to XML and vice versa
- Categories:
 - i)* Tool->Publishing->Converter->Bibliographic data
 - ii)* Best practice->XML Schema Language of W3C
- Tools: MARCEdit 4.5 or higher, MARC XML, MarcXML Converter, OCLC Connexion
- Status: Completed
- URL: <http://www.loc.gov/standards/marcxml/>

Project 2

- Name: MARC Relator terms in Dublin Core metadata
- Organization: LC and DCMI
- Description: Expression of a sub-set of MARC Relator Terms in RDF for use as refinements of the DC contributor elements
- Category: Tool->Publishing->Converter->Metadata
- Tools: Connexion, ready to export MARC XML into RDF files
- Status: Completed
- URL: <http://www.loc.gov/standards/marcxml/>

Project 3

- Name: MODS (Metadata Object Description Schema)
- Organization: LC
- Description: An XML schema for a bibliographic element set selected from MARC 21 records that may be used for a variety of purposes. It is a content model supporting a subset of MARC21, SRU-specified format, which has an extension to METS (Metadata Encoding and Transmission Standard)
- Categories:
 - i)* Tool->Publishing->Converter->Bibliographic data
 - ii)* Best practice->XML Schema Language of W3C
- Tools: MARCXML to MODS; MODS to MARCXML
- Status: MODS schema version 3.3 is under development
- URL: <http://www.loc.gov/standards/mods/>

Project 4

- Name: **MADS** (Metadata Authority Description Schema)
- Organization: LC
- Description: XML markup for selected authority data from MARC21 records as well as original authority data
- Categories:
 - i)* Tool->Publishing->Converter->Authority data
 - ii)* Best practice->XML Schema Language of W3C
- Tools: N/A
- Status: Completed
- URL: <http://www.loc.gov/standards/mads/>

Project 5

- Name: **EAD** (Encoded Archival Description)
- Organization: LC
- Description: XML markup designed for encoding finding aids
- Categories:
 - i)* Tool->Publishing->Metadata
 - ii)* Best practice->XML, Relax NG Schema (RNG), and W3C Schema (XSD)
- Tool: Archivists' Toolkit
- Status: Under development
- URLs:
 - i)* <http://www.loc.gov/ead/>
 - ii)* <http://archivalsoftware.pbworks.com/FrontPage>
 - iii)* http://clir.org/pubs/reports/spiro/spiro_Jan13.pdf

Project 6

- Name: METS (Metadata Encoding and Transmission Standard)
- Organization: LC
- Description: Structure for encoding descriptive, administrative, and structural metadata
- Category: Tool->Publishing->Metadata

- Tools: Archivists' Toolkit, DigiTool, Fedora and DSpace
- Status: Under development (Version 1.8 schema)
- URL: <http://www.loc.gov/standards/mets/>

Project 7

- Name: MIX (NISO Metadata for Images In XML Schema)
- Organization: LC
- Description: XML schema for encoding technical data elements required to manage digital image collections
- Categories:
 - i)* Tool->Publishing->Metadata->Digital Still Images
 - ii)* Best practice->Data Dictionary
 - iii)* Best practice->XML Schema Language of W3C
- Tools: JHOVE by JSTOR/Harvard Object Validation Environment
- Status: Under development (version 2.0)
- URL: <http://www.loc.gov/standards/mix/>

Project 8

- Name: PREMIS Data Dictionary for Preservation Metadata
- Organization: LC, OCLC, and other country libraries
- Description: Defines a set of semantic units that are implementation independent, practically oriented, and likely to be needed by most preservation repositories
- Categories:
 - i)* Tool->Digital Archiving->Metadata;
 - ii)* Best practice->Data Dictionary
 - iii)* Best practice->Preservation
- Tool: ECHODEP Hub and Spoke Interoperability
- Status: Completed version v.2.0 (March 2008)
- URLs:
 - i)* <http://www.oclc.org/research/projects/pmwg/>
 - ii)* <http://www.loc.gov/standards/premis/tools.html>

Project 9

- Name: TextMD (Technical Metadata for Text)
- Organization: LC
- Description: XML schema that details technical metadata for text-based digital objects
- Category: Tool->Publishing->Metadata
- Tool: JHOVE ASCII & UTF-8 Modules
- Status: Updated schema (3.0 Alpha)
- URL: <http://www.loc.gov/standards/textMD/>

Project 10

- Name: LCSH/SKOS - LC Controlled Vocabularies - Application to the Semantic Web
- Organization: LC

- Description: A web application which will make the Library of Congress Subject Headings available as SKOS. Convert LCSH/MARC to LCSH/SKOS, and deliver LCSH/SKOS as Linked Data.
- Category: Tool->Publishing->Controlled vocabulary
- Tools: cURL, wget
- Status: In 4-6 weeks LC will release the first offering the LC Subject Headings
- URLs:
 - i)* <http://id.loc.gov>
 - ii)* <http://id.loc.gov/authorities/techcenter/downloads.html>
 - iii)* <http://dcpapers.dublincore.org/ojs/pubs/article/viewArticle/916>

Project 11

- Name: RDF Topicmaps
- Organization: OCLC
- Description: The RDF Topicmaps project explores subject navigation of Web sites using semi-automatically generated finding aids.
- Categories:
 - i)* Tool ->Controlled vocabulary;
 - ii)* Tool ->Metadata;
 - iii)* Tool ->ETL-> RDF
- Tools: Web page harvester, topicmap generator and web-based interface to the topicmaps
- Status: Version 2 (2006)
- URLs:
 - i)* http://www.oclc.org/research/projects/rdf_topicmaps/default.htm
 - ii)* http://www.oclc.org/research/software/rdf_topicmaps/documentation/default.htm
 - iii)* http://www.oclc.org/research/software/rdf_topicmaps/default.htm

Project 12

- Name: WorldCat Identities
- Organization: OCLC
- WorldCat Identities is based on the concept of creating a summary page for every name in WorldCat. It uses an entity-relationship model of metadata for information object. It has a fuzzy searching module.
- Categories:
 - i)* Tool ->Publishing->Authority Data;
 - ii)* Tool ->Linked Data
 - iii)* Tool->Linked Data->FRBR grouping at the expression level
- Tools: SRU searches, NameFinder searches, OpenURL
- Status: beta
- URLs:
 - i)* <http://www.oclc.org/nextspace/006/research.htm>
 - ii)* <http://www.worldcat.org/identities/>
 - iii)* <http://outgoing.typepad.com/outgoing/2008/06/linking-to-worl.html>

Project 13

- Name: Terminology Services Project
- Organization: OCLC
- Description: This project provides Web-based services for controlled vocabularies. The Terminology Services prototype uses library and web standards to make the terms, relationships, descriptions, and other information in controlled vocabularies available as resources on the web
- Categories:
 - i)* Tool->Publishing->Controlled vocabulary;
 - ii)* Tool->Linked Data
- Tools:
 - i)* Search descriptions of controlled vocabularies
 - ii)* Search for terms/headings in a controlled vocabulary
 - iii)* Retrieve a single term/heading by its identifier
 - iv)* View relationships for a term/heading including equivalence, hierarchical, and associative
 - v)* Retrieve terms/headings in multiple representations, such as HTML, MADS, MARC XML, SKOS, and Zthes.SKOS¹⁰¹
- Status: Under development
- URL: <http://www.oclc.org/research/projects/termservices/>

Project 14

- Name: Curiouser
- Organization: OCLC
- Description: The Curiouser project will produce a prototype system that provides a works-based view of Open WorldCat (OWC) content
- Categories:
 - i)* Tool->Linked data;
 - ii)* Tool->Explanatory search
- Tools: Employs the OCLC FRBR work-set algorithm and FictionFinder
- Status: Completed (2005)
- URL: <http://www.oclc.org/research/projects/curiouser/default.htm>

Project 15

- Name: FictionFinder
- Organization: OCLC
- Description: A FRBR-based prototype for fiction in WorldCat
- Categories:
 - i)* Tool->Publishing->Bibliographic data;
 - ii)* Tool->Linked Data;
 - iii)* Tool->Explanatory search
- Tools: FRBR
- Status: Beta
- URLs:

¹⁰¹ <http://www.oclc.org/research/projects/termservices/>

- i)* <http://www.oclc.org/research/projects/frbr/fictionfinder.htm>
- ii)* <http://fictionfinder.oclc.org/>

Project 16

- Name: RDF-DC Interoperability
- Organization: OCLC
- Description: This project will bring together a substantial amount of data from different sectors in different countries, unified using RDF Schema declarations as described above. Participants will be recruited from the government sector, museums, the commercial sector, super-governmental organizations, and education. The resulting database will comprise a test-bed accessible to researchers and designers to demonstrate and experiment with an operational cross-disciplinary store. This test-bed will also provide a tutorial by example that will help to promote this approach to enhancing cross disciplinary interoperability.
- Categories:
 - i)* Tool->Publishing ->Metadata;
 - ii)* Tool->Publishing->RDF
- Tools: RDF schema
- Status: Phase I & II completed, Phase III under development
- URL: http://www.oclc.org/research/projects/rdf_interop/default.htm

Project 17

- Name: xISBN Web Service
- Organization: OCLC
- Description: xISBN Web Service supplies ISBNs and other information associated with an individual intellectual work that is represented in WorldCat. Submit an ISBN to this service, and it returns a list of related ISBNs and selected metadata.
- Category: Tool->Linked Data
- Tools: FRBR
- Status: Completed
- URL: <http://www.worldcat.org/affiliate/webservices/xisbn/app.jsp>

Project 18

- Name: ECHODEP
- Organization: University of Illinois with partners, founded by LC
- Description: The ECHO DEPOSITORY is a 5-year (2004-2009) digital preservation research and development project at the University of Illinois at Urbana-Champaign, funded by the Library of Congress under their National Digital Information Infrastructure and Preservation Program (NDIIPP).
- Categories:
 - i)* Tool->Digital Archiving->Metadata;
 - ii)* Best practice->Preservation
- Tools: N/A
- Status: Phase I completed, Phase II under development

- URL: <http://www.ndiipp.uiuc.edu/>

Project 19

- Name: Helping Interdisciplinary Vocabulary Engineering (HIVE)
- Organization: University of North Carolina, NESCent
- Description: HIVE is an automatic metadata generation approach that dynamically integrates discipline-specific controlled vocabularies encoded with SKOS.
- Categories:
 - i)* Tool->Publishing-> Controlled vocabulary->SKOS;
 - ii)* Tool->Publishing->Converter->Metadata
- Tools: Metadata generator; SKOS
- Status: Latest page consulted (2008)
- URLs: <http://ils.unc.edu/mrc/hive/>
https://www.nescent.org/sites/hive/Main_Page

Project 20

- Name: SIMILE (Semantic Interoperability of Metadata and Information in unlike Environments)
- Organization: MIT libraries, MIT CSAIL, W3C, and Hewlett-Packard Research Labs
- Description: Semantic Web initiative that seeks to enhance interoperability among digital assets, schemata, vocabularies, ontologies, metadata and services. The initiative is a joint project of the W3C, MIT Libraries and MIT Computer Science and Artificial Intelligence Laboratory. The SIMILE project has created an application called Piggy Bank that is a Firefox Web browser extension which allows existing information on the Web to be used in more useful and flexible ways.
- Categories:
 - i)* Tool->Publishing-> Converter-> Metadata;
 - ii)* Tool->Publishing-> Converter->RDF;
 - iii)* Tool->Explanatory search
- Tools: RDFizer, Gadget, Welkin, Longwell, Piggy Bank, Semantic Bank, Timeline, Timeplot, and etc.
- Status: Under development
- URLs: <http://dltj.org/article/simile/>
<http://simile.mit.edu/>

Project 21

- Name: RDA (Resource Description and Access) Vocabularies Project
- Organization: Joint Steering Committee for Development of RDA
- Description: Publishing Semantic Web vocabularies in Resource Description Framework (RDF). These vocabularies will provide machine-accessible versions of **RDA** Element and Value vocabularies, allowing for reuse in applications and citation via Uniform Resource Identifiers (URIs) (RDA Vocabularies Project, 2008)
- Categories:

- i)* Tool->Publishing->Bibliographic data;
 - ii)* Tool->Publishing->Metadata;
 - iii)* Tool->Linked data->Discovery and interchange;
 - iv)* Best Practice->Digital Library metadata standard
- Tools: Content model for resource description and access for all content types, media types, and carrier types
- Status: Under development
- URLs:
 - i)* RDA Online - <http://www.rdaonline.org/>;
 - ii)* Joint Steering Committee for Development of RDA <http://www.collectionscanada.gc.ca/jsc/rda.html>;
 - iii)* DCMI/RDA Task Group <http://dublincore.org/dcmirdataskgroup>;

Project 22

- Name: Semantic Personal Digital Libraries
- Organization: Divide Hoaro
- Description: The Semantic Personal Digital Library (SPDL) is an open source Java project helping users to **organize and retrieve documents in PDF format**. It is Personal Information Management (PIM) software works with the technologies of Semantic Web. User can insert/edit information on documents like author, title, description, subject and so on. The information is stored as RDF (Resource Description Framework) and uses standard properties like those defined in the Dublin Core metadata set. Users can also build some simple ontologies, assisting in classifying documents by the subjects or by other free classifications. Retrieving documents can be done by searching on content and metadata values or browsing classification schemes. SPDL uses the Lucene index and SPARQL to search documents. SPDL is a personal academic project to help users in personal information management, but it's also an example of how technologies of Semantic Web can be used in real project where information has to be organized and retrieved. (Hoaro, 2006)
- Categories:
 - i)* Tool->Publishing->Converter->Metadata;
 - ii)* Tool->Publishing->RDF;
 - iii)* Best practice->Semantic Web systems for libraries
- Tools: Semantic Personal Digital Library (an open source Java program)
- Status: Completed in 2006
- URL: <http://spdl.sourceforge.net/index.html>

Project 23

- Name: Semantic MEDLINE Project
- Organization: Lister Hill National Center for Biomedical Communications
- Description: The Lister Hill National Center for Biomedical Communications is developing a program, called Semantic MEDLINE, which integrates PubMed searching, advanced natural language processing, automatic summarization, and visualization into a single Web portal. Semantic Medline is intended to help users manage the results of

PubMed searches by normalizing core information in the citations retrieved. Semantic MEDLINE is a prototype Web application that summarizes MEDLINE citations returned by a PubMed search. Natural language processing is used to analyze salient content in titles and abstracts. This information is then presented in a graph that has links to the MEDLINE text processed.

- Category: Tool->Linked data->Semantified digital libraries
- Tools: Semantic Medline, PubMed
- Status: Under development
- URL: <http://skr3.nlm.nih.gov/SemMedDemo/>

Project 24

- Name: National Cancer Institute Thesaurus: Maryland Information and Network Dynamics Lab Semantic Web Agents Project
- Organization: National Cancer Institute
- Description: The NCI Thesaurus is a public domain description of logic-based terminology produced by the National Cancer Institute, distributed as a component of the NCI Center for Bioinformatics caCORE distribution [1]. It is deep and complex compared to most broad clinical vocabularies, implementing rich semantic interrelationships between the nodes of its taxonomies. The semantic relationships in the Thesaurus are intended to facilitate translational research and to support the bioinformatics infrastructure of the Institute. Topics described in the ontology include diseases, drugs, chemicals, diagnoses, genes, treatments, anatomy, organisms, and proteins. (National Cancer Institute, n.d.)
- Category: Tool->Publishing->Controlled vocabulary
- Tools: NCI Thesaurus
- Status: Under development
- URL: <http://www.mindswap.org/2003/CancerOntology/>

Project 25

- Name: DuraCloud
- Organization: DuraSpace
- Description: DuraCloud is a web-based service that will enable content replication to a network of cloud-based storage services, and provide an environment for running access and preservation services within the cloud.
- Categories:
 - i)* Tool->Open source->RDF repository;
 - ii)* Tool->Digital Archiving->Metadata
- Tools: Cloud storage, cloud computing
- Status: Under development, early release will be available for selected pilot partners in fall 2009
- URL: <http://www.duraspace.org/pressrelease.html>

Project 26

- Name: Knoodl
- Organization: Revelytix, Inc
- Description: Knoodl is an Ontology Web Language (OWL) editor. It serves as a semantic technology platform enabling users to build semantic applications using their ontology and knowledgebase.
- Category: Tool->Open source->RDF knowledgebase
- Tools: Mulgara (RDF store), Krule (rules engine)
- Status: Available on web
- URLs: <http://www.knoodl.com>
<http://www.revelytix.com/knoodl.php>

Project 27

- Name: ONIX (Online Information Exchange)
- Organization: EDItEUR (International)
- Description: ONIX is a standard format that publishers can use to distribute electronic information about their books to wholesale, e-tail and retail booksellers, other publishers, and anyone else involved in the sale of books
- Category: Best practice->Digital library metadata standard
- Tool: XML Schema Language of W3C
- Status: Completed
- URL: <http://www.bisg.org/documents/onix.html>

Project 28

- Name: NSDL Semantic Digital Library Services
- Organization: National Science Digital Library
- Description: NSDL is a collaborative, contributory semantic digital library. The platform NSDL 2.0 supports storing both content and metadata, and allows arbitrary relationships among resource and metadata objects: organization, annotation, citation.
- Categories:
 - i)* Tool->Open Source->RDF repository;
 - ii)* Tool->Digital Archiving->Metadata->Registry;
 - iii)* Tool->Explanatory search
- Tool: NSDL 2.0
- Status: Completed
- URL: <http://openarchives.org/fedora/ESWC-Fedora.zip>
http://en.wikipedia.org/wiki/National_Digital_Library

Project 29

- Name: **Drupal**
- Organization: Open Source
- Description: Expose the Content Construction Kit (CCK) site content model as an OWL ontology that reflects the site structure, where a site administrator can define a node as content types, and define fields for each content type at design time. As a result, semantics are fixed at design time based on the site structure, where after end users enter data,

it claims that linked data can be produced automatically. Any imported web of data that adhere to linked data principles and in OpenVocab¹⁰² format can be mapped to existing ontologies in Drupal. In the future, site administrator can export their site content model and data to the Web of data in RDF/XML using SKOS.

- Categories:
 - i)* Tool->Open Source->RDF;
 - ii)* Tool->Open Source->Content management systems
- Tools: Site Vocabulary, CCK, LODR.info, Arc RDF Store, SPARQL for Drupal 6, Semantic Search in Drupal
- Status: Release 6.12
- URLs:
 - i)* <http://drupal.org/project/rdf>;
 - ii)* <http://lodr.info/>;
 - iii)* http://drupal.org/project/arc_rdf_store;
 - iv)* <http://drupal.org/project/sparql>
 - v)* <http://semanticsearch.org/>

¹⁰² <http://open.vocab.org/about>

APPENDIX B

Library-related Semantic Projects in Europe

Project 1

- Name: LIBRIS
- Organization: National Library of Sweden and 175 libraries in the country
- Description: This is the national union catalog of Sweden with six million records from 175 libraries, including the National Library of Sweden, many academic and research libraries, and 20 public libraries. Released in early 2008, it is “the first union catalogue or national library catalogue to be published in its entirety as linked data” (Söderbäck & Malmsten, 2009). The center piece for the project is a RDF server wrapper which can output RDF in response to a query (Malmsten, 2008). Dublin Core is used for bibliographical data, FOAF for persons and organizations, SKOS for controlled vocabularies (mapping Swedish subject headings to LCSH.info), and bibliontology for the ontology of LIBRIS data set. This approach enables LIBRIS to extend its linked data to dbpedia and wikipedia. Using identifiers from MARC field 001, HTTP URIs for bibliographic and authority records were created, and those URIs can be de-referenced with as much useful information and as many links to other URIs as possible. This project created a web of linked data with semantic searching capability using SPARQL for “good and exhaustive database querying” (Söderbäck & Malmsten, 2009). The project is a real live example of library semantic catalog.
- Category: Tool->Linked data->OPAC
- Tools: LIBRIS
- Status: Completed in 2008
- URL: <http://libris.kb.se/?language=en>

Project 2

- Name: Cacao Project
- Organization: Nine partners. Different organizations from France, Germany, Hungary, Italy and Poland. 20 month project supported by eContentplus Programme of the European Commission.
- Description: A digital library system of catalogs and digital contents for European users. Its multilingual capabilities will allow a user to search in his or her native language and retrieve documents in any chosen language. The project aims to be “the largest European aggregation of digital libraries, i.e. The European Library will adopt the CACAO infrastructure. The library catalogues of the five partner libraries will be aggregated into a single multilingual access point. Three thematic portals aggregating several European collections (Mathematics, Medieval

literature, Geography) will be made available to the public” (Cacao Project, 2009)

- Category: Tool->Linked data->OPAC
- Tools: Software package
- Status: Under development
- URL: <http://www.cacaoproject.eu/>

Project 3

- Name: SPAR Project
- Organization: the National Library of France
- Description: This is a digital preservation repository (Distributed Preservation and Archiving System) based on the Fedora platform. One of the main components of the system is the data management module. It will use RDF data stored in a RDF triple store. The Bibliotheque nationale de France has chosen Atos Origin to build SPAR. The first release is planned to be delivered at the beginning of 2009.
- Categories:
 - i)* Best practice->Preservation;
 - ii)* Tool ->Linked data->OPAC
- Tools: SPAR
- Status: Under development
- URL:
http://www.bnf.fr/pages/version_anglaise/numerisation/num_spar_eng.htm

Project 4

- Name: Talia
- Organization: European Commission
- Description: “Talia is a web-based distributed digital library and publishing system, designed for scholarly research in philosophy.” Talia is based on Semantic Web technologies and is an open source semantic digital library management system. ActiveRDF and ontology are applied. It is “a novel implementation of a semantic digital web library system” and part of a Discovery project. “Talia deploys Semantic Web technologies and uses computational ontologies for the organization of knowledge, which can help the definition of a state-of-the-art research and publishing environment for philosophy.”
- Categories:
 - i)* Tool->Open source ->RDF;
 - ii)* Tool->Content management systems->Domain specific
- Tools: Talia
- Status: Under development
- URL: <http://www.discovery-project.eu/>

Project 5

- Name: SKUA Project (Semantic Knowledge Underpinning Astronomy)
- Organization: The Joint Information Systems Committee in the UK

- Description: The SKUA project supports research in astronomy by developing “a distributed network of semantically aware shared annotation services (in the form of RDF stores). This 'semantic layer' will support a cluster of applications that will either directly support users in finding and recovering useful resources, or indirectly support them by enhancing user-facing applications.” “The SKUA project aims to address a practical problem, and by exploiting lightweight but non-trivial semantics, providing an interface which developers can immediately integrate into their software. The SKUA project's principal product is a 'SAC' (semantic annotation collection), which is a local or remote service that provides a simple RESTful interface for storing annotations of various kinds, including bookmarks, ratings ('paper X is good') or classifications ('object X is a quasar').” “These annotations are stored in the SAC as RDF, and can subsequently be retrieved using SPARQL queries...” (Sandars, 2009).
- Categories:
 - i) Tool->Explanatory search;
 - ii) Tool->Semantic social digital annotation system->Domain specific
- Tools: Software including Qsac, Qsac Client and annotation services
- Status: Under development
- URL: <http://www.myskua.org>

Project 6

- Name: SWAD-Europe (Semantic Web Advanced Development for Europe)
- Organization: EU-funded project, part of the [Information Society Technologies](#) (IST-7) programme
- Description: The SWAD-Europe project ran from May 2002 to October 2004, and aimed to support W3C's Semantic Web initiative in Europe, providing targeted research, demonstrations and outreach to ensure Semantic Web technologies move into the mainstream of networked computing. Basically this project will provide experience and information for those who develop Semantic Web and digital libraries in Europe.
- Categories:
 - i) Tool->Explanatory search
 - ii) Best practice->Semantic Web systems for libraries->Research;
 - iii) Best practice->Semantic Web systems for libraries->Digital project management framework
- Tools:
- Status: Completed during 2002-2004
- URL: <http://www.w3.org/2001/sw/Europe/>

Project 7

- Name: CoinS (OpenURL Context Objects in Spain)
- Organization:
- Description: The project is not totally related to semantic web. However it uses OpenURL and embedding citation data in HTML. According the project web site,

- “Embedded citation metadata in web content may be useful in many other ways. It's not hard to imagine specialized indexing and search systems which make use of the embedded information to deliver new types of information retrieval services. "Semantic Web" systems could use embedded metadata to extract knowledge from large collections of documents.” Therefore it has potential use for semantic digital library development.
- Categories:
 - i)* Tool->Digital Archiving->Metadata;
 - ii)* Tool->Publishing->Converter->Metadata
 - Tools: COinS Generator
 - Status: Under development
 - URL: <http://ocoins.info/>

Project 8

- Name: UBit 2010
- Organization: the library of the Norwegian University of Science and Technology
- Description: The library of the Norwegian university of Science and Technology, NTNU Library started this project. According to a statement on the project web site, the aims of the project includes: 1. Use of Web-based library services at NTNU has increased by 50% (from the current 2008 level); 2. NTNU Library offers a wide array of library services based on new Web technologies*; 3. NTNU Library is the leading library in Norway for Web-based library technical skills and services. This seems to be an official attempt at improving library services. Semantic Web is not mentioned, but “new Web technologies” may include semantic development.
- Category: Tool->Explanatory search
- Tools:
- Status: Under development.
- URL: <http://ubit.blogs.bibsys.no/ubit2010/>

Project 9

- Name: SWED-Europe: Semantic Web Environmental Directory
- Organization: European Union
- Description: Part of a larger project SWAD-Europe. The Semantic Web Environment Directory (SWED) project is a SWAD-E demonstrator aimed at illustrating the nature of the semantic web and the practical issues in making use of it. It is a research project to experiment with semantic web. The focus of the demonstration is to create a directory of UK based environmental organizations.
- Categories:
 - i)* Tool->Explanatory search;
 - ii)* Tool->Semantic social digital annotation system->Domain specific
- Tools: SWED
- Status: Completed
- URL: <http://www.swed.org.uk/swed/index.html>

Project 10

- Name: JeromeDL (Social semantic digital library & e-Library with semantics)
- Organization: the Main Library of Gdansk University of Technology and DERI International
- Description: The project serves two purposes: it allows institutions to publish documents on the Web and allows users to contribute semantic annotations. The project used MarcOnt for bibliographical information, JeromeDL ontology for description of resource structure, and FOAFRealm ontology for information about user profile and authors. It has many other semantic features, including semantic searching.
- Categories:
 - i)* Tool->Linked data->Semantified digital libraries;
 - ii)* Tool->Open source->RDF;
 - iii)* Tool->Content management systems;
 - iv)* Best practice->Semantic Web systems for libraries
- Tools: JeromeDL
- Status: Under development
- URL: <http://www.jeromedl.org/>

Project 11

- Name: BRICKS
- Organization: European Commission
- Description: It is a software that you can download and install as open source and multilingual. It uses service oriented architecture (SAO: A system for linking resources on demand. In an SOA, resources are made available to other participants in the network as independent services) to connect digital resources/digital libraries together seamlessly on the Internet. Each resource or digital library has to install BRICKS as Bnode. BRICKS has several components: Content Management, Collection Management, Metadata Management, Digital Rights Management, User management, Authentication management, and security manager. It may require local client on the desktop to access those services. Its semantic features include: 1. Semantic searching; 2. Its metadata description uses RDF in Metadata Manager; 3. Schema definitions uses OWL in metadata Manager; 4. SPARQL (metadata queries over ontology and graph patterns in Metadata Manager
- Categories:
 - i)* Tool->Linked data->Semantified digital libraries;
 - ii)* Tool->Open source->RDF;
 - iii)* Tool->Open source->Content management systems;
 - iv)* Best practice->Semantic Web systems for libraries
- Tools: BRICKS
- Status: Under development
- URL: <http://www.brickcommunity.org>

Project 12

- Name: Corrib
- Organization: DERI International (a commercial company) and Gdansk University of Technology in Poland.
- Description: It is a cluster of projects. Its main projects include JeromeDL, FOAFRealm, MARCont, and some other non-library related semantic projects
- Categories:
 - i)* Best practice->Semantic Web systems for libraries;
 - ii)* Tool->Open source->RDF;
 - iii)* Tool->Content management systems
- Tools: JeromeDL, FOAF Realm, MARCont
- Status: Under development
- URL: <http://www.corrib.org>

Project 13

- Name: DELOS
- Organization: European Commission
- Description: The full name of the project is DELOS Network of Excellence (A network of excellence on digital libraries). The project focuses on Reference Management Module for future digital libraries. The following mission statement is from the project official web site: "The DELOS network is conducting a joint program of activities aimed at integrating and coordinating the ongoing research activities of the major European teams working in Digital Library - related areas with the goal of developing the next generation Digital Library technologies. DELOS also aims also at disseminating knowledge of digital library technologies to many diverse application domains." Semantic Web should be included in its mission even though it is not mentioned here.
- Categories:
 - i)* Best practice->Semantic Web systems for libraries->Research;
 - ii)* Best practice->Semantic Web systems for libraries->Digital project management framework
- Tools:
- Status: under development
- URL: <http://www.delos.info>

Project 14

- Name: Greenstone (the New Zealand Digital Library Project)
- Organization: the University of Waikato, and distributed in cooperation with UNESCO and the Human Info NGO in Belgium.
- Description: A suite of software for building and distributing digital library collections. It provides a new way of organizing information and publishing it on the Internet in the form of a fully-searchable, metadata-driven digital library. Greenstone version 3 was released with semantic technologies added. Its semantic features include Alerting service (alerting semantics), FRBR implementation, and Location based access. It is open source and

- multilingual. It runs on all versions of Windows, Unix/Linux, and Mac OS-X. The statistics at the project's web site indicates that 4500 sites are running Greenstone.
- Categories:
 - i)* Tool->Open source->RDF;
 - ii)* Tool->Digital Archiving->Metadata;
 - iii)* Tool->Content management systems;
 - iv)* Tool->Explanatory search;
 - v)* Tool->Linked data
 - Tools: 1. Metadata editor; 2. Document Plug-Ins for PDF, Postscript, and Word; 3. Metadata Plug-Ins for XML, MARC, METS, DSpace, SRW, and OAI; 4. Multimedia Plug-Ins for Images, and MP3 audio; 5. Generic Plug-Ins that can be configured for other media formats such as MPEG.
 - Status: Completed, but new releases
 - URL: www.greenstone.org

Project 15

- Name: Europeana
- Organization: European Commission and some other European organizations
- Description: This website created by the project is a prototype. Europeana Version 1.0 is being developed and will launch in 2010 with links to over 10 million digital objects. A European version of World Digital Library by the LC in the US, right now it is linked to 4 million digital items such as images, sounds, texts and videos. It is the creation of a virtual European library, aiming to make Europe's cultural and scientific resources accessible for all. The project uses a semantic search engine and it is still under development.
- Categories:
 - i)* Tool->Linked data->OPAC;
 - ii)* Tool->Open source->RDF;
 - iii)* Tool->Content management systems
- Tools: Europeana
- Status: 2 year project to finish the prototype. Went live in 2008. May make it operational in the future.
- URL: <http://www.europeana.eu/portal/>