

LINKED ARCHIVAL METADATA: A GUIDEBOOK

Executive Summary

=====

[The Executive Summary will list core objectives, anticipated outcomes, and implications that will provide administrators or other senior leaders with the information that they will need in order to understand the benefits and potential costs of this path.]

Introduction

=====

Linked Archival Metadata: A Guidebook provides archivists with an overview of the current linked data landscape, define basic concepts, identify practical strategies for adoption, and emphasize the tangible payoffs for archives implementing linked data. It focuses on clarifying why archives and archival users can benefit from linked data and will identify a graduated approach to applying linked data methods to archival description.

The Guidebook is a product of the Linked Archival Metadata planning project (LiAM), led by the Digital Collections and Archives at Tufts University and funded by the Institute of Museum and Library Services (IMLS). LiAM's goals include defining use cases for linked data in archives and providing a roadmap to describe options for archivists intending to share their description using linked data techniques.

Why linked data, and why now?

Linked data, or more recently referred to as "linked open data" for reasons to be explained later, is a proposed technique for generating new knowledge. It is intended to be a synergy between people and sets of agreed upon computer systems that when combined will enable both people and computers to discover and build relationships between seemingly disparate data and information to create and discover new knowledge. In a nutshell, this is how it works. People possess data and information. They encode that data and information in any number of formats easily readable by computers. They then put then make the encoded data and information available on the Web. Computers are then employed to systematically harvest the encoded data. Since the data is easily readable, the computers store the data locally and look for similarly encoded things in other locally stored data sets. When similar items are identified relationships can be inferred between the items as well as the other items in the data set. To people, some of these relationships may seem obvious and "old hat". On the other hand, since the data sets can be massive, relationships that were never observed previously may come to light, thus new knowledge is created.

Some of this knowledge may be trivial. For example, there might be a data set of places -- places from all over the world including things like geographic coordinates, histories of the places, images, etc. There might be another data set of people. Each person may be described using their name, their place of birth, and a short biography. These data sets may contain ten's of thousands of items each. Using linked data it would be possible to cross reference the people with the places to discover who might have met whom when and where. Some people may have similar ideas, and those ideas may have been generated in a particular place. Linked data may help in discovering who was in the same place at the same time and the researcher may be better able to figure out how a particular idea came to fruition.

Here's an example hitting closer to the home of archives and archivists. Suppose most archival finding aids were written in a format easily readable by computers. Let's call this format Encoded Archival Description. Let's suppose these finding aids were made available on the Web. Let's suppose one or more

computers crawled these archival sites harvesting the finding aids. Once done a computer program could be used to find all the occurrences of particular name and generate a virtual finding aid that is more complete and more comprehensible than any single finding aid on that particular person.

The amount of data and information accessible today is greater in size than it has ever been in human history. Using our traditional techniques of reading, re-reading, writing, discussing, etc. is more than possible to learn new things about the state of the world, the universe, and the human condition. By exploiting the current state of computer technology is possible to expand upon our traditional techniques and possibly exxcellerate the mass of knowledge.

How to use the Guidebook

The structure of the Guidebook supports readers moving through the text in a variety of ways. Like a travel book, it provides useful high-level information for users who only need the basics, as well as in-depth information for those planning an extended stay in LOD-land. The Guidebook is intentionally named, and will draw from the genre of actual travel guides (Fodors, etc.) providing readers easy access to both high-level information (know before you go, what to see if you're only there for a day) as well as in-depth details of for those staying in one place longer.

Synopses of the use cases developed by the LiAM project will be interspersed throughout the Guidebook to illustrate and frame the text. Each use case will be briefly described in 100-200 words with links to the full use cases on the LiAM website.

An initial release of the Guidebook will be in the form of a PDF document to be delivered to IMLS in fulfillment of the LiAM planning grant requirements as well as being shared with the public. However, the Guidebook's ongoing vitality will benefit from a more dynamic publication environment, and we therefore plan to publish it in a wiki connected to a code repository. This combination will enable updating of the resource to reflect changes in the field as well as providing a mechanism for sharing tools, scripts, and other code related to the project.

Much of the rest of the Guidebook, while providing a concise overview of today's linked data landscape and needs, would require ongoing updates, maintenance, and enhancement to describe implementation of LOD in the archival community over time.

Linked Data for Archives: a Primer, or "What is Linked Data and why should I care?"

=====

Linked Data is a process for manifesting the ideas behind Semantic Web. The Semantic Web is about encoding data, information, and knowledge in computer-readable fashions, making these encodings accessible on the World Wide Web, allowing computers to crawl the encodings, and finally, employing reasoning engines against them for the purpose of discovering and creating new knowledge. The canonical article describing this concept was written by Tim Berners-Lee, James Hendler, and Ora Lassila in 2001.

In 2006 Berners-Lee more concretely described how to make the Semantic Web a reality in a text called "Linked Data -- Design Issues". In it he outlined four often-quoted expectations for implementing the Semantic Web. Each of these expectations are listed below along with some elaborations:

1. "Use URIs as names for things" - URIs (Universal Resource Identifiers) are unique identifiers, and they are expected to have the same shape as URLs (Universal Resource Locators). These identifiers are expected to represent things such as people, places, institutions, concepts, books, etc. URIs are monikers or handles for real world or imaginary objects.

2. "Use HTTP URIs so that people can look up those names." - The URIs are expected to look and ideally function on the World Wide Web through the Hypertext Transfer Protocol (HTTP), meaning the URI's point to

things on Web servers.

3. "When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)" - When URIs are sent to Web servers by Web browsers (or "user-agents" in HTTP parlance), the response from the server should be in a conventional, computer readable format. This format is usually a version of RDF (Resource Description Framework) -- a notation looking much like a rudimentary sentence composed of a subject, predicate, and object.

4. "Include links to other URIs. So that they can discover more things." - Simply put, try very hard to use URIs that other people have used. This way the relationships you create can literally be linked to the relationships other people have created. These links may represent new knowledge.

In the same text ("Linked Data -- Design Issues") Berners-Lee also outlined a sort of reward system -- sets of stars -- for levels of implementation. Unfortunately, nobody seems to have taken up the stars very seriously. A person gets:

- * 1 star for making data available on the web (whatever format)
but with an open licence, to be Open Data
- * 2 stars for making the data machine-readable structured data
(e.g. excel instead of image scan of a table)
- * 3 stars for making the data available in non-proprietary
format (e.g. CSV instead of excel)
- * 4 stars for using open standards from W3C (RDF and SPARQL)
to identify things, so that people can point at your stuff
- * 5 stars for linking your data to other people's data to
provide context

The whole idea works like this. Suppose I assert the following statement:

The Declaration Of Independence was authored by Thomas Jefferson.

This statement can be divided into three parts. The first part is a subject (Declaration Of Independence). The second part is a predicate (was authored by). The third part is an object (Thomas Jefferson). In the language of the Semantic Web and Linked Data, these combined parts are called a triple, and they are expected to denote a fact. Triples are the heart of RDF.

Suppose further that the subject and object of the triple are identified using URIs (as in Expectations #1 and #2, above). This would turn our assertion into something like this with carriage returns added for readability:

```
http://www.archives.gov/exhibits/charters/declaration_transcript.html
was authored by
http://www.worldcat.org/identities/lccn-n79-89957
```

Unfortunately, this assertion is not easily read by a computer. Believe it or not, something like the XML below is much more amenable, and if it were the sort of content returned by a Web server to a Web browser (read "user-agent"), then it would satisfy Expectations #3 and #4 because the notation is standardized and because it points to other people's content:

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description rdf:about="http://www.archives.gov/exhibits/charters/declaration_transcript.html">
```

```
<dc:creator>http://www.worldcat.org/identities/lccn-n79-89957</dc:creator>
</rdf:Description>
</rdf:RDF>
```

Suppose we had a second assertion:

Thomas Jefferson was a man.

In this case, the subject is "Thomas Jefferson". The predicate is "was". The object is "man". This assertion can be expressed in a more computer-readable fashion like this:

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/">
  <rdf:Description rdf:about="http://www.worldcat.org/identities/lccn-n79-89957">
    <foaf:gender>male</foaf:gender>
  </rdf:Description>
</rdf:RDF>
```

Looking at the two assertions, a reasonable person can deduce a third assertion, namely, the Declaration Of Independence was authored by a man. Which brings us back to the point of the Semantic Web and Linked Data. If everybody uses URIs (read "URLs") to describe things, if everybody denotes relationships (through the use of predicates) between URIs, if everybody makes their data available on the Web in standardized formats, and if everybody uses similar URIs, then new knowledge can be deduced from the original relationships.

Unfortunately and to-date too little Linked Data has been made available and/or too few people have earned too few stars to really make the Semantic Web a reality. The purpose of this guidebook is to provide means for archivists to do their part, make their content available on the Semantic Web through Linked Data, all in the hopes of facilitating the discovery of new knowledge. On our mark. Get set. Go!

There are a number of challenges in the process. Some of them are listed below, and some of them have been alluded to above:

- * Create useful LOD, meaning, create LOD that links to other LOD. LOD does not live in a world by itself. Remember, the "L" stands for "linked". For example, try to include URIs that are the URIs used on other LOD data sets. Sometimes this is not possible, for example, le with the names of people in archival materials. When possible, they used VIAF, but other times they needed to create their own URI denoting an individual.
- * There is a level of rigor involved in creating the data model, and there may be many discussions regarding semantics. For example, what is a creator? Or, when is a term intended to be an index term as opposed reference. When does one term in one vocabulary equal a different term in a different vocabulary?
- * Balance the creation of your own vocabulary with the need to speak the language of others using their vocabulary.
- * Consider "fixing" the data as it comes in or goes out because it might not be consistent nor thorough.
- * Provenance is an issue. People – especially scholars – will want to know where the LOD came from and whether or not it is authoritative. How to solve or address this problem? The jury is still out on this one.
- * Creating and maintaining LOD is difficult because it requires the skills of a number of different types of people. Computer programmers. Database designers. Subject experts. Metadata specialists. Archivists. Etc. A team is all but necessary.

Objectives

[Management, access, and use and linked data affordances]

Overview of linked data concepts and vocabulary

Linked Data is a process for systematically and methodically exposing metadata on the Web. In many ways, it is the re-articulation of a thing called the Semantic Web first outlined more than a decade ago. Linked Data (and the Semantic Web) are efforts to increase the "sphere of knowledge" through the use of computer technology.

Increasingly you will hear of linked data being qualified as "linked open data". The "open" qualifier alludes to the important distinctions between truly free data/information and licensed data/information which comes with some strings attached. Truly "open" linked data comes with no licensing restrictions.

When you hear of linked data and the Semantic Web, the next thing you often hear is "RDF" or "Resource Description Framework". First and foremost, RDF is a way of representing knowledge. It does this through the use of assertions (think, "sentences") with only three parts: 1) a subject, 2) a predicate, and 3) an object. Put together, these three things create things called "triples". The subject of each assertion is expected to be a Universal Resource Identifier (or URI, but think URL), and this URI is expected to represent a thing -- anything. (Really, anything.) The predicate is some sort of relationship such as equals or is a sub-part of or contains or is a description of, or is the name of, etc. Predicates are the vocabulary of linked data, and you will find an abundance of vocabularies from which to choose when creating Linked Data. Finally, objects come in two forms: 1) more URIs (pointers to things) or literal values such the names of people, places, or things. Examples of literals include "Lancaster, PA", "Thomas Jefferson", or "Musée d'Orsay".

RDF is not to be confused with RDF/XML or another other type of RDF "serialization". Remember, RDF describes triples, but it does not specify how the triples are express or written down. On the other hand, RDF/XML is an XML syntax for expressing RDF. Some people think RDF/XML is too complicated and too verbose. Consequently, other serializations have manifested themselves including N3 and Turtle.

Brief overview of the history of LOD-LAM

Examples

Linked Data Today

=====

Projects

[Brief descriptions with an emphasis on tangible benefits and outcomes of each]

Trends in LOD-LAM

With great interest I read the Spring/Summer issue of Information Standards Quarterly where there were a number of articles pertaining to linked open data in cultural heritage institutions. [0] Of particular interest to me where the various loosely enumerated challenges of linked open data. Some of them included:

- * the apparent Tower Of Babel when it comes to vocabularies used to describe content
- * dirty, inconsistent, or wide varieties of data integrity
- * persistent URIs
- * the "chicken & egg" problem of why linked data if there is no killer application

Getting Started: Strategies and Steps

=====

Defining your strategy

[Articulate goals, objectives, and metrics to measure success.]

Is your archival description LOD-ready?

Identify building blocks

[metadata components in archival description that are (or nearly are) ready for linking.]

Readiness

[Making small changes in practice to make your description LOD-ready.]

What you can do now if you have

EAD

EAC-CPF

MARC

METS, MODS, and perhaps more

[This section will provide illustrative examples of what can be done with the linked data building blocks already included in archival description.]

On Your Way: Next Steps

=====

Integration into daily practice

Three Cs: Cleanup, Conversion, Consistency

Creating and maintaining metadata is a never-ending process. The items being described can always use elaboration. Collections may increase in size. Rights applied against content may change. Things become digitized, or digitized things are migrated from one format to another. Because of these sorts of things and many others, cleanup, conversion, and consistency are something every metadata specialist needs to keep in mind.

Cleanup, conversion, and consistency means many things. Does all of your metadata use the same set of one or more vocabularies? Are things spelled correctly? Maybe you used abbreviations in one document but spelled things out in another? Have you migrated your JPEG images to JPEG2000 or TIFF formats? Maybe the EAD DTD has been updated, and you want (need) to migrate your finding aids from one XML format to another? Do all of your finding aids exhibit the same level of detail; are some "thinner" than others? Have you used one form of a person's name in one document but used another form in a different document? The answers to these sorts of questions point to the need for cleanup, conversion, and consistency.

Tools

* Fusion Tables (<http://www.google.com/drive/apps.html>) - Bust your data out of its silo! Combine it with other data on the web. Collaborate, visualize and share.

* OpenRefine (<https://github.com/OpenRefine/>) - OpenRefine is a free, open source power tool for working with messy data and improving it

Looking Ahead: Advanced Tools and Visualizations

=====

Tools for archivists (data preparation, cleanup, management)

What's available now

* ckan (<http://ckan.org>) - The open source data portal software

* CouchDB (<http://couchdb.apache.org>) - CouchDB is a database that completely embraces the web. Store your data with JSON documents. Access your documents with your web browser, via HTTP. Query, combine, and transform your documents with JavaScript. CouchDB works well with modern web and mobile apps. You can even serve web apps directly out of CouchDB. And you can distribute your data, or your apps, efficiently using

CouchDB's incremental replication. CouchDB supports master-master setups with automatic conflict detection.

- * Curl (<http://curl.haxx.se>) - curl is a command line tool for transferring data with URL syntax, supporting DICT, FILE, FTP, FTPS, Gopher, HTTP, HTTPS, IMAP, IMAPS, LDAP, LDAPS, POP3, POP3S, RTMP, RTSP, SCP, SFTP, SMTP, SMTPS, Telnet and TFTP. curl supports SSL certificates, HTTP POST, HTTP PUT, FTP uploading, HTTP form based upload, proxies, cookies, user+password authentication (Basic, Digest, NTLM, Negotiate, kerberos...), file transfer resume, proxy tunneling and a busload of other useful tricks.

- * Data.gov (<http://www.data.gov>) - The purpose of Data.gov is to increase public access to high value, machine readable datasets generated by the Executive Branch of the Federal Government. As a priority Open Government Initiative for President Obama's administration, Data.gov increases the ability of the public to easily find, download, and use datasets that are generated and held by the Federal Government. Data.gov provides descriptions of the Federal datasets (metadata), information about how to access the datasets, and tools that leverage government datasets. The data catalogs will continue to grow as datasets are added. Federal, Executive Branch data are included in the first version of Data.gov.

- * Data.Gov.uk (<http://data.gov.uk>) - The [British] Government is releasing public data to help people understand how government works and how policies are made. Some of this data is already available, but data.gov.uk brings it together in one searchable website. Making this data easily available means it will be easier for people to make decisions and suggestions about government policies based on detailed information.

- * Datahub (<http://datahub.io/>) - the free, powerful data management platform from the Open Knowledge Foundation

- * DBpedia (<http://dbpedia.org/>) - DBpedia is a crowd-sourced community effort to extract structured information from Wikipedia and make this information available on the Web. DBpedia allows you to ask sophisticated queries against Wikipedia, and to link the different data sets on the Web to Wikipedia data. We hope that this work will make it easier for the huge amount of information in Wikipedia to be used in some new interesting ways. Furthermore, it might inspire new mechanisms for navigating, linking, and improving the encyclopedia itself.

- * ead2rdf (<http://data.archiveshub.ac.uk/xslt/ead2rdf.xsl>) - The "transform" process is currently performed using XSLT to read an EAD XML document and output RDF/XML, and the current version of the stylesheet is now available:

- * elasticsearch (<http://www.elasticsearch.org>) - flexible and powerful open source, distributed real-time search and analytics engine for the cloud

- * GeoNames (<http://www.geonames.org>) - The GeoNames geographical database covers all countries and contains over eight million placenames that are available for download free of charge.

- * MySQL (<http://www.mysql.com>) - The world's most popular open source database

- * Protégé (<http://protege.stanford.edu>) - Protégé is a free, open source ontology editor and knowledge-base framework The Protégé platform supports modeling ontologies via a web client or a desktop client. Protégé ontologies can be developed in a variety of formats including OWL, RDF(S), and XML Schema Protégé is based on Java, is extensible, and provides a plug-and-play environment that makes it a flexible base for rapid prototyping and application development.

- * Solr (<http://lucene.apache.org/solr/>) - Solr is the popular, blazing fast open source enterprise search platform from the Apache LuceneTM project. Its major features include powerful full-text search, hit highlighting, faceted search, near real-time indexing, dynamic clustering, database integration, rich document (e.g., Word, PDF) handling, and geospatial search. Solr is highly reliable, scalable and fault tolerant, providing distributed indexing, replication and load-balanced querying, automated failover and recovery, centralized configuration and more. Solr powers the search and navigation features of many of

the world's largest internet sites.

- * TemaTres (<http://www.vocabularyserver.com>) - The open source way to manage formal representations of knowledge

- * ULAN (<http://www.getty.edu/research/tools/vocabularies/ulan/>) - The Union List of Artist Names ® (ULAN), the Getty Thesaurus of Geographic Names ® (TGN), the Art & Architecture Thesaurus ® (AAT), and the Cultural Objects Name Authority ® (CONA) (in development) are structured vocabularies that can be used to improve access to information about art, architecture, and material culture.

- * VIAF (<http://viaf.org>) - The VIAF™ (Virtual International Authority File) combines multiple name authority files into a single OCLC-hosted name authority service. The goal of the service is to lower the cost and increase the utility of library authority files by matching and linking widely-used authority files and making that information available on the Web.

- * W3C RDF Validation Service (<http://www.w3.org/RDF/Validator/>) - Enter a URI or paste an RDF/XML document into the text field above. A 3-tuple (triple) representation of the corresponding data model as well as an optional graphical visualization of the data model will be displayed.

Gaps: What is needed

Tools for users: visualizations, interfaces, etc.

What's available now

- * D3.js (<http://d3js.org>) - D3.js is a JavaScript library for manipulating documents based on data. D3 helps you bring data to life using HTML, SVG and CSS. D3's emphasis on web standards gives you the full capabilities of modern browsers without tying yourself to a proprietary framework, combining powerful visualization components and a data-driven approach to DOM manipulation.

- * Gephi (<http://gephi.org>) - Gephi is an interactive visualization and exploration platform for all kinds of networks and complex systems, dynamic and hierarchical graphs.

- * Tableau Public (<http://www.tableausoftware.com/public>) - With Tableau Public you can create interactive graphs, dashboards, maps and tables from virtually any data and embed them on your website or blog in minutes.

Gaps: What is needed

Further reading

=====

This is a list of links and citations to get one started on Linked Open Data

- * "About - LOCAH Linked Archives Hub Test Dataset." Accessed August 16, 2013.
<http://data.archiveshub.ac.uk/about.html>.

- * Becker, Christian, and Christian Bizer. "Exploring the Geospatial Semantic Web with DBpedia Mobile." Web Semantics: Science, Services and Agents on the World Wide Web 7, no. 4 (December 2009): 278-286.
doi:10.1016/j.websem.2009.09.004.

- * Belleau, François, Marc-Alexandre Nolin, Nicole Tourigny, Philippe Rigault, and Jean Morissette. "Bio2RDF: Towards a Mashup to Build Bioinformatics Knowledge Systems." *Journal of Biomedical Informatics* 41, no. 5 (October 2008): 706–716. doi:10.1016/j.jbi.2008.03.004.
- * Berners-Lee, Tim. "Linked Data - Design Issues." Accessed August 4, 2013. <http://www.w3.org/DesignIssues/LinkedData.html>.
- * Berners-Lee, Tim, James Hendler, and Ora Lassila. "The Semantic Web." *Scientific American* 284, no. 5 (May 2001): 34–43. doi:10.1038/scientificamerican0501-34.
- * "BIBFRAME.ORG :: Bibliographic Framework Initiative - Overview." Accessed September 2, 2013. <http://bibframe.org/>.
- * Bizer, Christian, Tom Heath, and Tim Berners-Lee. "Linked Data - The Story So Far:" *International Journal on Semantic Web and Information Systems* 5, no. 3 (33 2009): 1–22. doi:10.4018/jswis.2009081901.
- * Bizer, Christian, Jens Lehmann, Georgi Kobilarov, Sören Auer, Christian Becker, Richard Cyganiak, and Sebastian Hellmann. "DBpedia - A Crystallization Point for the Web of Data." *Web Semantics: Science, Services and Agents on the World Wide Web* 7, no. 3 (September 2009): 154–165. doi:10.1016/j.websem.2009.07.002.
- * Buneman, Peter, Sanjeev Khanna, and Tan Wang-Chiew. "Why and Where: A Characterization of Data Provenance." In *Database Theory – ICDT 2001*, edited by Jan Bussche and Victor Vianu, 1973:316–330. Berlin, Heidelberg: Springer Berlin Heidelberg. Accessed September 4, 2013. http://www.springerlink.com/index/10.1007/3-540-44503-X_20.
- * Carroll, Jeremy J., Christian Bizer, Pat Hayes, and Patrick Stickler. "Named Graphs." *Web Semantics: Science, Services and Agents on the World Wide Web* 3, no. 4 (December 2005): 247–267. doi:10.1016/j.websem.2005.09.001.
- * "Ckan - The Open Source Data Portal Software." Accessed August 14, 2013. <http://ckan.org/>.
- * "Content Negotiation." Wikipedia, the Free Encyclopedia, July 2, 2013. https://en.wikipedia.org/wiki/Content_negotiation.
- * Correndo, Gianluca, Manuel Salvadores, Ian Millard, Hugh Glaser, and Nigel Shadbolt. "SPARQL Query Rewriting for Implementing Data Integration over Linked Data." 1. ACM Press, 2010. doi:10.1145/1754239.1754244.
- * Cudré-Mauroux, Philippe, Parisa Haghani, Michael Jost, Karl Aberer, and Hermann De Meer. "idMesh: Graph-based Disambiguation of Linked Data." 591. ACM Press, 2009. doi:10.1145/1526709.1526789.
- * "Curl and Libcurl." Accessed August 6, 2013. <http://curl.haxx.se/>.
- * Cyganiak, Richard, Holger Stenzhorn, Renaud Delbru, Stefan Decker, and Giovanni Tummarello. "Semantic Sitemaps: Efficient and Flexible Access to Datasets on the Semantic Web." In *The Semantic Web: Research and Applications*, edited by Sean Bechhofer, Manfred Hauswirth, Jörg Hoffmann, and Manolis Koubarakis, 5021:690–704. Berlin, Heidelberg: Springer Berlin Heidelberg. Accessed September 4, 2013. http://www.springerlink.com/index/10.1007/978-3-540-68234-9_50.
- * David Beckett. "Turtle." Accessed August 6, 2013. <http://www.w3.org/TR/2012/WD-turtle-20120710/>.
- * Dunsire, Gordon, Corey Harper, Diane Hillmann, and Jon Phipps. "Linked Data Vocabulary Management: Infrastructure Support, Data Integration, and Interoperability." *Information Standards Quarterly* 24, no. 2/3 (2012): 4. doi:10.3789/isqv24n2-3.2012.02.
- * Elliott, Thomas, Sebastian Heath, and John Muccigrosso. "Report on the Linked Ancient World Data

Institute." *Information Standards Quarterly* 24, no. 2/3 (2012): 43. doi:10.3789/isqv24n2-3.2012.08.

* Fons, Ted, Jeff Penka, and Richard Wallis. "OCLC's Linked Data Initiative: Using Schema.org to Make Library Data Relevant on the Web." *Information Standards Quarterly* 24, no. 2/3 (2012): 29. doi:10.3789/isqv24n2-3.2012.05.

* Gruber, Ethan. "Eaditor." GitHub. Accessed August 12, 2013. <https://github.com/ewg118/eaditor>.

* Halpin, Harry, Patrick J. Hayes, James P. McCusker, Deborah L. McGuinness, and Henry S. Thompson. "When owl:sameAs Isn't the Same: An Analysis of Identity in Linked Data." In *The Semantic Web – ISWC 2010*, edited by Peter F. Patel-Schneider, Yue Pan, Pascal Hitzler, Peter Mika, Lei Zhang, Jeff Z. Pan, Ian Horrocks, and Birte Glimm, 6496:305–320. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. http://www.springerlink.com/index/10.1007/978-3-642-17746-0_20.

* Hartig, Olaf. "Querying Trust in RDF Data with tSPARQL." In *The Semantic Web: Research and Applications*, edited by Lora Aroyo, Paolo Traverso, Fabio Ciravegna, Philipp Cimiano, Tom Heath, Eero Hyvönen, Riichiro Mizoguchi, Eyal Oren, Marta Sabou, and Elena Simperl, 5554:5–20. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. http://www.springerlink.com/index/10.1007/978-3-642-02121-3_5.

* Hartig, Olaf, Christian Bizer, and Johann-Christoph Freytag. "Executing SPARQL Queries over the Web of Linked Data." In *The Semantic Web – ISWC 2009*, edited by Abraham Bernstein, David R. Karger, Tom Heath, Lee Feigenbaum, Diana Maynard, Enrico Motta, and Krishnaprasad Thirunarayan, 5823:293–309. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. http://www.springerlink.com/index/10.1007/978-3-642-04930-9_19.

* Heath, Tom, and Christian Bizer. "Linked Data: Evolving the Web into a Global Data Space." *Synthesis Lectures on the Semantic Web: Theory and Technology* 1, no. 1 (February 9, 2011): 1–136. doi:10.2200/S00334ED1V01Y201102WBE001.

* Heath, Tom, and Enrico Motta. "Revyu: Linking Reviews and Ratings into the Web of Data." *Web Semantics: Science, Services and Agents on the World Wide Web* 6, no. 4 (November 2008): 266–273. doi:10.1016/j.websem.2008.09.003.

* Isaac, Antoine, Robina Clayphan, and Bernhard Haslhofer. "Europeana: Moving to Linked Open Data." *Information Standards Quarterly* 24, no. 2/3 (2012): 34. doi:10.3789/isqv24n2-3.2012.06.

* Kobilarov, Georgi, Tom Scott, Yves Raimond, Silver Oliver, Chris Sizemore, Michael Smethurst, Christian Bizer, and Robert Lee. "Media Meets Semantic Web – How the BBC Uses DBpedia and Linked Data to Make Connections." In *The Semantic Web: Research and Applications*, edited by Lora Aroyo, Paolo Traverso, Fabio Ciravegna, Philipp Cimiano, Tom Heath, Eero Hyvönen, Riichiro Mizoguchi, Eyal Oren, Marta Sabou, and Elena Simperl, 5554:723–737. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. http://www.springerlink.com/index/10.1007/978-3-642-02121-3_53.

* Le-Phuoc, Danh, Axel Polleres, Manfred Hauswirth, Giovanni Tummarello, and Christian Morbidoni. "Rapid Prototyping of Semantic Mash-ups through Semantic Web Pipes." 581. ACM Press, 2009. doi:10.1145/1526709.1526788.

* LiAM. "LiAM: Linked Archival Metadata." Accessed July 30, 2013. <http://sites.tufts.edu/liam/>.

* "Linked Data." Wikipedia, the Free Encyclopedia, July 13, 2013. http://en.wikipedia.org/w/index.php?title=Linked_data&oldid=562554554.

* "Linked Data: Evolving the Web into a Global Data Space." Accessed September 4, 2013. <http://linkeddatabook.com/editions/1.0/>.

* "Linked Open Data." Europeana. Accessed September 12, 2013. <http://pro.europeana.eu/web/guest;jsessionid=09A5D79E7474609AE246DF5C5A18DDD4>.

- * "Linked Open Data in Libraries, Archives, & Museums (Google Group)." Accessed August 6, 2013.
<https://groups.google.com/forum/#!forum/lod-lam>.
- * "Linking Lives | Using Linked Data to Create Biographical Resources." Accessed August 16, 2013.
<http://archiveshub.ac.uk/linkinglives/>.
- * "LOCAH Linked Archives Hub Test Dataset." Accessed August 6, 2013. <http://data.archiveshub.ac.uk/>.
- * LODLAM. "Zotero | Groups > LOD-LAM." Accessed August 6, 2013. <https://www.zotero.org/groups/lod-lam>.
- * "LODLAM - Linked Open Data in Libraries, Archives & Museums." Accessed August 6, 2013.
<http://lodlam.net/>.
- * Möller, Knud, Tom Heath, Siegfried Handschuh, and John Domingue. "Recipes for Semantic Web Dog Food – The ESWC and ISWC Metadata Projects." In *The Semantic Web*, edited by Karl Aberer, Key-Sun Choi, Natasha Noy, Dean Allemang, Kyung-Il Lee, Lyndon Nixon, Jennifer Golbeck, et al., 4825:802–815. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.
http://www.springerlink.com/index/10.1007/978-3-540-76298-0_58.
- * "Notation3." Wikipedia, the Free Encyclopedia, July 13, 2013.
<http://en.wikipedia.org/w/index.php?title=Notation3&oldid=541302540>.
- * "OWL 2 Web Ontology Language Primer." Accessed August 14, 2013.
<http://www.w3.org/TR/2009/REC-owl2-primer-20091027/>.
- * "PELAGIOS: About PELAGIOS." Accessed September 4, 2013.
<http://pelagios-project.blogspot.com/p/about-pelagios.html>.
- * Quilitz, Bastian, and Ulf Leser. "Querying Distributed RDF Data Sources with SPARQL." In *The Semantic Web: Research and Applications*, edited by Sean Bechhofer, Manfred Hauswirth, Jörg Hoffmann, and Manolis Koubarakis, 5021:524–538. Berlin, Heidelberg: Springer Berlin Heidelberg. Accessed September 4, 2013.
http://www.springerlink.com/index/10.1007/978-3-540-68234-9_39.
- * "RDF/XML." Wikipedia, the Free Encyclopedia, July 13, 2013.
<http://en.wikipedia.org/w/index.php?title=RDF/XML&oldid=561972340>.
- * "RDFa." Wikipedia, the Free Encyclopedia, July 22, 2013.
<http://en.wikipedia.org/w/index.php?title=RDFa&oldid=561972699>.
- * "Semantic Web." Wikipedia, the Free Encyclopedia, August 2, 2013.
http://en.wikipedia.org/w/index.php?title=Semantic_Web&oldid=566813312.
- * "SPARQL." Wikipedia, the Free Encyclopedia, August 1, 2013.
<http://en.wikipedia.org/w/index.php?title=SPARQL&oldid=566718788>.
- * "SPARQL 1.1 Overview." Accessed August 6, 2013. <http://www.w3.org/TR/sparql11-overview/>.
- * "Spring/Summer 2012 (v.24 No.2/3) - National Information Standards Organization." Accessed August 6, 2013. <http://www.niso.org/publications/isq/2012/v24no2-3/>.
- * "TemaTres Controlled Vocabulary Server." Accessed August 14, 2013. <http://www.vocabularyserver.com/>.
- * "The Protégé Ontology Editor and Knowledge Acquisition System." Accessed September 2, 2013.
<http://protege.stanford.edu/>.
- * Tim Berners-Lee, James Hendler, and Ora Lassila. "The Semantic Web." Accessed September 4, 2013.

<http://www.scientificamerican.com/article.cfm?id=the-semantic-web>.

* "Transforming EAD XML into RDF/XML Using XSLT." Accessed August 16, 2013.
<http://archiveshub.ac.uk/locah/tag/transform/>.

* "Turtle (syntax)." Wikipedia, the Free Encyclopedia, July 13, 2013.
[http://en.wikipedia.org/w/index.php?title=Turtle_\(syntax\)&oldid=542183836](http://en.wikipedia.org/w/index.php?title=Turtle_(syntax)&oldid=542183836).

* "VIAF." Accessed August 27, 2013. <http://viaf.org/>.

* Volz, Julius, Christian Bizer, Martin Gaedke, and Georgi Kobilarov. "Discovering and Maintaining Links on the Web of Data." In *The Semantic Web - ISWC 2009*, edited by Abraham Bernstein, David R. Karger, Tom Heath, Lee Feigenbaum, Diana Maynard, Enrico Motta, and Krishnaprasad Thirunarayan, 5823:650-665. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009.
http://www.springerlink.com/index/10.1007/978-3-642-04930-9_41.

* Voss, Jon. "LODLAM State of Affairs." *Information Standards Quarterly* 24, no. 2/3 (2012): 41.
doi:10.3789/isqv24n2-3.2012.07.

* W3C. "LinkedData - W3C Wiki." Accessed August 4, 2013. <http://www.w3.org/wiki/LinkedData>.

* "Welcome - the Datahub." Accessed August 14, 2013. <http://datahub.io/>.

* "Welcome to Euclid." Accessed September 4, 2013. <http://www.euclid-project.eu/>.

--

Eric Lease Morgan
September 21, 2013