

Learning Linked Data: Final Project Report

IMLS Grant no.: LG-51-11-0147-11, October 2011 through September 2012

Project Partners:

University of Washington (lead)

Kent State University

University of North Carolina

JES & Company

Part 1: Project Narrative

Project Overview

The "Learning Linked Data" project, funded from October 2011 through September 2012 by a one-year IMLS planning grant led by the Information School at the University of Washington, envisions an online learning environment in support of educating library and museum professionals in the analysis and processing of Linked Data. A core project group of twenty instructors, students, and technology experts met to develop an Inventory of Learning Topics outlining a target set of analytical skills relevant to a wide range of pedagogical contexts. This draft was posted on a blog for input from a larger circle of colleagues and subsequently refined. The group envisions a follow-on project using the finished Inventory to support the creation of tightly-scoped "micro-tutorials" -- screencasts, how-to texts, and annotated code snippets -- linked to well-defined Learning Objectives, and usable in constructing Learning Trajectories customized for either classroom or individualized instruction for a wide range of target audiences.

Description of Project Activities

Linked Data is data that can fit into a "cloud" of interconnected data sources -- whether those sources are published world-readably on the Web (Linked Open Data) or behind corporate or institutional firewalls (Linked Enterprise Data). For our purposes, Linked Data is data published in a form compatible with the Resource Description Framework (RDF) model of the World Wide Web Consortium (W3C). Compatibility with RDF and practices common to Linked Data has become an important objective of most current initiatives for standardization and service development in the library and museum worlds.

"Learning Linked Data," a project funded under the IMLS program National Leadership Grants for Libraries from October 2011 through September 2012, aimed at planning the development of a software-supported environment for learning the principles and practice of Linked Data. The project aimed at designing a "language lab" for the trainers and university faculty who are teaching current and future library and museum professionals. In order to elicit the requirements for this environment, the project recruited a community of twenty active project participants from a wide range of relevant backgrounds, encompassing university LIS faculty and professional trainers, LIS graduate students, software developers, metadata application implementers, instructional technology practitioners, IT consultants, and Semantic Web experts.

February 2012 Workshop

The project convened a meeting of its core participants on 2-3 February 2012 at the Information School on the University of Washington campus in Seattle (see List of Participants in the Appendix). In preparation for the workshop, several concept papers were discussed on a mailing list. The workshop focused on refining and restructuring the project's key text, an Inventory of

Learning Topics. In the Inventory, each learning topic is associated with examples of software tools needed for accomplishing specific analytical or data-processing tasks. After the workshop, the Inventory was posted on a blog for comment by colleagues. A final version of the Inventory, appended to this report, is intended to serve as the starting point of a follow-on project for a Learning Lab.

In-scope for the Learning Lab: "basic" learning topics. Mindful of the risks of taking on too many topics, at too superficial a level, the participants felt that the follow-on project should focus, at least initially, on the fundamentals of understanding and interpreting Linked Data.

Out of scope: Specific pedagogical approaches and learning outcomes. Each participant came to the workshop with particular pedagogical scenarios in mind -- courses for particular learners, from particular backgrounds and experience, and with correspondingly specific expected learning outcomes. An early draft of the Inventory of Learning Topics tried to classify the topics into Beginning, Intermediate, and Advanced levels. However, a Beginning topic for a library science student may be considered Advanced for a computer science student, and vice versa. Narrowing the focus to a specific audience, it was felt, risked favoring one audience at the expense of others. Any particular syllabus offered by the project would inevitably need to be customized for a diverse range of instructional goals and instructor preferences and would, in complex ways, go quickly out of date. Specific pedagogical approaches customized for particular audiences were therefore considered to be out of scope for the Learning Lab per se; rather, the Learning Lab was seen as something that would support the construction of courses or curricula according to a wide range of pedagogical approaches.

Out of scope: Support for data modeling. Sound modeling is an essential foundation for creating high-quality Linked Data. However, the group recognized that methodologies for designing data models from scratch are very diverse and that selection among them depends on target audience. Students may approach the problem with backgrounds in databases, UML modeling, formal ontologies, or traditional knowledge organization systems. Due to the broad diversity of approaches (and relative lack of suitably mature software tools), the group felt that support for data modeling should be a second-order priority in the follow-on project.

Partially in-scope: implementation practicalities. While the Learning Lab was envisioned as focusing on the fundamentals of understanding and interpreting Linked Data, rather than on particular implementation technologies -- subject, as they are, to continual change and obsolescence -- the project recognized that several topics straddled the line between supporting instruction about conceptual underpinnings versus imparting implementation skills. For example:

- **Publishing linked data "as Linked Data".** Simply publishing a dataset on the Web as a ZIP file does not make the contents of that dataset available for linking. The publication of Linked Data involves exposing one or more RDF-compatible representations of the contents of a dataset, possibly via content negotiation based on browser settings or user preferences.

Publication is a practical implementation skill, but the concepts underlying content negotiation are an academic topic.

- **Visualizing Linked Data.** Visualization of Linked Data relationships was seen not just as a software function, but as a Learning Topic in itself. At the most basic level, node-and-arc diagrams are used to visualize RDF graphs -- webs of RDF statements. However, diagrams can be generated at higher levels of granularity (as in "Linked Data Cloud" diagrams), and sophisticated statistical techniques can be used to depict clusters of related resources or generate "cloud" diagrams of high-level relationships between datasets.
- **Storing Linked Data.** Linked Data can be exposed on the Web as plain-text files holding RDF triples "serialized" in one of several interchangeable syntaxes; in the Inventory, these fall under the learning topic "Creating and manipulating RDF data." RDF triples can be indexed for retrieval by storing them in specifically optimized databases ("triple stores"). Setting up a triple store is relatively straightforwardly an implementation-related skill. However, the participants recognized that the rapid evolution of approaches and software for storage could potentially be seen in the follow-on project as a learning topic in its own right.

The "language lab" metaphor. The metaphor guiding development of this project was that of a "language lab" for learning Linked Data, and the workshop's goal was to specify how the language lab should be equipped. While the language metaphor was seen as useful because it emphasizes the nature of RDF as a conceptual model rather than just a specific data format or concrete syntax, workshop participants felt that it should not be featured (as in the initial concept papers for the February workshop) in the place of native RDF terminology.

Use of native RDF terminology. The native terminology of RDF makes few distinctions between types of data; for RDF, "everything is just data." Distinctions common in other fields between, say, "element sets," "value vocabularies," and "datasets" have no exact equivalents in RDF. Ontologies, RDF vocabularies, and SKOS concept schemes – the conceptual structures of Linked Data – are themselves considered as "just data" and expressed with the same formatting as the instance metadata using those structures. Workshop participants felt that the Inventory of Learning Topics, and any future Learning Lab, should take care to label topics primarily with terminology native to RDF, drawing analogies to terms from other fields only as needed. Cloaking the principles underlying RDF in the terminology of, say, library science, may help students in that discipline to grasp concepts in the short term but does not prepare them well for working with RDF outside of the library-science context. The participants felt it to be the job of instructors, rather than the Learning Linked Data project itself, to shape the material into a form intellectually accessible to specific groups of students. As one participant noted, "Grounding [the project] in RDF allows instructors to use whatever metaphor or mechanism makes sense to them and their students -- whether linguistics, math, or programming logic."

A "kitchen" metaphor. The collection of software-supported methods to be provided by the project was seen as providing a "palette" of functionality from which instructors could draw in

realizing particular learning goals and pedagogical approaches. The metaphor of a "kitchen" was proposed, describing the challenge in terms of equipping a workspace with utensils that would allow cooks to prepare a wide range of "courses."

An Inventory of Learning Topics

The resulting list of topics was organized under five categories:

- Understanding Linked Data
- Searching and querying datasets
- Creating and manipulating RDF data
- Visualizing webs of data
- Implementing a Linked Data application

For each Learning Topic, the workshop participants characterized what type of software tools instructors and learners would need to use in the course of learning, illustrating each type of software, where possible, with a known exemplar.

Following the workshop, the Inventory was posted on a University of Washington blog^[1] for public comment. Workshop participants helped disseminate this invitation to specific colleagues as well as to mailing lists and communities where they were active (see the Outreach list in the Appendix). The resulting feedback greatly enriched the Inventory with pointers to additional tools and with potential user scenarios. Much of this input was documented as blog comments, and much arrived via email and other private channels. All comments were compiled into a document that will inform the development of a proposal for a Learning Lab.

Target Users of a Learning Lab

The workshop, follow-up discussion, and feedback on the blog identified three broad categories of potential users for tutorials on learning Linked Data:

Working professionals seeking to solve problems on-the-job. Potential users in professional settings may bring diverse sets of needs. For example, working librarians, archivists, and others may need to map metadata records in diverse formats (MARC, EAD, and Dublin Core) into a common record structure. Educators face challenges to integrate diverse sources of data about faculty expertise, scholarly research, grant applications, and publications. Institutions want to link journals, conference proceedings, committee bulletins with topic-relevant information from external sources or provide unified access to functionally separate repositories. Such scenarios require working professionals to learn new approaches to familiar problems while remaining on the job, pointing to the need both for professional training seminars and for resources to support self-instruction.

Students seeking formal qualification. In today's environment, people move into the library and museum profession from educational backgrounds ranging from the social sciences and humanities to computer science, both academic and applied. The availability of generic

instructional materials -- usable in the context of widely differing pedagogical approaches -- would support formal university courses at all levels.

Faculty, trainers, or consultants instructing either of the above. Instructors in courses about, or touching on, Linked Data themselves need to continually update their knowledge and skills. Indeed, the planning project has been animated in large part by faculty members and professional trainers who feel challenged by the rapid pace of technological change. In order to keep ahead of their students, they see a need for a forum where they can learn about new concepts and tools and pool experience among themselves about their use in teaching.

Planning the Learning Lab

The planning project initially envisioned the "language lab" for learning Linked data as a software platform -- a set of analytical and data-processing tools, with documentation on how to use those tools for teaching, possibly packaged for download, and ideally integrated with an "orchestrator" interface. The participants concluded, however, that a project heavy on software development would be risky to undertake, subject to continual obsolescence, and thus unsustainably expensive to maintain.

Rather, the project converged on the concept of a learning environment comprising, as a recent New York Times article on "continual learning" puts it: "bite-size instructional videos, peer-to-peer forums, and virtual college courses" ^[2]. (Indeed, the article highlights the case of a metadata librarian needing to keep up with evolving technology.)

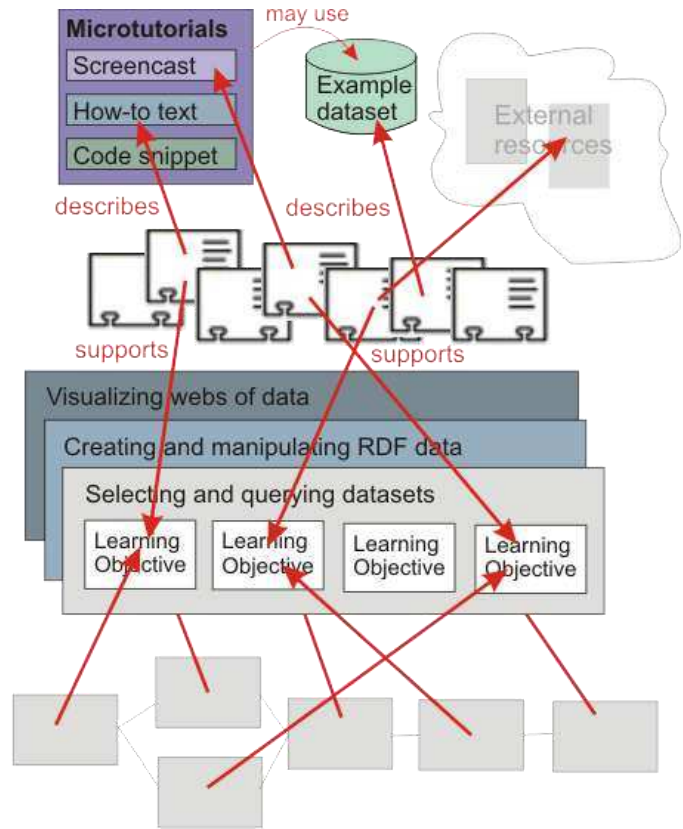
Virtual courses. The project lead, University of Washington, participates in Coursera, a pioneer of "massively open online courses" ^[3]. Online courses and videos offered by MIT, Stanford, Cornell, and Khan Academy ^[4] (with support from the Bill & Melinda Gates Foundation) exemplify this trend. The trend towards virtual courses, however, also involves support for individualizing learning, sometimes through "flipping" the classroom experience by assigning lectures as homework and using class time for targeted assistance to students with what used to be homework exercises.

A "competency-based" approach. Achieving the ideal of individualized instruction involves a new and emerging approach to course design based on the definition of competencies to be achieved as outcomes of instruction (Learning Objectives). The National Science Digital Library (NSDL), for example, defines learning objectives related to the Benchmarks for Scientific Literacy of the American Association for the Advancement of Science (AAAS). These learning objectives are linked to supporting resources -- lesson plans, videos, exercises, quizzes, and the like -- which may be selectively combined in Learning Trajectories (in NSDL terminology "strand maps" ^[5]). Learning Trajectories chart educational pathways customized for different audiences, or even for individual learners.

Using Linked Data to implement the competency-based approach. Another of the IMLS planning project partners, JES & Company, is a pioneer in the use of Linked Data to create

machine-readable webs of Learning Resources, linked with Learning Objectives, for use in creating Learning Trajectories. Created with support from the National Science Foundation and from the Bill & Melinda Gates Foundation, the Achievement Standards Network (ASN)^[6] provides access to RDF representations of learning objectives in the form of specific "learning outcome statements," identified with URIs, as linking

Learning Resources
Metadata Descriptions
Learning Objectives
Learning Trajectories



targets for related learning resources. This approach has been adopted for support of United States Common Core State Standards and by related endeavors such as the Learning Resource Metadata Initiative, a project co-led by the Association of Educational Publishers and Creative Commons^[7], the Machine Readable Australian Curriculum^[8], and the European Education Resource Network, EdReNe^[9].

Proposal for a Learning Lab about Linked Data. The primary deliverable of the IMLS planning project, an Inventory of Learning Topics in the area of Linked Data, is proposed as the starting point for a learning environment structured according to the competency-based approach. The Learning Lab would initially focus on the three topic areas which fall between prerequisite knowledge ("Understanding Linked Data") and skills related to specific software packages ("Implementing a Linked Data application") -- the topic areas about searching, querying, creating, manipulating, and visualizing Linked Data.

From Topics to Learning Objectives. The project will begin by "unpacking" (or "deconstructing") the three general topic areas into more-specific Learning Objectives, using an online cataloging tool provided by the Achievement Standards Network to describe the objectives and assign them URIs. Addressable Learning Objectives, so described, will serve as targets to which supporting tools and resources can be linked by means of Metadata Descriptions using, at least initially, a cataloging tool provided by The Gateway^[10], a repository of education resources for US teachers. Ideally, the definition of more-specific objectives under general topics

will result from discussion by a community of interested instructors and over time, the topic areas covered will evolve both in breadth and in depth.

From Learning Objectives to Microtutorials. Feedback on the University of Washington blog to the Inventory of Learning Topics reinforced the notion that learners seek "playable" tutorial materials which visibly demonstrate the use of tools for analysing or processing Linked Data. In order to establish an Inventory of (addressable) Learning Objectives as a linking target for such resources, the project will use well-specified Learning Objectives as concrete specifications for screencasts and to-do texts ("microtutorials"), to be solicited by the project from software developers in order to show specific uses of their tools. Longer-term, the goal is bootstrap the creation of a growing set of well-specified Learning Objectives to which other tools and resources can maintainably be mapped -- as common reference points to help developers communicate with users in the education world about the capabilities of their tools, and as the glue joining discrete instructional resources to larger pedagogical packages.

The role of a peer-to-peer forum. The IMLS planning grant has begun to create a community of university and professional instructors collectively interested in improving the quality of education about Linked Data. The Dublin Core Metadata Initiative, a non-profit organization dedicated to promoting and educating about the use of open metadata standards, has offered to provide infrastructural support, a permanent home, and long-term-persistent URIs for the activities and products -- Learning Objectives, Learning Resources, and the Metadata Descriptions linking the two -- of a Learning Linked Data community. This community, it is envisioned, can serve as a platform for discussing courses and curricula built on the foundation of the Learning Objective approach, perhaps expanding its scope in the medium term beyond the initial scope of Linked Data to encompass other related topics of metadata best practice.

References

1. <http://lld.ischool.uw.edu/wp/>
2. <http://www.nytimes.com/2012/09/22/business/to-stay-relevant-in-a-career-workers-train-nonstop.html>
3. <http://ischool.uw.edu/news/2012/07/barbara-endicott-popovskys-course-one-first-offered-new-uwcoursera-agreement>
4. <http://www.khanacademy.org/>
5. <http://strandmaps.nsd.org/>
6. <http://asn.jesandco.org/>
7. <http://wiki.creativecommons.org/LRMI>
8. <http://www.australiancurriculum.edu.au/Technical/Introduction>
9. <http://edrene.org/>
10. <http://www.thegateway.org/>

Part 2: Quantitative Report

Institution Name: The Information School, University of Washington

Grant #: JG - 51 - 11 - 0147 - 11

A. SITE SPECIFIC PROJECT ACTIVITY: Workshop on resources for Linked Data instruction, 2-3 Feb. 2012 at UW iSchool, Seattle, Wash.; follow-up workshop to finalize project deliverables, 17-18 Sept. 2012 at UW iSchool, Seattle, Wash.

1. ___ n.a. ___ Total # of collection items conserved, relocated to protective storage, rehoused, or for which other preservation-appropriate physical action was taken.
2. ___ n.a. ___ Total # of collection items digitized, scanned, reformatted, or for which other electronic or digital preservation action was taken.
3. ___ n.a. ___ Total # of collection items with new or enhanced accessibility (include items that were cataloged or for which finding aids or other records were created or computerized) [includes ___ n.a. ___ items made accessible to users other than grantee staff for the first time, ___ n.a. ___ items with new or enhanced access for staff only].
4. ___ 2 ___ Total # of lectures, symposia, demonstrations, exhibits, readings, performances, concerts, broadcasts, Webcasts, workshops, multi-media packages, or other learning opportunities provided for the public (do not include PSAs or other promotional activities) [includes ___ n.a. ___ out-of-school or after-school programs, ___ n.a. ___ exhibits].
5. ___ n.a. ___ Total # of tools created, improved, or produced for searching, information management, or information analysis by users other than or in addition to grantee staff.
6. ___ 2 ___ Total # of conferences, programs, workshops, training sessions, institutes, classes, courses, or other structured educational events provided.
7. ___ 3 ___ Total # of internships, apprenticeships, mentoring opportunities, or other extended educational opportunities provided.
8. ___ n.a. ___ Total # of degrees/certificates earned as a result of the grant [includes ___ n.a. ___ Master's, ___ n.a. ___ Ph.D. degrees, ___ n.a. ___ other (specify): ___ n.a. ___].
9. ___ n.a. ___ Total # technology upgrades or improvements (specify): ___ n.a. ___

10. Other activities not covered by the categories above:

Two workshops, one with 16 participants -- 4 from University of Washington iSchool, 3 from the Dublin Core Metadata Initiative, 2 from Kent State University, 1 each from University of North Carolina - Chapel Hill, University of Illinois iSchool, New York University Libraries, Library of Congress, 1 each from three private consulting practices in information systems

serving libraries, not-for-profit organizations, and corporate enterprise – the second with 5 participants.

B. PORTABLE PRODUCTS (relating to the activity named in section A.)

11. __ n.a.__ Total # of research reports, papers, books, reprints, or other publications generated.

12. ___1___ Total # of Web sites developed or improved [include URLs/addresses:
___http://lld.ischool.uw.edu/wp/___].

13. ___2___ Total # of learning resources produced [includes __ n.a.__ oral histories, __ n.a.__ curriculum resources, __ n.a.__ curriculums, __ n.a.__ Web-based learning tools, or ___2___ other (specify): _Inventory of Learning Linked Data Topics, Glossary_].

14. ___5___ Total # of key management documents created
[includes __ n.a.__ emergency plans, __ n.a.__ conservation surveys, __ n.a.__ strategic plans,
___5___ other (specify): __planning wiki, 3 blog posts, 1 implementation project narrative__].

15. If your grant created one or more quantifiable products not covered by the categories above, please briefly identify and quantify them here. Attach another sheet if necessary.

1 Inventory of Learning Topics (B.13): 6 main categories, each with 8-13 subcategories
1 panel proposed summarizing the project activity for the ALISE 2013 conference
1 panel proposed for currently scheduled session for the ALA Midwinter conference

C. PARTICIPANTS/VISITORS/USERS/AUDIENCE (relating to the activity named in section A.)

16. ___>45___ Total # of **community organization partners** [includes __>30__ informal partners, __15__ formal partners].

17. __ n.a.__ Total # of **schools** (pre-K through grade 12) that used services provided by your grant (include only schools that actively participated, not those to which material was simply distributed or made available) [includes __ n.a.__ students participating in field trips].

18. ___ n.a.__ Total # of **teachers** supported, trained, or otherwise provided with resources to strengthen classroom teaching or learning.

19. ___ n.a. ___ Total # of **pre-K through grade-12 students** served [includes _ n.a._ youth 9-19 who used, participated, visited, or otherwise interacted with activities, experiences, resources, or products offered by your grant].

20. ___ n.a. ___ Total # of **viewers and listeners** for radio, television, and cable broadcasts (for series, include total actual audience for all broadcasts; do not include audience for PSAs or other promotional activities or Webcasts; do not report potential audience).

21. ___ >200 ___ Total # of **users of Web-based resources** provided by your grant (include all individuals the project served). Choose the measure that best represents your use rate (choose only one): ___ visits (hits), ___ unique visitors, ___ registered users, ___ other measure (specify): ___ 53 page comments from 24 individuals, plus >12 comments by email from 8 individuals ___.

22. ___ >300 ___ Total # of **individuals** benefiting from your grant (include all those from questions 18-21 plus others the project served, including staff or others in your field). Only include those who actually participated or used your project services in some way.

23. This number includes: ___ >300 ___ **professionals**, _____ **non-professionals or pre-professionals**, _____ **docents or interpreters**, _____ **volunteers**, _____ **staff** that received services provided by your grant.

24. If your grant served one or more quantifiable audiences not covered by the categories above, please briefly identify and quantify them here. Attach another sheet if necessary.

6 individuals representing 4 organizations provided Linked Data use cases
>20 software tools and instructional resources identified

Appendix

Workshop Attendees

- Mike Crandall, Joseph Tennis, Randy Orwin, and David Talley from the University of Washington iSchool
- Tom Baker, Diane Hillmann, and Stuart Sutton from the Dublin Core Metadata Initiative
- Riley Stormer and Marcia Zeng from the Kent State University School of Library and Information Science
- Craig Willis (representing Jane Greenberg) from the School of Information and Library Science at the University of North Carolina at Chapel Hill
- Karen Wickett from the University of Illinois iSchool
- Independent consultants Joseph Busch, Karen Coyle, and Marjorie Hlava
- Library software developers Corey Harper and Ed Summers