

# Learning Linked Data: Project Workshop, 2-3 February 2012 -- Minutes

---

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

## Document Scope

This document represents a condensed version of a full transcript (40+ pages in length) of discussions during the 2-day workshop.

## Workshop Attendees

Tom Baker, Joseph Tennis, Marjorie Hlava, Riley Stormer, Marcia Zeng, David Talley, Joseph Busch, Karen Wickett, Corey Harper, Ed Summers, Craig Willis, Diane Hillmann, Karen Coyle, Randy Orwin, Stuart Sutton, Mike Crandall

## Goal of this workshop

The goal of this day-and-a-half workshop was to agree on an analytical inventory of Learning Topics related to Linked Data, and to use these Learning Topics as the basis for outlining a software tool platform to be created (in a 2013+ IMLS project) to support the teaching and learning of the Learning Topics.

The guiding metaphor was that of a "language lab" for learning Linked Data and, in this respect, the goal was to specify how the language lab should be equipped.

The backgrounds of attendees included: academia, IT development and deployment, software, LIS, metadata design and application.

A first-draft inventory of Learning Topics was discussed point-by-point and refined into a second-draft inventory of Learning Topics associated with supporting Software Tools.

## Scope of workshop discussion

The workshop began with a discussion of scope and objectives:

- The Learning Topics cover the basics of Linked Data -- topics that provide learners with a springboard both for learning more advanced topics and for applying the skills in practice.
- The participants agreed to prioritize the analysis of existing Linked Data on the Web over the more ambitious goal of designing data models from scratch and creating Linked Data applications. In particular, the group agreed that tool support for the design workflow implied by the Singapore Framework for Dublin Core application profiles was too specific to one methodology; some students will be better-prepared for UML or database approaches. Moreover, design of mixed-vocabulary profiles can quickly become quite complex. Given the state of software tools, it would be more realistic to defer the requirements of application profile design to future projects. "The Singapore Framework is a layer on top of RDF that helps a particular group understand how it applies to them".
- "Learning outcomes", detailing specific skills or knowledge to be mastered by learners in specific courses, were considered out of scope. An "outcome" is typically formulated with a sentence that starts "At the completion of this course, students will be able to..."

- The intended audience of the Learning Topics and software tool platform were acknowledged to be very diverse. Although participants all came to the workshop with particular pedagogical scenarios in mind -- courses for particular learners, from particular backgrounds and experience (e.g., "baseline" versus "competent"), with particular expected outcomes -- specific pedagogical approaches, curricular concepts, and target audiences were considered beyond the scope of this discussion. (See the ACTION below on all participants to articulate these use cases as follow-up to the workshop.)
- The discussion was agnostic about how such a software tool platform might be realized in a future project -- whether through packaging a set of tools ("snapshot"), more tightly integrating that set of tools in a common interface with "orchestrator" functionality, or more simply by focusing on documenting the potential uses of the tools in learning environments -- filling the gap between how the software works and how it may be used for learning Linked Data.
- The point is not to promote a particular set of tools or vocabularies, but to put a well-documented set of tools into the hands of instructors in this area, allowing them more efficiently to focus on their teaching.

## Discussion of learning topics

The results of this discussion have been incorporated into the revised Inventory of Learning Topics, not repeated here. A few highlights:

- While the participants acknowledged that the "language" metaphor is particularly apt for RDF, and usefully underlines the nature of RDF as a model in contrast to specific data formats or concrete syntaxes, consistently using the language metaphor as upper-level headings for the Learning Topics (e.g., "grammar", "composition", "translation"... ) risks imposing an interpretation not generally used in writing about RDF. It was agreed, therefore, that topic headings should use the native jargon of RDF.
- In the native jargon of RDF, "everything is just data". In other words, distinctions made in other fields between, say, "element sets", "value vocabularies", and "datasets" are not native to RDF. Indeed, ontologies, RDF vocabularies, and SKOS concept schemes are themselves considered data. The Inventory of Learning Topics -- and the narrative in which the future tool platform will be wrapped -- should therefore take care to label topics primarily with native-RDF terminology and point to corresponding terminology from other fields by way of explanation. Cloaking RDF in the terminology of, say, library science, may help students grasp concepts in the short term but will not prepare them well for working with RDF in other contexts. The participants felt that it is the job of instructors (not of the platform itself) to shape the material into a form that their students will grasp.
- "Until we get to a neutral point of language, the communities will talk past each other."
- "When you go out there looking for tools for Linked Data, they're not going to be described in linguistic terms. One introduces dissonance by talking about it in a way that doesn't match the tools."
- "Grounding it in RDF allows instructors to use whatever metaphor or mechanism makes sense to them and their students, whether linguistics, math, or programming logic."

- Participants agreed that Visualization constituted a Learning Topic, not just a type of tool. Visual node-and-arc diagrams, for example, are one of the modes in which RDF graphs can be expressed.
- On the formulation "publishing Linked Data 'as Linked Data'": "If you just took your data, put it into a ZIP file and put it out there, it's not Linked Data in the sense of being linked to anything. We're really talking about the idea of exposing URIs on the Web." In principle, however, Linked Data can use non-dereferenceable URIs, such as info URIs.
- "Storing", as a learning topic, should cover different types of storage that people are using. It is important to teach that data does not need to be stored in a triple store to be Linked Data. However, it makes sense to include triple stores in the tool platform because storing data in triple stores is relatively easy and straightforward. We need to distinguish between having a triple store in the tool platform versus triple stores as a Learning Topic.
- On the topic "Implementation": "Here are the actual things you might need to know in order to implement this in practice." We shouldn't lead people to the edge then leave them without the ability to actually put it into practice.

## Use cases

- "Collecting use cases could be an important component of the platform, allowing people to match use cases to their own situation".
- Corey: Merge MARC, EAD, and Dublin Core records into "NYU Core" records. Taking five records, extracting parts of them to triples, merging those into a graph, and turning that into a record.
- Joseph B: "Companies want consultants to produce deliverables. What you're doing here is creating an environment for people to teach themselves. A corporate problem would be that they can't efficiently get the data they need because it's dispersed in all these different places. I educate them, get their data management people to stop doing 'data management,' and think about how their data should be made Linked Data ready without building another system. A lot of this is how to educate people in the real world who are IT professionals, library professionals, to think differently about problems they're already familiar with. This tool is focused on one part of that educational process."
- Marjorie: "My need is a little different; I have to deliver the actual product. I might start with specifications, but we have to take data, massage it, and put it into a system where it will work. A science association might have journals, conferences, committees, etc. on a topic. If someone has an article they discovered, they might want to link out to other things in different formats or venues that have to do with that data. The use case is being able to present on a website to the searcher who finds that article the upcoming conference, etc. It's a great use case for Linked Data capability."
- Marjorie: "In records management, I have a client that's a large insurance company, with 2 terabytes of undifferentiated data lying around like a bomb. It costs them over a million dollars a year to maintain the drive. The application there is to figure out the retention schedule for the data. We have a controlled vocabulary and a retention schedule, and you link the two and apply it in a crawling mechanism to assign retention schedules to every

information object on the shared server. They have legal holds on some data that would otherwise fall under some of those schedules."

## **Associating software tools with learning topics**

The workshop concluded by determining, for each Learning Topic, what type of software would be needed by instructors and learners. Each type of software was illustrated, where possible, with an exemplar.

The resulting inventory of Learning Topics and supporting Software Tools provides the rough outline of a software tool platform needed by a "language lab" for learning Linked Data. The software tool platform is seen as providing a palette of functionality on which instructors can draw in realizing a particular course concept. The metaphor of a "kitchen" was proposed, the challenge being that of equipping a workspace with utensils that will allow cooks to prepare a wide range of "courses".

The original Inventory of Learning Topics tried to classify topics into "beginning", "intermediate", and "advanced" levels. The discussion ignored these as distinctions among topics in acknowledgement of the diversity of backgrounds from which learners will approach the topics. In the discussion of tools, however, the participants recognized a rough distinction between tools usable by anyone with a mouse and tools that required command-line input or writing lines of code. It was therefore suggested that the software platform might need to distinguish tools by required IT competence.

## **ACTIONs taken**

ACTION 2012-02-03: Marjorie, Corey, and possibly Ed to write a paragraph or two on "non-RDF Linked Data".

ACTION 2012-02-03: Everyone, starting with Stuart, to write up a brief use case describing a pedagogical approach to teaching (or learning) Linked Data, specifying how the elements of their approach map to the Learning Topics. "If we do our homework on the use cases, we'll be able to fill out details about the intended audience".

ACTION 2012-02-03: Tom to revise the Inventory of Learning Topics, providing definitions of basic concepts where necessary.

## **Putting the Inventory of Learning Topics out for comment**

- First, review within this group. As a next step, maybe selected reviewers, then go to a blast:
- ALISE - ALA's Association for Library and Information Science Education
- DCMI
- DERI
- Talis
- UKOLN

- Individual iSchools
- ASIS Information Architecture list is active (with alot of professors)
- LODLAM list, because people are trying to learn this stuff outside of an educational environment.
- Make a list of people we know are teaching in this area.
- W3C Library Linked Data list
- ISKO, especially the European chapters
- We will want to look for commitments (or interest) from people to use the platform we envision.