

# LIS 640: Linked Data

Information School  
University of Wisconsin-Madison  
Spring 2018

Dorothea Salo (please call me “Dorothea”)  
Office address: 4261 Helen C. White Hall  
Course URL: <https://canvas.wisc.edu/courses/82244>  
Instructional mode: all online

salo@wisc.edu  
Office Hours: by appointment  
Special course attributes: none

## Course Description and Objectives

Students completing this course will earn one credit hour. One credit is the learning that takes place in at least 45 hours of learning activities, which include time in lectures or class meetings, in person or online, labs, exams, presentations, tutorials, reading, writing, studying, preparation for any of these activities, and any other learning activities.

This course has a co-requisite of LIS 602.

Upon completion of this course, students will be able to:

- Insert schema.org microdata correctly into an HTML document or template
- Read a relatively simple RDF graph
- Read RDF triples in N-triple, Turtle, and RDF/XML formats; hand-author triples in N-triple and Turtle formats
- Recognize and use syntax for common RDF datatypes and notations (e.g. URIs, strings, dates, language)
- Recognize and read a few RDF vocabularies common in information agencies (e.g. DC, SKOS, PCDM)
- (Optional extra unit) Query RDF datastores with SPARQL

This course is designed to assess student progress in the following iSchool program-level learning outcomes: 3a and 3d.

## Course Policies

**I intend to fully include persons with disabilities in this course. Please let me know within one week how I can best meet your needs. I will try to maintain the confidentiality of this information.**

**Academic Honesty:** I follow the academic standards for cheating and plagiarism set forth by the University of Wisconsin.

### Readings and software

There is no textbook for this course; all readings are listed in this syllabus. Check Canvas for required software installations.

### Weekly schedule

Course weeks run from Tuesday to Monday this semester. Due dates and new-content releases are calibrated accordingly. *Course content opens a week early*, to help those who are able to work ahead.

### Contacting me

For any difficulty with the course that is not private or confidential, please ask in class or use the Canvas help forum; *I will not answer such questions by email*. Please also do your best to assist your classmates. I am not available weekends; otherwise, I do my level best to answer forum questions and email within two business days.

I have no travel planned during this course, but please be aware that I am teaching two other regular courses and one continuing-education course at the same time. Help me juggle when you can!

### Prerequisite skills

LIS 640 will require you to go beyond the basic computer skills needed in other courses. I assume that you have successfully mastered the following computer skills, ergo *will not teach them*:

- Hand-author basic HTML (what you learned in LIS 500, 602, or 644 should be sufficient)

- Be able to save files to and retrieve them from external storage (USB drives, cloud storage, etc.) Know the difference between `http://` and `file://` URLs, and what that means for whether someone other than you can access a given file.
- Be able to create, find, and use directories/folders.
- Be able to use a Web browser, perform Internet searches, and save a file to your computer from a hyperlink.
- Be able to save files to a specific location and e-mail files to yourself.
- Understand the difference between a plain-text editor and a word processor.
- Understand filename extensions and how they affect use of a file.

These are the *minimum* skills required to start the course. If any of the above confuses you, you are expected to work it out on your own *immediately*.

## Week 1: Context. RDF use in information agencies.

*Learning objectives: Why integrating documents and data from different sources is hard, but often useful. History of the Semantic Web; transition to and context of the linked-data movement. Five stars of linked data.*

Hilton. "Rise of the machines." <http://blog.wellcomelibrary.org/2013/12/rise-of-the-machines/>

Campbell & MacNeill. "The Semantic Web, Linked and Open Data." [http://publications.cetis.org.uk/wp-content/uploads/2011/02/The\\_Semantic\\_Web.pdf](http://publications.cetis.org.uk/wp-content/uploads/2011/02/The_Semantic_Web.pdf)

Kelley. "How the W3C has come to love library linked data." <http://lj.libraryjournal.com/2011/08/technology/how-the-w3c-has-come-to-love-library-linked-data/>

"Library Linked Data Incubator Group Final Report." Sections 2, 3, 4.1, 4.4. <http://www.w3.org/2005/Incubator/11d/XGR-11d-20111025/>

Saló. "Linked data in the creases." <http://lj.libraryjournal.com/2013/12/opinion/peer-to-peer-review/linked-data-in-the-creases-peer-to-peer-review/>

"5 star Open Data." <http://5stardata.info/>

If you need a rapid introduction to HTML, get it THIS WEEK: "HTML Beginner Tutorial" <http://htmldog.com/guides/html/> (you won't need Tables or Forms, but do pick up "Span and Div" from the Intermediate tutorial) or "HTML Essential Training" <https://www.lynda.com/HTML-tutorials/HTML-Essential-Training/170427-2.html?org=wisc.edu> From there, use the "HTML element reference" <https://developer.mozilla.org/en-US/docs/Web/HTML/Element> to look up useful tags.

## Week 2: Identifiers and RDF. Microdata in HTML.

*Learning objectives: Identifiers, unique identifiers. Why strings are lousy identifiers, and what that means for library/archive authority control. URIs, URLs, URNs. Identifier/URI correlation; owl:sameAs and similar constructs. Identifier sources. How search engines use schema.org microdata. Microdata syntaxes; using them in HTML. Using microdata in templated web applications (CMSes, ILSes, digital libraries and archives, repositories). (RDF) Classes.*

*Linklist: <http://pinboard.in/u:dsalo/t:identifiers>, <http://pinboard.in/u:dsalo/t:schemaorg>*

"Falsehoods programmers believe about names." <http://www.kalzumeus.com/2010/06/17/falsehoods-programmers-believe-about-names/> (pay special attention to 7, 12-13, 21-22, 37-40!)

Dempsey. "Names and identities: looking at Flann O'Brien." <http://orweblog.oclc.org/archives/002212.html>

Look me (and/or your favorite author) up in <http://viaf.org/> and the name authority search at <http://id.loc.gov>. Check out the available RDF!

Custer & Joyner. "Approaching authority." <http://www.slideshare.net/steganogram/approaching-authority-a-preliminary-implementation-of-encoded-archival-context-eaccpf-at-east-carolina-university> (Skim this for what they're trying to accomplish and how they did it.)

Ronallo. "HTML5 microdata and schema.org." *code4lib journal*. <http://journal.code4lib.org/articles/6400> (You may stop when you reach the tutorial, but keep this article in mind for examples that will help you do your homework.)

Coyle. "Schema.org: where it works." <http://kcoyle.blogspot.com/2014/10/schemaorg-where-it-works.html>

Hellman. "Spoonfeeding library data to search engines." <http://go-to-hellman.blogspot.com/2011/07/spoonfeeding-library-data-to-search.html>

Arnold. "Improving search engine traffic to DIMES." <http://rockarch.org/programs/digital/bitsandbytes/?p=826>

## Week 3: RDF graphs. N-triples. Simple RDF vocabularies.

*Learning objectives: RDF as data model with many serialization syntaxes. Microdata as RDF. Subject, predicate, object. Reading RDF graphs. String-literal objects; marking language on literals. N-triples syntax. Basic RDF via the Yarn Game. FOAF. Portland Common Data Model (PCDM).*

"Introducing Graph Data." <http://www.linkeddatatools.com/introducing-rdf>

Gonzalez. "RDF 101." <http://web.archive.org/web/20140327205008/http://www.cambridgesemantics.com/semantic-university/rdf-101> (Don't try to use the version on the Cambridge Semantics website; all the images have dropped out of it, and it's useless without them.)

Gonzalez. "RDF Nuts & Bolts." <http://web.archive.org/web/20131030053957/http://www.cambridgesemantics.com/semantic-university/rdf-nuts-and-bolts>

FOAF. "FOAF vocabulary specification." <http://xmlns.com/foaf/spec/#sec-standards> (through "FOAF Auto-Discovery")

Giarlo. "Hydra-in-a-Box and PCDM." [http://hydrainabox.samvera.org/2016/01/28/hydra-in-a-box\\_and\\_pcdm.html](http://hydrainabox.samvera.org/2016/01/28/hydra-in-a-box_and_pcdm.html)

## Week 4: Turtle. RDF/XML. Inference engines; classes; domain/range. Common RDF vocabularies. Information-agency-specific RDF vocabularies.

*Learning objectives: Turtle abbreviations. Reading RDF/XML. More on RDF classes. Domain and range in RDF. Inference. SKOS. Dublin Core. FOAF. BIBFRAME. BL model.*

Coyle. "Classes in RDF." <http://kcoyle.blogspot.com/2014/11/classes-in-rdf.html>

"Introducing RDF/XML." <http://www.linkeddatatools.com/introducing-rdf-part-2>

Gutteridge. "What you need to know about RDF+XML." <http://blog.soton.ac.uk/webteam/2010/11/08/what-you-need-to-know-about-rdfxml/>

W3C. "SKOS primer." <http://www.w3.org/TR/skos-primer/> (through section 3; ignore sections 4 and 5)

## Week 5: RDFizing other metadata.

*Learning objectives: Atomicity, and why a lot of library/archive metadata doesn't have it. Transforming metadata in other formats to linked data. Limitations of such transformations (especially with respect to string values). Reconciliation as a step toward linked data.*

Coyle. "Linked data first steps & catch-21." <http://kcoyle.blogspot.com/2013/07/linked-data-first-steps-catch-21.html>

Stevenson. "Archives Hub and VIAF Name Matching." <http://archiveshub.ac.uk/blog/2013/08/hub-viaf-namematching/>

Schilling. "Transforming library metadata into linked library data." <http://www.ala.org/alcts/resources/org/cat/research/linked-data>

Weeks. "OMG! My metadata is as fresh as the Backstreet Boys!" <http://www.slideshare.net/rascalwhale/using-google-refine-long-version>

Heller. "A librarian's guide to OpenRefine." <http://acrl.ala.org/techconnect/?p=3276>

Nguyen. "Using Google Refine to clean messy data." <http://www.propublica.org/nerds/item/using-google-refine-for-data-cleaning>

## OPTIONAL EXTRA UNIT: SPARQL.

*(For those who have taken or are taking LIS 751)*

Prud'hommeaux. "SPARQL vs. SQL: Intro." <https://web.archive.org/web/20170227201138/http://www.cambridgesemantics.com/semantic-university/sparql-vs-sql-intro>

Lincoln. "SPARQL for humanists." <http://matthewlincoln.net/2014/07/10/sparql-for-humanists.html>

"Querying semantic data." <http://www.linkeddatatools.com/querying-semantic-data>

## Assignments

**NO ASSIGNMENTS MAY BE SUBMITTED AS WORD-PROCESSING DOCUMENTS OR PDFS** unless this is specifically permitted in the assignment description. Assignments submitted in this fashion will *automatically receive zeroes*. HTML and all RDF serializations are *plain-text* formats; install and use a text editor (sometimes called a "programmer's editor") to work with them.

I expect and encourage collaboration among students in this course on major projects as well as weekly assignments. Students who work and study with partners generally find the assignments easier. However, all homework assignments will be submitted, reviewed, and graded individually.

### Assignments

Valid HTML5 résumé

### Percentage

10%

### Due date

start of Week 2

HTML résumé with added microdata

20%

start of Week 4

RDF graph and RDF from résumé

20%

Final day of course

Weekly assignments

40%

(throughout course; 8% each week)

Timely peer review of weekly assignments 10% (weeks 3 and 4; 5% each week)

Grading scale: 100-93.5 A; 93.4-89.5 AB; 89.4-83.5 B; 83.4-79.5 BC; 79.4-73.5 C, 69.5-73.4 D, below 69.5 F

## Weekly assignments

Weekly assignments are part of each week's Canvas content. If no other due dates are given, the following schedule applies:

- Initial assignment draft: Friday
- Peer review of draft assignment, where applicable: the next Monday
- Revised assignment: Tuesday of the next week

Late assignments and late peer reviews will be penalized one point per day or fraction thereof late. (You *don't want to get behind* in this class. You really, really don't.)

Many weekly assignments will be of the form “Do—Check—Fix.” Once you DO the work, turn in your first (draft) version to the weekly assignment box. Next, CHECK your work, sometimes via a test to pass (as with XML validation or linked-data crosswalking), sometimes via peer review. For peer-review CHECKs, submit your peer review by the due date. FIX any errors you find, then turn in the corrected version to the weekly dropbox. I will only look at the draft assignment if I suspect academic-honesty issues.

Canvas's peer review feature is not perfectly intuitive. Learn how it works from <http://freshmancomp.com/2013/02/19/conducting-peer-review-assignments-in-canvas/> or <https://community.canvaslms.com/videos/1133> (transcript available).

## HTML5 résumé

Reformat your résumé into HTML5. (You may add CSS to make your résumé look more attractive, but this is *not required*; your HTML5 résumé need not visually resemble your printed one.) Minimally, the HTML markup must use heading tags and (as appropriate) list and paragraph tags, and it must use them consistently; any other tags used must be sensible in context. Do not use HTML table tags just to align material visually in the browser, please. You will lose half credit on this assignment if your HTML does not validate via <https://validator.w3.org> (warnings usually okay; fix all errors).

## HTML5 résumé with microdata

Next, add microdata to your résumé indicating that it is about you.

Now, for each person (self, reference, supervisor, etc.) and organization (school/college, workplace, professional organization, etc.) in your résumé, add Person or Organization microdata respectively from the schema.org schemas. Where possible (there will not always be an applicable property), tie yourself to each Person or Organization via `itemprop` with an appropriate schema.org property. All available person/organization information in your résumé that can be marked up with microdata should be, but you are not required to add information just to mark it up with microdata. Use *subclasses* of Person and Organization and their properties wherever appropriate; at minimum, educational organizations should be given an appropriate subclass of Organization (there are several!). You may add additional HTML tags to create scope, if you need to.

Word to the wise: if you've made CreativeWorks of some kind, you can mark them up fairly easily (go ahead!), but trying to tie them back to you as creator is cumbersome. Don't bother.

I have made you an example derived from my own CV at <https://github.com/dsalo/MicrodataResume>. (To use the Structured Data Sniffer on it, click on GitHub's “Raw” tab for the file DSaloResume.html, save the result to disk, reopen it in your browser, and click the dog head.)

## RDF graph and RDF from résumé

Pull out the people (references, supervisors, instructors, etc; minimum of 3, add people if you need to), organizations (educational institutions, workplaces; minimum of 2, add to yours if you need to), and work products/projects (books, articles, presentations, e-portfolio, projects, etc; minimum 1, fake it if you need to) mentioned in your résumé OR involving people mentioned in your résumé. Using at minimum three different linked-data vocabularies (that is, you should need at least three `@prefix` declarations if using Turtle) as sources of subject/property/value URIs, create a single RDF graph (that is, containing no triples completely isolated from the larger graph) containing a minimum of 30 triples. Look up and use appropriate URIs for people and organizations wherever possible. You may use whichever RDF serialization you prefer. Draw and turn in a graph of the resulting triples, either by hand or with the W3C's graph/validator tool (<http://www.w3.org/RDF/validator/>), RDF Distiller (<http://rdf.greggkelllogg.net/distiller>) or similar RDF-graph-generation tool.

<b>SLIS Program-level Learning Outcomes</b>	<b>640 Objectives</b>	<b>640 Measurable Outcomes</b>
3a. Students organize and describe print and digital information resources.	Recognize and read a few RDF languages common in libraries and archives (e.g. RDFS, DC, SKOS)	Weekly assignments test student ability to recognize and use information in these description languages.
3d. Students understand and use appropriate information technologies.	All objectives.	Weekly assignments designed to familiarize students with linked data. Graded on syntactic correctness, understanding of RDF serializations. Projects measure student ability to be generative (rather than solely reactive) with the tools and techniques learned in class.