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:

- Recognize microdata in HTML
- Add information to Wikidata
- 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: 2, 5. It assesses student progress in the following former 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 will be out of town March 12-13 at a Digital POWRR Institute. I will check email and Canvas, but may fall behind.

Week 1: Context. RDF use in information agencies. Identifiers. Wikidata.

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. 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; Wikidata.

Cagle. "The Semantic Web Comes of Age." https://www.forbes.com/sites/cognitiveworld/2018/08/03/the-importance-of-schema-org/

Salo. "Linked data in the creases." https://web.archive.org/web/20170318073902/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/

"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!)

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.)

Allison-Cassin and Scott. "Wikidata: a platform for your library's linked open data." http://journal.code4lib.org/articles/13424

Week 2: Reading RDF graphs. Microdata.

Learning objectives: RDF graphs. Triple structure: subject, property, value. Where URIs and literals fit. Blank nodes. How search engines use schema.org microdata. FOAF.

Linklist: http://pinboard.in/u:dsalo/t:schemaorg

"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.)

Ronallo. "HTML5 microdata and schema.org." *code4lib journal*. http://journal.code4lib.org/articles/6400 (Stop when you reach the tutorial.)

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

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

Week 3: RDF syntaxes: microdata, N-triples, Turtle, RDF/XML. Simple RDF vocabularies.

Learning objectives: RDF as data model with many serialization syntaxes. N-triples syntax. Turtle. RDF/XML.

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

"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/

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

Learning objectives: More on RDF classes. Domain and range in RDF. Inference. SKOS. Dublin Core. BIBFRAME. PCDM.

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

DCMI Metadata Terms. http://dublincore.org/documents/dcmi-terms/ (For reference. If you like, you can read them in Turtle: http://dublincore.org/2012/06/14/dcterms.ttl)

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

"Overview of the BIBFRAME 2.0 Model." https://www.loc.gov/bibframe/docs/bibframe2-model.html

Giarlo. "Hydra-in-a-Box and PCDM." http://hydrainabox.samvera.org/2016/01/28/hydra-in-a-box_and_pcdm.html

Week 5: RDFizing non-RDF metadata. Reconciliation.

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. Enhancement via linked data.

Linklist: http://pinboard.in/u:dsalo/t:reconciliation (several real-world projects there!)

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

Smith-Yoshimura. "Metadata reconciliation." http://hangingtogether.org/?p=5710

Stevenson, "Archives Hub and VIAF Name Matching." http://archiveshub.ac.uk/blog/2013/08/hub-viaf-namematching/
Heller. "A librarian's guide to OpenRefine." http://acrl.ala.org/techconnect/?p=3276

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	Percentage	Due date
Wikidata assignment	10%	start of Week 2
RDF graph and RDF from résumé	20%	start of Week 4
Clean and reconciled metadata	20%	Final day of course
Weekly assignments	50%	(throughout course; 10% 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, they are due the Monday after they were assigned at 11:59 pm. Late assignments 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). 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.

Wikidata assignment

Based on a prompt in Canvas, you will:

- Create one new Wikidata page, adding at least three statements to it.
- Add new information to at least three more Wikidata pages.

Show your work by turning in screenshots (most image formats or PDF fine; please no Photoshop-native files):

- a screenshot of a Wikidata search showing that the page you intend to create did not exist
- another screenshot of the page once you finish creating it
- "before" and "after" screenshots of the pages you alter RDF graph and RDF from résumé

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 least 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/`), Easy RDF (`http://www.easyrdf.org/converter`) or similar RDF-graph-generation tool.

The graph may be submitted as an image file or PDF. The triples must be plain text.

Clean, reconciled, enhanced metadata

On Canvas you will find a spreadsheet containing metadata that needs some cleanup. Using OpenRefine, clean it up as needed, and reconcile the columns designated in Canvas with Wikidata. Once reconciliation is complete, enhance the metadata by adding at least two more columns based on Wikidata information. Turn in screenshots of the completed OpenRefine project.

iSchool Old Program-level Learning Outcomes	640 Objectives	640 Measurable Outcomes
3a. Students organize and describe print and digital information resources.	Recognize, read, and use 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.

iSchool New Program-level Learning Outcomes	640 Objectives	640 Measurable Outcomes
2. Students apply principles of information organization.	Recognize, read, and use 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.
5. Students demonstrate competency with information technologies important to the information professions.	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.