

LIS 632

Metadata Standards and XML

Information School, University of Wisconsin-Madison
Fall 2022

Dorothea Salo (please call me “Dorothea”)
Office address: 4261 Helen C. White Hall
Course URL: <https://canvas.wisc.edu/courses/305578>
Special course attributes: none

salo@wisc.edu, 4261 Helen C. White Hall
Student Hours: Zoom, 1:30-3:30 pm US Central time Mondays
Instructional mode: online asynchronous

Course description

This course requires graduate standing and has LIS 602 as a prerequisite/co-requisite. No prior technology or computer-science experience is assumed.

Upon completion of the course, students will:

- Understand standards, concepts and types of metadata
- Recognize various metadata schemes, controlled vocabularies, and element sets and evaluate their appropriateness in specific domains
- Apply Dublin Core and MODS metadata to digital objects, using basic XML and RDF encodings
- Crosswalk metadata between two metadata schemes, and document that crosswalk for others
- Understand the structure of and be able to add information to Wikidata
- Select and document a metadata scheme for use with a particular collection or digital library, based on an understanding of the information needs and behaviors of individuals and user communities
- Select and implement appropriate controlled vocabularies; assess and respond to representation and inclusion challenges in controlled vocabularies
- Be prepared to manage digital-collections metadata, including project leadership, needed skills, project management approaches, quality control, sustainability, labor, and collaboration

This course is designed to assess the following iSchool program-level learning outcomes: 2, 5.

Course Policies

I aim to make this course as accessible as possible to all students. Students seeking accommodations for lecture or assignments must obtain a McBurney Center Faculty Notification Letter. For more information, see <https://mcburney.wisc.edu/apply-for-accommodations/>.

Preferred name/pronouns: Your name or gender may have been reported to me incorrectly. Please let me know your pronouns and preferred given name or nickname as you are comfortable. My pronouns are she/her/hers. UW-Madison lets students indicate a preferred name: https://registrar.wisc.edu/preferred_name.htm Canvas does as well, adding pronoun specification: <https://kb.wisc.edu/luwmad/page.php?id=108069> Zoom also has a pronouns setting: <https://kb.wisc.edu/105475>

Contacting me

READ THE SYLLABUS before asking a question, please; the syllabus may answer it! For any difficulty with the course that is not private or confidential, please use the Canvas help forum; *I will not answer such questions by email*. Please also do your best to assist your classmates on the help forum.

Student hours (“office hours”) are listed, but appointments outside them are also fine. Please make such an appointment directly on my O365 calendar at a time that works for you. I am typically around between 8AM and 4PM Central time M-F, but if you need to meet with me outside those hours, it’s fine. Let me know your preferred communication mode: phone, text chat, and online video conference are all fine.

Should you see dead links (it does happen, usually with no notice), weird due dates, or other syllabus problems, please post them to the “Syllabus questions and problems” forum on Canvas as soon as you see them.

Course schedule

Each calendar week of the course will begin no later than Wednesday morning at midnight, though I will try to open modules a week (or perhaps even more) ahead. Assignment due dates will be Tuesday night at 11:59pm (unless otherwise noted). Please ask in the Canvas syllabus-problems forum if a due date appears misplaced; when this happens, the due date will be either the Canvas or syllabus due date, *whichever is later*.

Summer: to accommodate the compressed summer schedule, we will do two modules per week EXCEPT the week of July 4 and the final week of the course, which will contain one module each.

GETTING BEHIND IN THIS COURSE OR ITS ASSIGNMENTS IS A VERY BAD IDEA, ESPECIALLY IN SUMMER TERM.

Topics build on one another, and there is technical content for you to assimilate from the very beginning. If you do not believe you can stay current with course demands (**summer:** taking the compressed schedule into account), this may not be the right course for you.

Notes from a pandemic

These are not usual times, I'm acutely aware. I am absolutely willing to accommodate sudden unforeseen challenges. Please let me know what you need as soon as you can. If I come down with a symptomatic COVID case (and I assure you, I am up-to-date on vaccinations and trying hard not to), this course should not be hugely affected; I've taught it several times before, such that almost all course content is already built. Grading and discussion-forum responses from me will likely be delayed, but aside from that, we should be fine.

Textbooks

We formerly relied heavily on Steven J. Miller's *Metadata for Digital Collections*, which came out in a new edition summer 2022. It is a very useful book (prior students in this class agree) and I recommend you purchase a copy if metadata forms an important part of your career aspirations. That said, the book is expensive and I haven't been able to do a thorough read of the new edition yet, so (except for one chapter from the first edition on e-reserve) I have removed it from the course. Instead, we will use Richard Gartner's *Metadata: Shaping Knowledge from Antiquity to the Semantic Web*, available as a library ebook from <https://search.library.wisc.edu/catalog/9912245153802121> as well as other readings available online.

ASSIGNMENTS

Exact due dates are in Canvas (to avoid discrepancies between Canvas and this syllabus). If you are comfortable working ahead, feel free.

Assignments at-a-glance

	% of final grade	Due date:
Per-module technology work (See each module's description for a brief description of the tech work for that module.)	56% (4% per module)	First day of next module
Schema in 10 Points	14%	End of Module 8
Final project		
Collection build	2%	End of Module 2
Case studies	3%	End of Module 5
Three complete metadata records	5%	End of Module 9
More practice records	5%	End of Module 10
Crosswalk documentation	5%	End of Module 11
Documentation compare/contrast	5%	End of Module 12
Final assignment	5%	End of course

No extra credit opportunities are available in this class.

Final grade 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

Assignment descriptions

Unless otherwise stated here or in Canvas, turn in assignments to the appropriate homework dropbox on Canvas. Assignments employing a plain-text format such as XML or N-Triples/Turtle must be uploaded as plain-text files. If you paste them into a word-processing document or spreadsheet, *you have destroyed them* and **you will receive a zero** for the assignment.

Once we begin working with XML, I expect that you will **parse and** (as appropriate) **validate all XML files** without my having to ask! "XML" files that don't parse will lose half credit automatically. If you run into XML validation problems you

can't solve (this is common and expected!) you will only lose points if you DO NOT include the validation errors in the memo field and do your best to explain why you're stuck.

Per-module technology work

Modules will contain hands-on exercises that apply knowledge and skills from readings and lecture. Topics are listed in this syllabus under the module to which they pertain, are worth four points per module, and are due on the first day of the next module.

Schema in 10 Points

In the XML module, you will read “XML in 10 points”, a document that synthesizes the most important things you need to know about XML in ten (actually seven) points. For this assignment, you will research a content standard, structure standard, or controlled vocabulary *not otherwise discussed in this class* (don't worry, there's no shortage of these!) and prepare a similar training document, presented as a Canvas discussion post. (Note: a placeholder post consisting of your document as a Word or PDF attachment *will receive an automatic zero*. You may, however, attach appropriate supplemental files, e.g. sample records.) You are encouraged to include (with credit/citation as appropriate) images, links and other material that you think your audience will find helpful.

Your intended audience is your classmates: early-career professionals who need to know the essentials of a metadata schema so that they can a) interpret records in that schema that they encounter in their work b) decide whether that schema is an appropriate choice for a metadata creation project and c) make sensible decisions about how much future training to pursue related to that schema.

Schemas and vocabularies you might choose from (many others are emphatically possible!):

- TEI (text encoding initiative; focus on the TEI Header)
- Learning Object Metadata (LOM)
- DarwinCore
- NDLTD thesis/dissertation metadata
- METS
- Mukurtu (this is not technically a schema, but can be used for this assignment; focus on Mukurtu's specialized metadata use)
- EAC (Encoded Archival Context) and/or EAC-CPF
- EAD (only if you are on the archives track and have not yet encountered it)
- ISAD(G) (archives track!)
- DDI (Data Documentation Initiative)
- Ecological Metadata Language
- FGDC Content Standard for Digital Geospatial Metadata (if you dare!)
- ISO 19115 (geospatial metadata)
- MIX (image technical metadata)
- schema.org microdata
- RDF/XML

Final Project – Metadata Schema Choice and Documentation

This project simulates the extremely common real-world situation of a brand-new digitization/data project where nobody is sure which metadata is desirable, what it should look like, or how to collect it.

Due end of Module 2: Collection list. Build a digitized or born-digital collection around a coherent theme of your choice consisting of roughly fifteen *heterogeneous* objects. (By “heterogeneous,” I mean that the collection must contain examples of *at least three types* from the DCMI Type Vocabulary <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/#section-7>, “Collection,” “Event,” and “Service” excluded.) You may certainly put together a “Frankencollection” — a combination of items from more than one source — if you like. Your collection should include born-digital objects, digitized objects, and not-yet-digitized realia. Please avoid published books and journal articles! Manuscripts, letters, maps, websites, and other types of documents found in records-management and archives contexts are acceptable, however.

Ideally, this collection will consist mostly or entirely of un- or underdescribed objects. Family-history collections are encouraged, as are existing un- or underdescribed digital collections. I understand if this isn't feasible, however!

Turn in a list or spreadsheet of the items, for each item giving its DCMI type and a *brief* (think “HTML alt text”) narrative description (e.g. “Image, photo from a family wedding”).

Due end of Module 5: Wisdom From Those Before Us. For *each DCMI type* of object in the collection you built, find at least one **case study** from the LIS literature (blogs from practitioners or cultural-heritage organizations count!) for a project that includes that object type *and* discusses metadata use and decisionmaking surrounding it. Cite (APA, MLA, or Chicago is fine)

the case study and summarize its *metadata-related* practices, decisions, and difficulties in a paragraph or two, including all content standards, structure standards, and controlled vocabularies mentioned. For *each* mentioned content standard and controlled vocabulary (but not any structure standards), include a link to its primary documentation. (Word to the wise: Bookmark any workflow documentation you find as you work with your case studies; you'll need it later in the course.)

Due end of Module 9: Three Specialized Records. Create the **completest metadata records** (largely descriptive, but include as much administrative and technical metadata as possible without inventing information you don't have or committing tag abuse in a record) you can for *one example object of each of three DCMI types* from your collection, in one or more schema(s) that are **not Dublin Core**. (A Dublin Core expansion such as PBCore or NDLTD thesis/dissertation metadata is fine as long as it's appropriate to the object, however.) Record each description in one of the metadata structure standards we have learned in this course (an XML-based language, N-Triples or Turtle, or other as appropriate; *a spreadsheet/CSV file is not good enough*), and make sure you validate the structure and syntax as best you can.

You may use different content standards, controlled vocabularies, and structure standards for each object, if you think that best, or even if you just want the practice! (This is subject to a basic sniff test for appropriateness: for example, if you use VRA Core on a software object, I will not be best pleased.)

This is a lab practical, not a quiz or exam. *Of course* you may use any documentation you can find! *Of course* you may ask questions when that documentation is unclear!

Due end of Module 10: More Metadata. For each of the remaining objects in your collection (up to 20 items; don't overwork yourselves!), create a record that **EITHER** aligns with one you used for your three example objects, **OR** is a Simple Dublin Core record in the XML-based Dublin Core structure documented at <https://www.dublincore.org/specifications/dublin-core/dc-xml-guidelines/> Again, you may mix and match approaches — this wouldn't be feasible in most real-world projects, but here it's fine.

Due end of Module 11: Crosswalk Documentation. Create **documentation for a crosswalk** from the standards used in your three metadata records to records conforming to either (not both) of the Europeana Publishing Guide <https://pro.europeana.eu/post/publication-policy> or the Recollection Wisconsin Metadata Essentials <https://recollectionwisconsin.org/wp-content/uploads/2016/12/RecollectionWisconsinMetadataEssentials.pdf>. Focus on content and controlled vocabularies over structure (that is, you may assume that the programming script that actually performs the crosswalking is Somebody Else's Problem), but try to note any issues that cannot be solved computationally and will require manual intervention.

You may use any of the example crosswalks in <https://guides.lib.utexas.edu/metadata-basics/crosswalks> as a guide for your work.

Due end of Module 12: Workflow Documentation. Locate any existing **documentation for metadata creation and/or metadata workflows** for the three (or more) case studies you found. If not all three have relevant documentation (it happens!), fill in by locating one or more set(s) of metadata-creation documentation (written by libraries, archives, museums, consortia, or professional organizations) that could plausibly be applied to your collection, in whole or in part. Compare, contrast, and critique the documentation sets you located on the following features:

- Comprehensiveness
- Comprehensibility
- Appropriate use of content standards and controlled vocabularies (do they? should they be using a standard they're not? are they twisting a standard or CV in unjustifiable ways?)
- Usability (if you had a question while working on metadata, how easy would it be to find the answer?)
- Visual/graphic design (is it helpful to metadata creators? is it accessible?)
- Quality and quantity of examples

Due at end of course: Wrapping Up. Your final assignment is a reflection and augmentation of your work to date, in a rough case-study format. Think of it as informing your start-of-course self how to do this project! It should include

- an executive summary (absolutely no more than one page) with the most important things someone attempting a project like this should know about metadata
- all previous assignments — revised as needed; sometimes you learn things between your first turn-in and the end of the course
- metadata-relevant characteristics and affordances of the objects you studied
- metadata-relevant characteristics of at least one primary user community for the collection you built, including their information behavior and needs
- the metadata records you created, as separate plain-text files (please don't paste them into a word-processing document! that ruins them!)

I would guess this final document to be about 5-7 pages (not including the actual records), but depending on how you design it and how much you know about relevant communities, it could be fewer... or much more. I don't grade by weighing your papers — brevity is the soul of wit! If you think you've completed the assignment adequately, it's likely you have.

Course schedule

Linklists are for enrichment and reference; you are **NOT** required to read everything on them.

Unit 1: Building blocks

Module 1: Course introduction. What is metadata? Metadata jobs.

Topics: Course structure, assignments, assessment, policies. Defining metadata, metadata vs cataloging, metadata is for users, types of metadata. Metadata labor. Standards. Basic best practices in metadata creation: authority control and name representations, date representations (ISO 8601), place names.

Tech topics: Text editors, spreadsheets, XML editors, and other tools in metadata; software installation. When and why Excel is the absolute worst. The importance of plain text; creating and looking at CSV files from spreadsheets.

Linklist(s): <https://pinboard.in/u:dsalo/t:metadata>, <https://pinboard.in/u:dsalo/t:csv/>

Watters/Oates. “The search for a minimum viable record.” <https://www.forbes.com/sites/oreillymedia/2011/06/01/the-search-for-a-concise-and-all-encompassing-library-record/?sh=ba2aebc43332>

Gartner. *Metadata*, chapters 1 and 3. (Chapter 2 is super-fun, but not required.)

Gilliland. “Setting the Stage.” https://www.getty.edu/research/publications/electronic_publications/intrometadata/setting.html
(This is a classic, and for good reason.)

van den Burg. “Handling messy CSV files.” <https://towardsdatascience.com/handling-messy-csv-files-2ef829aa441d>

Optional, but eye-opening: Cargill. “Why standardization efforts fail.” <http://quod.lib.umich.edu/j/jep/3336451.0014.103?rgn=main;view=fulltext>

Module 2: Types of metadata standards. What goes into a (metadata) standard?

Topics: Interoperability, and how it drives standards construction. Models; data models. Content vs. structure standards (Gartner calls these “content rules” and “syntax”). Descriptive, technical, administrative, and structural metadata. Controlled vocabularies; enumerations, taxonomies, thesauri, ontologies, synonym rings, authority files. Identifiers; why personal and geographic names are terrible ones; DOIs, handles, ARKs, PURLs.

Tech topics: Identifiers in HTML, relational databases, and RDF/linked data. Identifier sources. URIs as identifiers.

Linklist(s): <https://pinboard.in/u:dsalo/t:identifiers>, <https://pinboard.in/u:dsalo/t:authoritycontrol>, <https://pinboard.in/u:dsalo/t:standards>

Gartner, *Metadata*, chapter 5.

Bilder. “DOIs unambiguously and persistently identify published, trustworthy, citable online scholarly literature. Right?” <https://www.crossref.org/blog/does-unambiguously-and-persistently-identify-published-trustworthy-citable-online-scholarly-literature-right/>

McKenzie. “Falsehoods programmers believe about names.” <http://www.kalzumeus.com/2010/06/17/falsehoods-programmers-believe-about-names/>

OCLC. “Registering researchers in authority files.” <https://www.oclc.org/content/dam/research/publications/library/2014/oclcresearch-registering-researchers-2014.pdf>

Meadows. “Everything you ever wanted to know about ORCID.” <http://crln.acrl.org/index.php/crlnews/article/view/9428/10644>

“About the ULAN.” <https://www.getty.edu/research/tools/vocabularies/ulan/about.html> (skip History)

Paskin. “On making and identifying a ‘copy.’” <http://www.dlib.org/dlib/january03/paskin/01paskin.html> (As you read, keep the question “what exactly does an identifier actually identify?” in mind.)

Fast/Leise/Steckel. “What is a controlled vocabulary?” <http://boxesandarrows.com/what-is-a-controlled-vocabulary/>

Rayburn. “Taxonomies and thesauri.” <https://web.archive.org/web/20160305124415/https://www.ischool.utexas.edu/~i385e/readings/Warner-aTaxonomyPrimer.html>

Module 3: Content standards for descriptive metadata. Dublin Core.

Topics: ID3v2, PBCore, VRA Core, CCO. EAD. History and goals of Dublin Core. One-to-one principle, “dumb-down principle” (apologies for the ableism; I did not invent this term). Simple vs. qualified DC, DC application profiles. Challenging elements, limitations of Dublin Core.

Tech topics: Creating Dublin Core records. Examining them in spreadsheets, CSV, and OAI-PMH XML.

Linklist(s): <https://pinboard.in/u:dsalo/t:dublincore>, <https://pinboard.in/u:dsalo/t:pbcore>, <https://pinboard.in/u:dsalo/t:ead>

ID3.org “Welcome” and “FAQ.” <http://id3.org/Home> and <http://id3.org/FAQ>

“PBCore FAQs.” <http://pbcore.org/faqs>

“Cataloging Cultural Objects: Part I, General Guidelines” <http://vraweb.org/wp-content/uploads/2020/04/CatalogingCulturalObjectsFullv2.pdf>

Baker. “A grammar of Dublin Core.” <http://www.dlib.org/dlib/october00/baker/10baker.html>

Scott. “How much description is enough? A brief history on the debate over the Dublin Core Metadata Initiative.” http://www.eliotscott.com/documents/dublin_core.pdf

Urban. “Principle paradigms: revisiting the Dublin Core 1:1 principle.” <https://hdl.handle.net/2142/31109> (Abstract only.)

Chapter 3, “Using Dublin Core” in *The metadata manual: a practical workbook*. (See Canvas.)

Coyle/Baker. “Guidelines for Dublin Core Application Profiles.” <https://www.dublincore.org/specifications/dublin-core/profile-guidelines/>

Module 4: Administrative metadata: preservation, technical, and rights metadata

Topics: Defining administrative metadata, rights management metadata, preservation metadata. EXIF, PREMIS, provenance metadata, event, entity, agent, activity. Quick tools for determining rights for cultural objects, RightsStatements.org, localcontexts.org.

Tech(ish) topics: Reading PREMIS records. Extracting EXIF data. Machine-readable licenses (e.g. from CC). Examining copyright/licensing-metadata issues in a deed of gift.

Linklist(s): <https://pinboard.in/u:dsalo/t:metadata/t:copyright>, <https://pinboard.in/u:dsalo/t:premis>

“EXIF data explained.” <http://www.photographymad.com/pages/view/exif-data-explained>

Chastain. “About EXIF, IPTC, and XMP use in graphics software.” <https://www.lifewire.com/what-is-metadata-1701735> (N.b. I STRONGLY DISAGREE that photo metadata shouldn’t be worried about!)

“Rights statements.” <https://rightsstatements.org/en/>

Whalen. “Rights Metadata Made Simple.” http://www.getty.edu/research/publications/electronic_publications/intrometadata/rights.pdf

“A primer on PREMIS and PREMIS Rights.” <https://archival-integration.blogspot.com/2016/02/a-primer-on-premis-and-premis-rights.html>

Phillips/Alemneh. “Case Study: Implementing an Open-Source and In-House Developed PREMIS Events and Agents System.” https://doi.org/10.1007/978-3-319-43763-7_17

“Recommendations for standardized international rights statements.” http://rightsstatements.org/files/160208recommendations_for_standardized_international_rights_statements_v1.1.pdf

“Traditional Knowledge (TK) labels.” <https://localcontexts.org/labels/traditional-knowledge-labels/>

“Social Humans.” <https://www.docnow.io/social-humans/> (Click through to the SH-C and SH-A label sets.)

Module 5: Structure standards; storing metadata structures

Topics: Data (structure) models: log files, tabular data, relational databases, (meta) markup, RDF-linked data. Advantages and disadvantages of these data models. Sources of metadata: manual generation, automatic generation, harvested metadata, converted metadata, user-contributed metadata. External storage, internal storage. Functional requirements, search and browse options.

Tech topics: Introduction to relational data modeling and relational databases.

Linklist(s): <https://pinboard.in/u:dsalo/t:models>, <https://pinboard.in/u:dsalo/t:751>, <https://pinboard.in/u:dsalo/t:sql>

Hooland/Verborgh. “Chapter 2: Modelling,” sections 1 through 6. *Linked data for libraries, archives and museums: how to clean, link and publish your metadata*. (Library ebook: <https://search.library.wisc.edu/catalog/9911170142102121>.)

Gilbert/Mobley. “Breaking Up With CONTENTdm: Why and How One Institution Took the Leap to Open Source.” <http://journal.code4lib.org/articles/8327>

Zeng/Qin, *Metadata*, Section 3.5 “Metadata Sources” (see Canvas)

Zeng/Qin, *Metadata*, Section 3.6 “Metadata Storage” (see Canvas)

Microsoft. “NCSA Logging.” <https://docs.microsoft.com/en-us/windows/win32/http/nca-logging>

Unit 2: Making metadata

Module 6: A (very) short introduction to linked data.

Topics: Escaping information silos. RDF. Ontologies. Linked data. Triples as statements. URIs as identifiers. The “open world assumption.” Reconciliation. SKOS.

Tech topics: RDF, N-Triples syntax, Turtle syntax, prefix declarations. Wikidata. (N.b. We will not be authoring linked data in this course — only learning to read it — and I am not even touching JSON-LD or RDF/XML. Take my linked-data one-credit if you want exposure to these.)

Linklist(s): <https://pinboard.in/u:dsalo/t:linkeddata>

Roke/Kitchin Tillman. "Pragmatic principles for archival linked data." <https://doi.org/10.17723/2327-9702-85.1.173>

Tauberer, "What is RDF and what is it good for?" <https://github.com/JoshData/rdfabout/blob/gh-pages/intro-to-rdf.md#>
(Sections 1-4; pay special attention to Tauberer's answers to the second question in his title, please.)

Noy and McGuinness. "Ontology development 101." <http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness-abstract.html>

Kelley, Michael. "How the W3C has come to love library linked data." <https://www.libraryjournal.com/story/how-the-w3c-has-come-to-love-library-linked-data>

W3C. "SKOS primer." <https://www.w3.org/TR/skos-primer/> (Through section 3; ignore the rest.)

Module 7: Introduction to XML.

Topics: history of XML, function vs. format (structure vs. presentation, markup vs. makeup), well-formed, valid, element, attribute, DTD/schema, namespace, common encoding errors

Linklist(s): <https://pinboard.in/u:dsalo/t:xml>

Tech topics: Well-formed and valid XML documents.

Kitchin Tillman. "Understanding EAD and XML." <https://eadiva.com/understanding-ead-and-xml/>

W3C. "XML in Ten Points." <https://www.w3.org/XML/1999/XML-in-10-points-19990327>

Hawkins. "Introduction to XML for Text." <http://www.ultraslavonic.info/intro-to-xml/>

Harold/Means. "How XML Works." *XML in a Nutshell*. <https://search.library.wisc.edu/catalog/9912614244102121>

Harold/Means. "XML Fundamentals." *XML in a Nutshell*.

Module 8: Applications of XML: MODS and family.

Topics: comparing MARCXML, MODS and DC; MARC and MODS; MODS, Dublin Core in XML

Tech topic: XML namespaces. XML schemas. Schema vs. namespace URLs; building xsi:schemaLocation. Using language documentation to avoid common misunderstandings about MODS and VRA Core rules.

Linklist(s): <https://pinboard.in/u:dsalo/t:mods>

"DLF/Aquifer Implementation Guidelines for shareable MODS records." https://wiki.dlib.indiana.edu/download/attachments/24288/DLFMODS_ImplementationGuidelines.pdf (Introduction and Summary of Requirements and Recommendations; the rest is for your reference.)

"MODS Guidelines Levels of Adoption." <https://wiki.dlib.indiana.edu/display/DLFAquifer/MODS+Guidelines+Levels+of+Adoption>
(Honestly, level 5 is hugely expensive overkill. Aim for level 4 as you create MODS records, but sometimes level 3 with pieces of level 4 will be the best you can do.)

Allain. "Converting spreadsheets into MODSXML using Open Refine." <http://digitalscholarship.utoronto.ca/content/blogs/convert-spreadsheets-modsxml-using-open-refine> (DO NOT lose yourself in the technical details, please! Work to understand Allain's goals and how OpenRefine makes them possible. I'll have a simplified example of this technique later in the course for you. If you'd like to try this technique as your MODS creation tool, however, go for it! Allain includes links to the spreadsheet and template, and there is another at https://github.com/calhist/mods_xml/blob/master/openrefine_mods_template.xml.)

Unit 3: Applying metadata

Module 9: Designing/choosing/documenting a metadata schema.

Topics: Focusing on user needs, user behavior, corporate culture. Selecting and developing an element set, adapting vs. starting from scratch, general vs. domain specific element sets, factors in choice of metadata element set, mandatory vs. optional. Content guidelines, documenting your schema. Examples of design process and documentation. Copyright and metadata.

Tech topic: Creating MODS and VRA Core records.

Miller. Chapter 10: Designing and Documenting a Metadata Schema. (See Canvas.)

Bowen. "RTFM? How to write a manual worth reading." <https://opensource.com/business/15/5/write-better-docs>

"Principles for evaluating metadata standards." <https://web.archive.org/web/20160807051928/http://metaware.buzz/2016/08/04/principles-for-evaluating-metadata-standards/>

Godby, Smith, and Childress. "Toward element-level interoperability in bibliographic metadata." <https://journal.code4lib.org/articles/54> (Now imagine all the non-bibliographic metadata out there...)

Greenberg et al. "Metadata ownership and metadata rights." <https://www.slideshare.net/slideshow/view?login=chelciasansmerci&preview=no&slideid=1&title=metadata-ownership-metadata-rights>

Module 10: Managing metadata quality.

Topics: What is data quality? Functional requirements, quality and granularity. Measuring data quality. Tools for quality control. Training. Acceptably bad quality. Metadata remediation; metadata enhancement; OpenRefine.

Tech topic: Regular expressions.

Linklist(s): <https://pinboard.in/u:dsalo/t:openrefine>

Gentry et al. "Survey of benchmarks in metadata quality: initial findings." https://dlfmetadataassessment.github.io/assets/WhitePaper_SurveyofBenchmarksinMetadataQuality.pdf (Read Results sections 2 through 4, Discussion, and Conclusion.)

"The Quartz guide to bad data." <https://github.com/Quartz/bad-data-guide> (What can go wrong? Any or all of this!)

Hooland/Verborgh. "Chapter 3: Cleaning." In *Linked Data for Libraries, Archives and Museums: How to Clean, Link and Publish Your Metadata*. Chicago: ALA Editions, 2014.

Bruce/Hillmann. "Metadata quality in a linked data context." <https://blog.law.cornell.edu/voxpath/2013/01/24/metadata-quality-in-a-linked-data-context/>

Phillips/Tarver/Frakes. "Implementing a collaborative workflow for metadata analysis, quality improvement, and mapping." <http://journal.code4lib.org/articles/9199>

Module 11: Sharing, crosswalking, and harvesting metadata.

Topics: federated searching vs aggregation, crosswalks, mapping, how the mapping process works (you win some, you lose some), sharing metadata - solutions, OAI-PMH, METS, PCDM.

Tech topic: Crosswalking.

Helland. "The power of babble." <https://queue.acm.org/detail.cfm?ref=rss&id=3003188>

Woodley. "Crosswalks, metadata harvesting, federated searching, metasearching." In *Introduction to metadata* <https://www.getty.edu/publications/virtuallibrary/0892368969.html> (Ignore the "Metasearching" section. Metasearching is a hack that doesn't work. Skip it and go straight to the four case studies and conclusion.)

Shreeves/Riley/Milewicz. "Moving towards shareable metadata." <http://firstmonday.org/ojs/index.php/fm/article/view/1386>

Neatrou et al. "A clean sweep: the tools and processes of a successful metadata migration." <https://doi.org/10.1080/19322909.2017.1360167>

Gregory/Williams. "On being a hub: some details behind providing metadata for the Digital Public Library of America." <http://www.dlib.org/dlib/july14/gregory/07gregory.html>

Gueguen et al. "Metadata aggregation." <https://dp.la/info/2015/01/28/metadata-aggregation-webinar-video-and-extended-qa/> (Just the Q&A; you do not need to watch the webinar.)

Module 12: Metadata standards and practices outside GLAM

Topics: DarwinCore; eagle-i ontology as cautionary tale (moral: do the simplest thing that can possibly work). DDI. TEI and the TEI Header. CodeMeta. ONIX; BISAC. Data dictionaries. Corporate metadata.

Tech topic: OpenRefine 1.

UW-Madison Research Data Services. Metadata. <http://researchdata.wisc.edu/metadata/>

"About eagle-i." <https://open.catalyst.harvard.edu/products/eagle-i/>

EDItEUR. "ONIX FAQs." <http://www.editeur.org/74/FAQs/>

Tarsala. "BISAC basics." <https://web.archive.org/web/20151214154651/https://cbtarsala.wordpress.com/2014/11/16/bisac-basics/>

CodeMeta. "CodeMeta Developer Guide." <https://codemeta.github.io/developer-guide/> and take a look at a couple of its crosswalks <https://codemeta.github.io/crosswalk/>

Briney. "README.txt" <http://dataabinitio.com/?p=378> and "Data dictionaries." <http://dataabinitio.com/?p=454>

Digital Curation Center. "Disciplinary metadata." <http://www.dcc.ac.uk/resources/metadata-standards> (and skim a few from the lists: minimally, DDI, DarwinCore, EML.)

Hawkins. "Introduction to the TEI Header." <https://www.ultraslavonic.info/intro-to-tei-header/>

Emmelhainz. "Adding metadata to a qualitative data project." <https://web.archive.org/web/20181110182800/https://databrarians.org/2015/11/adding-metadata-to-a-qualitative-data-project/>

Mannheimer. "Using data dictionary creation as the teaching moment for metadata." <https://rethinking.clir.org/blog/using-data-dictionary-creation-as-the-teaching-moment-for-metadata/>

Roszkiewicz. "Enterprise metadata management: How consolidation simplifies control." <https://doi.org/10.1057/dam.2010.32>

Inmon, William H, Bonnie K O'Neil, and Lowell Fryman. "Chapter 1: Introducing Business Metadata." *Business Metadata: Capturing Enterprise Knowledge*. (Library ebook: <https://search.library.wisc.edu/catalog/9911039061202121>)

Unit 4: The human side of metadata

Module 13: Equity and inclusion in metadata

*Topics: Inclusion as a quality measure; the power of naming; bias in controlled vocabularies; euphemism in controlled vocabularies; diverging from controlled vocabularies that won't fix themselves. Inclusive vocabularies; vocabularies designed in and by marginalized communities. Fixing *-ist legacy metadata. Authority control and deadnaming. Automated metadata and its inclusion pitfalls; the many and repeated inclusion failures of ML/AI in metadata generation. Ethics in metadata labor; unethical labor practices (hope labor, student labor, incarcerated labor, precarious labor, unpaid labor from marginalized communities); crowdsourcing.*

Tech topic: OpenRefine 2. Linked-data reconciliation.

Linklist(s): <https://pinboard.in/u:dsalo/t:metadata/t:inclusion>, <https://pinboard.in/u:dsalo/t:controlledvocab/t:inclusion>, <http://pinboard.in/u:dsalo/t:601/t:labor>, <https://pinboard.in/u:dsalo/t:metadata/t:tradcultures>

Reidsma. "Algorithmic bias in library discovery systems." (So we're going to use similar algorithms to automate metadata creation? Not on my watch we're not.) <https://matthew.reidsrow.com/articles/173>

Johnson/Forsythe. "Disability and accessibility language in subject headings and social tags." https://cdn.ymaws.com/www.cilip.org.uk/resource/collection/2547D534-4057-449C-BBFF-0095E4CF5AEF/ci_197_johnson_forsythe_disability.pdf

Lellman et al. "Guidelines for inclusive and conscientious description." <https://wiki.harvard.edu/confluence/display/hmschommanual/Guidelines+for+Inclusive+and+Conscientious+Description>

Trans Metadata Collective. "Metadata best practices for trans and gender diverse resources." <https://zenodo.org/record/6686841>

Geraci. "Programmatic approaches to bias in descriptive metadata." <https://osf.io/9uehx/> (Following linked resources strongly recommended! Lots of good ones in here!)

Duarte/Belarde-Lewis. "Imagining: Creating Spaces for Indigenous Ontologies." <http://doi.org/10.1080/01639374.2015.1018396>

Mattson. "Creating metadata by hand: musings on the limits of automation in archives." <https://pushcartcatalog.wordpress.com/2017/05/25/creating-metadata-by-hand-musings-on-the-limits-of-automation-in-archives/> (Because you will very likely have to defend the value of your labor.)

Logsdon. "Ethical digital libraries & prison labor?" <https://osf.io/yqpk/>

Collective Responsibility Labor Forum et al. "Collective equity! A handbook for designing and evaluating grant-funded positions." <https://doi.org/10.26207/6p4a-md61>

DiPressi et al. "A student collaborators' bill of rights." <https://humtech.ucla.edu/news/a-student-collaborators-bill-of-rights/>

Module 14: Training and human resources for metadata.

Topics: Networks of practice, technological frames, project management, generational issues, MARC to non-MARC transition. Metadata careers: trends in careers and job postings. Planning your post-graduation professional development, self-teaching.

Tech topics: JSON and JSON-LD. Translating between RDF syntaxes (not by hand!!!! life is too short for that nonsense).

Chopey. "Planning and Implementing a Metadata-Driven Digital Repository." <http://scholarspace.manoa.hawaii.edu/bitstream/10125/337/1/0409CCQPlanningImplementing.pdf>

Khoo/Hall. "Managing Metadata: Networks of Practice, Technological Frames, and Metadata Work in a Digital Library." *Information and Organization* 23:2: 81–106. <https://doi.org/10.1016/j.infoandorg.2013.01.003> (Start at section 3.2 and skip section 4; the literature review and methods are uninteresting for our purposes.)

Thompson. "What If I Break It?" Project Management for Intergenerational Library Teams Creating Non-MARC Metadata. <http://journal.code4lib.org/articles/10395>

Phillips. "Managing metadata editing for telecommuting." <https://vphill.com/journal/post/6212/>

Mackeil-Pepin. "On addressing weaknesses." <https://hacklibraryschool.com/2015/01/28/on-addressing-weaknesses/>

Bureaucratic stuff you mostly don't care about that the Powers that Be make me put in my syllabi

Credit hours and regular and substantive student-instructor interaction

Students completing this course will earn three credit hours. 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 class carries the

expectation that students will work on course learning activities (reading, writing, problem sets, studying, etc) for about 9 hours each module.

The credit standard for the course is met by an expectation of a total of 135 hours of student engagement with the course learning activities (at least 45 hours per credit). Students should expect 150 minutes per week (in Canvas video) of lecture, full-class and small-group discussion on Canvas moderated by the instructor, and individual or small-group activities with outcomes reportable to (and assessed by) the instructor. Students should expect to work on course learning activities (reading, writing, studying, etc) for about three hours out of the classroom for each class period, doing readings, individual assignments and projects and other student work as described in the syllabus.

This course will provide regular and substantive student-instructor interaction in the following ways:

- The instructor provides direct instruction weekly in lecture;
- The instructor will assess and provide substantive feedback on student coursework regularly as assignments are due;
- The instructor will regularly provide information and respond to questions about the content of a course through weekly instruction in lecture, regular office hours, and communication through email and Canvas;
- The instructor will facilitate group discussions and group critiques related to course content through weekly direct instruction on Canvas.

Learning outcomes table

iSchool program-level learning outcome	How will mastery of learning outcomes be assessed?
2. Students apply principles of information organization.	Dublin Core, MODS, VRA Core assignments in per-module technology work Metadata creation as part of Final Project – Metadata Schema Design and Documentation
5. Students demonstrate competency with information technologies important to the information professions.	Per-module technology work (includes work on relational-database modeling, various XML languages, Excel/CSV, regular expressions, and OpenRefine).