SEPTEMBER–DECEMBER 2021

01–09 CLASS / LECTURE NOTES

10–25 STUDENT FEEDBACK

EMMA METCALFE HURST
ARST 556L: METADATA
DR. JULIA BALLARD
FINAL PROJECT

CLASS / LECTURE NOTES & STUDENT FEEDBACK

CLASS / LECTURE NOTES

Sept 9, 2021 - Class 1 - Notes

- Individual Plan
  - Customize your individual student plan
- Job Search
  - Look for job ads with metadata - note competencies to help inform what you want to get out of this course
- Skills to take note of:
  - Familiarity with DAMs and systems (FileMaker Pro, Islandora, Atom, Arca)
  - Cataloguing and cataloguing systems
  - Collaboration skills
  - Stats & reports
  - Decolonial frameworks
  - Digital collections
  - Documentation, training, consultation
  - Policies & terminology
  - Instructions and consultations
  - Metadata standards & vocabularies
  - Archival standards & principles
  - Analyzing / remediating large sets of metadata using tools (OpenRefine, MarcEdit, OCLC Connection, XML Editor)
  - "Sustainable, extensible, linked open-data (LOD)
  - .csv data entry files + Omeka S
  - Creating authority records
  - Project mgmt => workflows
  - FIle mgmt systems => Sharepoint, Microsoft Teams
  - Data formats => .csv, .xml, .json
  - Object metadata + metadata tags
  - Taxonomies / standards
  - Tools for remediating and analyzing data (Open Refine)
  - MS Office Tools
  - Lots of metadata jobs in healthcare, STEM, information analysis and cloud computing

CLASS / LECTURE NOTES

- Metadata standards and vocabularies (DublinCore, MODS, OCLC-FAST, AAT, GeoNames, Wikidata, RDA, DACS)
  - Python and SQL Language
  - Data visualization
  - Cloud-based digital record mgmt system
- Goals:
  - Familiarity with metadata standards
  - Understand role of metadata in a particular context

Sept 16, 2021 - Class 2 - Notes

- Terms from job search:
  - LOD Model - Linked (Open) Data
  - DAMs
  - Oxygen
  - OCLC Connexion
  - Open Refine
  - Marc Edit
  - XML Editor
  - PDCore
  - JSON
  - Data migration
  - Data Quality Assurance
  - Crosswalking
  - DACS
  - EAD
  - DCRM-B
  - Object Metadata
  - RDA - AACR2 - MARC - LSCH - TEI - .CSV (Library Descriptive work)
  - Atom built on EAD
    - Critical for user-generated metadata, crowd-sourced metadata to make sure it's consistent
  - DAMS => can work with Collective Access next week
- "Remediating" - "perform" - "knowledge of" - "familiarity with" - "experience with" - "provide technical training" - "collaborate with"

- "negotiate with" - "assists in" - "working on" - "languages" (highly desirable for metadata) - "training" (hands-on training, metadata compliance, training) - "knowledge of" / "familiarity with" (know the nature of thing, how it relates to a job interview / know what it is)
- What is your ideal version of this course? What do you want to be able to show ./ have at the end?
- Ideas:
  - Review / assessment of The Marks Project
    - Controlled vocabulary, clean up metadata
    - Recommend DAMs
    - Searchability functionality
  - Collective Access
  - Assessment of what DAMs work best for visual metadata
    - Accommodation of wider character sets
    - Non-Roman character sets / unicode
- Consider crowd-sourcing / AI for generating metadata

Lecture Notes - Sept 15, 2021 - Types of Metadata

- Metadata Wishlist - one wish to search
  - I wish I could determine the maker of ceramic pots
  - I wish I could find recipes for fruits / vegetables / food that I currently have in my fridge
  - I wish I could determine how many times my podcast episodes were played
- Types of Metadata (Typologies)
  - Stratford - Layne - Image Attributes
  - NISO's Types
    - Descriptive - Structural - Administrative - Markup Languages
- Pomerantz lays out schemas:
  - Descriptive - Administrative - Technical (structural) - Use
- Schemas
  - Adoption because of different technologies => quick acceleration => eventually becomes ubiquitous.
    - Relationships between tech and metadata

- "negotiate with" - "assists in" - "working on" - "languages" (highly desirable for metadata) - "training" (hands-on training, metadata compliance, training) - "knowledge of" / "familiarity with" (know the nature of thing, how it relates to a job interview / know what it is)
- What is your ideal version of this course? What do you want to be able to show ./ have at the end?
- Ideas:
  - Review / assessment of The Marks Project
    - Controlled vocabulary, clean up metadata
    - Recommend DAMs
    - Searchability functionality
  - Collective Access
  - Assessment of what DAMs work best for visual metadata
    - Accommodation of wider character sets
    - Non-Roman character sets / unicode
- Consider crowd-sourcing / AI for generating metadata

Lecture Notes - Sept 15, 2021 - Types of Metadata

- Metadata Wishlist - one wish to search
  - I wish I could determine the maker of ceramic pots
  - I wish I could find recipes for fruits / vegetables / food that I currently have in my fridge
  - I wish I could determine how many times my podcast episodes were played
- Types of Metadata (Typologies)
  - Stratford - Layne - Image Attributes
  - NISO's Types
    - Descriptive - Structural - Administrative - Markup Languages
- Pomerantz lays out schemas:
  - Descriptive - Administrative - Technical (structural) - Use
- Schemas
  - Adoption because of different technologies => quick acceleration => eventually becomes ubiquitous.
    - Relationships between tech and metadata

CLASS / LECTURE NOTES

- Types of Metadata
  - What types of info fall into the category
  - Its relationship to a current owner/provider
- **1. Descriptive Metadata**
  - Everything you can say about a resource:
    - what you're doing to do / can do with it
    - its relationship to current owner / provider
  - "Catch-all" category which covers "everything"
  - Doesn't cover relationships; only the <u>thing</u>
  - Includes: TiHe, Aboutness, Depicted, Type (format), item-item relationships
    - Exp. Dublin Core - primarily concerned with resource discovery
      - Search by title, subject
        - Judgement / quality of things
- **2. Administrative Metadata**
  - Represents and helps manage object "life cycle"
  - Naming rights, permissions, ownership, use
  - Answers Qs about histories and agents
  - Who does item belong to / provenance / source
    - Exp. Traditional Knowledge Lables
      - Represents Indigenous communities relationship
      - Dublin Core usually uses Creative Commons
        - Exp. Wikipedia edits and what and by who
- **3. Technical Metadata**
  - Metadata that is automatically collected / created
  - Pertains to technicality and materiality of object
    - Often connected to preservation
    - Exp. size, file format, encoding, history of software programs (develop with Admin metadata)
      - Physical nature of things
- **4. Structural Metadata**
  - Sometimes categorized as Technical Metadata subtype
  - Structures within / among items in a collection
    - Exp. Digitized books having chapter in books

- **5. Use Metadata**
  - Most popular type of metadata
  - Is it <u>being</u> used? Not <u>can they</u> use it (Administrative Metadata)
  - <u>Who</u> is using it, <u>when</u>, and <u>for what</u> purpose?
    - Exp. Wiki usage info ("paradata" - "data exhaust")

Lecture Notes - Sept 22, 2021 - Metadata Design

- Main considerations when you design, create, or reuse metadata schema
- Info / Tech / PPL
- Metadata Design = Layers of perspectives + Compromises + is often re-designed (adapting technologies and systems)
- Starts abstract and conceptual
- Diagram that looks at the entities and relationships in library catalogues
- Institutional Context => Software => Metadata Schema
- Conceptual Constraint => Technical Constraints => Human Info Interaction (People creating metadata; users)
  - Metadata focus is often top-level, other two are influential
  - How concepts interact with software, data formats, communication)
- Conceptual Work
  - What are the key attributes (components) to represent about the object?
    - What makes the object distinct?
    - What attributes align to users' expected needs and interests?
  - What are the relationships among attributes about the object?
  - What is the authoritative source of information about these attributes?
    - Provenance
  - What existing standards and approached exist that address all or the most critical of these concerns?
  - How much would have to be changed from existing standards in institution to capture key attributes of these resources?

- Interoperability
- Technical Constraints and Affordances
  - What limitations (technology) are in place and need to be implemented for consistency?
  - What limitations should be put on:
    - Types of input for values
    - The length of input for values (min+max)
  - Can attributes be made required, options, or recommended?
  - What kinds of connections can be drawn between entities and attributes?
  - What needs are there for particular values to be present, formatted, parsed for interoperability?
  - What tech infrastructure is already in place and for similar collections?
  - What character sets and syntax (adaptations) are comparable to add? Is it modular?
- Human Information Interaction (Human Context)
  - How much judgement can individual workers have and how much trust? In interpreting the object and it's attributes
  - How should metadata be restricted, protected, open, retained, destroyed
  - What agency should future workers and users have to revise individual records, attributes, and parameters, and oversight of scheme => how will it adapt?
  - How can design work of the system be made visible?
  - What expectation do users have of how items will de described and made available?
  - What training do workers have on existing standards and systems for metadata?
  - How much support do users and workers in this context have for learning new systems and support users
    - Exp. DTES RAP
      - Metadata re-design project
      - Conceptual work: Research Objects, Controlled Vocab, Access Attributes, Relational Attributes

- Technical constraints:
  - Built on UBC's institutional repository, CIRCLe and has to conform to metadata guide - interest in displaying more objects directly from records - need to streamline jumps between related works
- Human Information Interaction:
  - Researchers and local community (main users) => needs may be different; what will this mean for the collection of use metadata and ongoing labour of portal? Relevance of research to community and ethical Qs of collecting data

Class 3 - Notes - September 23, 2021

- Areas you could focus on for individual project plan:
  - Conceptual
    - Controlled vocab
  - Technical
    - Systems you have to work with => may restrict
    - Character sets => documentation of vendor products
  - Human Computer Interaction
    - Proposal for research => consultation project with community members, budget, scope
  - What is needed here, what are the relationships?
  - Re-explaining how something works / didn't work is okay
  - What is worth describing? What training is needed? What capabilities?

Extra
- RDF Framework
- Linked Data - best practices for using the web to link information using RDFs and IRIs
- Semantic Web - Coined by Time Berners-Lee, made www.
  - "The web will reach it's full potential when it becomes an environment where data can be shared and processed by automated tools as well as by ppl"
    - Built on XML and RDF and Ontologies as building blocks

CLASS / LECTURE NOTES

- IRIs - Internationalized Resource Identifiers
    - Fundamental components of the current web and foundation of semantic data
    - RDF subjects and predicates should be IRIs and URIs
- RDF Datatypes:
    - RDF / JSON
    - RDF / XML
    - TTL - RD triple language
- RDF Triple Scores
    - No SQL, flexible
    - Subjects and objects stored as nodes, predicates as edges
    - Ability to interpret data and uncover hidden relationships

STUDENT FEEDBACK

To: Amber,

1.Generally, did you find searching easy or difficult? Was there anything you found confusing or challenging?

Design wise, I thought the user interface was very clean and legible. I also thought the choice to put the search bar in the centre of the database header encouraged me as a user to engage with it because it was highlighted as a main feature. At first, after clicking on the author's name, I found some of the language a bit confusing when trying to differentiate between "List of Authors," "Creator," and "Editor." I wonder if it would consider substituting "Creator" with "Author" and maybe "Contributing Authors" instead of "List of Authors" for more clarity? Given you are working with sci fi books / publishing, it would make sense (to me!) to use the language users would be more familiar with instead of "archives speak."

2. What do you think of the location of the Advanced Search function in the Main Menu?

I thought the advanced search function was too buried. As a librarian, I usually go to the advanced search function right away if there is a specific item I am looking for. I would make it available close to the main search bar on the homepage.

STUDENT FEEDBACK

3.Are there any search terms you would like to see for the Advanced Search "search by value" function that have not been included?

 I think including an "Author" (sub for "Creator") value would be very helpful! Even after playing with it, I'm not sure I have a clear idea of what "is/is not resource with ID" does. Given you already have a generous handful of search options in this field, I would consider removing these to avoid confusion.

4.Are there any other fields you think should be browsable other than anthology, person, story, publication date, and publisher?

 Theme or subject tags would be a great way to introduce new readers to short stories / sci fi genre! Curated collections might also be an interested way of filtering through all of the information you have.

5.Did you find any broken links?

 Nope!

STUDENT FEEDBACK

To: Jade,

What a fun project! I enjoyed looking through your drawings to animate some of the material that we've learned in this class. Most of the examples and analogies were clear to me and the drawings supported the concepts well. You used a lot of familiar, everyday examples to make the concepts relatable for your readers. I enjoyed the house analogy to explain databases and the organization of information.

The tone and writing seem more geared towards a middle-school audience level than high-schoolers to me. I would aim for that age range as your target audience.

I'm wondering if you've thought at all about what the final project will look like? Will it be in print? Digital? In a book (zine?) format or letter size pages? Will you do drawings for every section? Include text boxes for the dialogue? I think it would also be beneficial to give your drawings more space because there is currently a lot of text! Another idea that came to mind was potentially including page numbers and an index that allows the reader to locate key concepts.

Excited to see what you come up with for the final product!

STUDENT FEEDBACK

To: Jeremy,

Great to get an introduction to your work / interests in zines! This looks like a fun topic to research as a final project for this class and your robust bibliography affirms that there are a lot of resources for you to look at.

To your questions, the table looks as clear as it can to me (I am not super familiar with MARC field tags unfortunately!), but I am curious to know why you chose to make the "Description" field and "Dimensions" field not required?

Given the unique formal qualities of zines, I would imagine the more physical descriptive information you could provide about the zine would be beneficial to the user and would honour the diverse nature of zines as a format. For example, what if your user is an zinester/artist looking for examples of miniature books? It would be great if they could pull a list of zines from the collection that are unconventionally "small" using the dimensions / descriptive fields. Further, a "Subject" field would also help refine searches. Just a thought!
Lastly, I wonder if you would consider including a cover page image as well? While not textual data, image-based data would also be a great resource to have for reference and also seems to fall in line with the effort to honour the format.

Good luck with your project and excited to see what you do next!

STUDENT FEEDBACK

To: Joy,

It's nice to see a personal touch to this assignment! It must have been fun and nostalgic going back through these images and applying the metadata schema. I am of the same mind as everyone else in that your schema overall is very straightforward and thorough. To your questions, I've added a few responses below:

1.Does the schema makes sense for this use case?
Yes, very much. I would also recommend including an example just to give users a quick and fast idea of what the formatting would look like without having to dig too far into the document.

2. Are there are other ways of searching (aka other properties) I should consider and include into the schema?
Not that I could think of. I believe you included all of the main search properties: date, place(s), subject(s), creator. The "keywords" also broaden the range of unique search terms which is an additional asset. Copyright might be a consideration if you are looking to make these collections public at some point, but I believe you clarified it is for a private album for family members.

3. Do the tables suffice as instructions to the users on how to fill out the property fields? If not, what should

STUDENT FEEDBACK

be changed and how should it be formatted?
In my experience, all of the instructions were clear and concise. My only question was around the inclusion of locations in the "keywords" as that might be redundant information given you've already got multiple, separate property fields for location. Again, examples underneath the property descriptions would likely help expedite the learning process for your users. I also have the same question as Kelly regarding how the controlled vocabularies for the dropdown menus are being developed – particularly for the neighbourhoods. It would be helpful to include a source for your users to know where that vocabulary / understanding is being pulled from.

4. Does the order of things make sense or would it make more sense for some of the properties to be moved before/after others?
Yes! The way you've currently organized it has a nice flow.

5. Try using the schema on one of your own and let me know what was difficult or needed more clarification.
The process worked well for me!

6. Any additional thoughts?
One quick thought: I am wondering if the "Date Created" section belongs in the EXIF table? I assume that information would be embedded in the technical metadata and would likely be less prone do human error if so!
Best of luck with the remainder of your project!

STUDENT FEEDBACK

To: Hannah,

Great to see the return of the SFFSSDBA project! Here is the link to the spreadsheet:
https://docs.google.com/spreadsheets/d/1VhSn8ovZkXFHoeQ CxM892mDnBeTRmI1tWdUT_BVrKjg/edit?usp=sharing.

Additionally, to your questions:

1.Generally, was the schema easy to follow? Why or why not?

Yes, I found the schema that you provided easy overall with the exception of a couple specific formatting questions which I address below.

2. More specifically, is there any parts of the schema that are particularly hard to follow?

This has already been brought up, but I wonder how you would like information entered when it's a list? I used a semicolon (;) to distinguish between entries, but that would be useful to include in the instructions. Further, formatting for Date entries would also be helpful to ensure consistency. Lastly, when including information that is sourced elsewhere (such as Bio) I wonder if it would make sense to include a Source / Link column?

STUDENT FEEDBACK

3. Any wording that needs changing?

"Read Online" in the Short Story spreadsheet was a bit broad. Do you mean where you can ready the short story online / in an e-book format? More clarification / information about what you're looking for in the column would be helpful. On that note, it might also be useful to provide an example of what you're looking for (links? anthology titles?) for the "Other Translation(s)" column as well.

4. Are there any fields that you think are missing from any of the record types?

Subject terms would be helpful for searching!

5. Any other comments on the schema as a whole?

It's looking good to me :)

STUDENT FEEDBACK

To: Hans,

Great to see this come together and to get a better sense of the project that you worked on at the Centre. What an amazing resource and still so much potential for further use and access! To your questions:

1.Is it written persuasively as opposed to informatively?

Yes! I think you have a great balance between providing useful information, clear statements, and outcomes for why the project would benefit both the collection and the Centre. While I too am not familiar with how to write an executive summary, I wonder if the length is appropriate? I know business-models tend to prioritize brevity (I'm thinking of an elevator pitch!), so, in that vein, I would consider cutting it down to make it super succinct. I would also support Estelle's suggestion of using bullet points, specifically on page 3, to make your points more concise and to identify key headers for your reader.

2. Is it written in active tense? Where passive tense is used, is it fine or can/should it be changed to active tense?

I didn't overtly pick up on this. I think it's fine the way it is!

STUDENT FEEDBACK

3. Is it client-centred enough? (Writing in second person was recommended to support this.)

I wonder if "you" could be substituted with "The Centre" as it makes it more clear who you're addressing and acknowledges key outcomes and successes.

4. Is the value proposition of implementing metadata clear to what could be a variety of stakeholders/gatekeepers, some of whom will likely not be as knowledgeable about it as us? Did I make it plain enough to understand? The value proposition of something to a non-profit versus a for-profit is also a little different — did I align the proposal to the client's organizational mission effectively?

Yes, very much so!

I think my last remaining suggestion would be to highlight how this project would create more opportunities to access the collection digitally, which responds to the increased demand for digital resources due to COVID-19. Looking forward to seeing the final result and good luck on the pitch!!

STUDENT FEEDBACK

To: Milan,

Your project is very interesting and you've identified a very significant gap! As collections increasingly move into the digital realm and with access being at the forefront of information organizations' objectives, I think multilingual functionalities will become more and more important. All of that to say, your project seems very relevant and important to the information field at this point in time. This also points to the inherent English-language bias.

1. Your personal experiences with (lacking) foreign language metadata.

While I have not worked directly with non-English metadata, I have been looking more into DAMs that support diacritics which are often used in Indigenous languages. Mukurtu was one such example.

2. Ways out of this situation that you have contemplated or used.

n/a

3. Theoretical approaches that you would suggest to developers of multilingual resources

I would imagine a of plug-in translation feature that could be

STUDENT FEEDBACK

 added to a database would be the most beneficial to translate metadata. While I assume this would not be completely accurate and may improve over time, it would likely be the most efficient way of supporting multi-language functionality without having a person do the manual translation labour. I imagine AI software might also be put to good use to identify specific elements or objects. Lastly, another example that I stumbled across was the use of parent-child metadata fields that would allow a search to take place in two languages, in to different, but linked fields. Adding a search filter for a specific language might also be a useful feature. If you are working with a small collection, perhaps developing a controlled vocabulary of translations may help support multi-language searches and functionality.

Best of luck with this important research and the development of your final project!

STUDENT FEEDBACK

To: Sammy,

Great work on this project!

I think you did a really good job building off of the tutorial that we did in class in a very effective and productive way. It was very helpful to have a step-by-step tutorial of the whole process (finding, downloading, and uploading datasets to Excel and OpenRefine) and I appreciated that you included screenshots as well. I also thought your design was tasteful, minimal, and clear to read. All of the headers were well selected and titled. One suggestion I did have was possibly adding a table of contents for users to be able to navigate to the section they're looking for assistance with more easily. It would be an easy and a useful addition, given you are at the 16 page mark!

Unfortunately, I don't have any extra tips to pass on about OpenRefine, but I thought you did a really great job describing the processes clearly and succinctly! I can definitely imagine myself returning to this document for reference if I find myself using OpenRefine again in the future.

Best of luck with the final product!

STUDENT FEEDBACK

To: Shannon,

Great to see your work with the DTES RAP! They're doing some really interesting projects and I'm glad to see that you are contributing to that work. To your questions:

1. I don't have much experience writing formal project proposals of this type, so my first question is regarding the format. Does the format make sense? Is the order logical and do you think I am missing any important sections? If yes, what? Are there any sections that seem unnecessary?

The format you're using works well. The layout is clear and the index is a useful addition. I would suggest for the header "Expected Results or Impact" would benefit from a slight change to "Results and Impacts" as this makes it makes you sound more assertive and confident in what you're proposing. "Objectives and Expected Results" could also benefit from taking out "expected."

2. Does the project seem reasonable within the given timeframe and stated objectives?

I believe so! With these types of projects, there is always more that can be done, but I think the timeframe that you proposed would give you a good chunk of time to start things off. It might also be relevant to include a caveat that

STUDENT FEEDBACK

relationship-building takes time, so this project will be the beginning of a longer and ongoing process.

3. Does the project seem sufficiently motivated? Is my argument solid? Yes or no and why?

Yes, I think your argument is solid and presents a nice range of potential benefits that would convince your client to undertake this initiative.

4. Anything else to or add or any ideas as to how I could improve the current schema or, alternatively, do you know of any other collections that contain similar items?

Given you've framed this is a community engaged project, I wonder if you are building in opportunities to interface with your users (community members, researchers, students, etc.) to get their feedback on what key words / searches would be beneficial to include in order to meet their search needs. I also think your recommendation to determine the feasibility of accessible image / icon descriptions is a great idea and falls in line with the effort to support visual-based research summaries.

Thanks for sharing, and I look forward to seeing how your project turns out!

STUDENT FEEDBACK

To: Theresa,

I enjoyed you spinning this project into a topic of personal interest and use! I think that enthusiasm helped bring your voice (and humour!) into your draft, making it for an enjoyable read, in addition to it being clearly laid out.

Example 1: Unfortunately I am not a MARC metadata whiz, but it seems to me that you are also addressing a larger problem in the publishing industry; more specifically, the standards (or lack thereof) for acknowledgement. With this in mind, I was wondering if the way you framed your examples (Request; Verdict; Discussion) could be supplemented by another Header, possibly "Recommendations" as a way to acknowledge and call upon industry standards to change? This could also be accompanied by a list that outlines benefits for including this information as metadata. Other folks have also mentioned adding visuals that show your searches in the databases that you explored and what the search results yielded.

Example 2: Again, a Recommendations entry could be interesting to include to share any insights you may have about whether or not the search feature could be refined by IMDB even further! This would be great following the final discussion of the minor issues you address in the last paragraph of Example 2.

Good luck with the rest of your project!