Appendix - Metadata remediation and the MODS to RDF mapping process at Whitman College

Introduction

Early on the Whitman College team made the decision to standardize their metadata fields across both the institutional repository, which contains faculty scholarship and theses, as well as the digital archives, as much as possible. Their users were not making the same distinctions between their digital collections, and they wanted their users to have a fairly seamless searching experience across the entire platform. After doing an analysis of their metadata, the team determined they had 158 unique fields in Islandora 7. Their plan was to analyze each field to determine which fields could be deleted, which fields could be consolidated, what to name each field, and how the use of each field should be standardized. In order to keep track of this sprawling yet highly-detailed project, they used a Trello Board to track their progress and decisions. Each field was assigned a card on the board that linked out to a template form that they used to define the field and how it was used, as well as documentation they pulled together from other public metadata manuals, primarily from colleges and universities, to help guide their own decision-making process.

Metadata Field Remediation

Once a complete list of metadata fields was established, each field was assigned to a Trello card. It was easier to tackle similar fields at once, so they created a single Trello card for each cluster of related fields. Each Trello card had three attachments:

1. An Excel template (IMI_template_use). The first step was to create an Excel sheet to compare side-by-side uses in Islandora. This helped the team get a better understanding of how these fields were used in the various collections, as well as compare against each other.
   a. A combined excel document was created to keep track of all fields, what collections they were used in, and examples of their use (IMI Info List-3.11.20)
   b. A simplified spreadsheet (Arminda Fields) kept track of which fields were used in which collection. If no collection used that field, that field was a candidate to be removed upon discussion.
   c. The Second Trello attachment was a Google Doc template (IMI/template card). After evaluating each field, they focused on what they wanted it to look like moving forward, and noted those decisions in the Google doc template.
2. The third attachment was an RDF mapping template (IMI/template_RDFMapping) (explained in part II).

Tracking Clean-up Actions

After deciding what a particular field should look like going forward, the team discussed what a clean up would require, taking notes on which fields had to undergo which action (Fields that need maintenance done by Paige) This spreadsheet separated each collection out and indicated the action needing to be done by the metadata librarian, and the status of that action.

1. Changes were documented in a Google Doc and Google Sheet (PAM Cards Information) so all stakeholders, including their future selves, had something to reference.
2. The separate documents shared information for specifically targeted people, so as to not overwhelm them with information, and allowed relevant information to remain in the forefront. For example, the document for the systems librarian indicated changes that impacted the TWIG template. The cataloger's document indicated the removal, the addition and standard changes to Theses (which they input into Arminda).
3. A "Catch-all" Spreadsheet was created (ARMINDA Crosswalk) to gather various field information together in prep for mapping
   a. "Crosswalk" Tab is the accumulation of information on fields (cells were populated with equations pulling from various spreadsheet tabs as to insure information was updated"
   b. "ALL" tab is the technical information on specific fields
   c. "IMIG" tab is MIG recommended mappings correlated with our fields
   d. "MODS Fields as of 3.3.20" tab are the list of top-level MODS fields we contained
   e. "Solr" tab is the full list of solr fields in Islandora
4. Because the collections were so different from one another, it became important to keep track of all the moving parts including tracking which various documents and when they were last updated. A spreadsheet to track document changes was created (ARMINDA documentation sync).
5. For the various fields in various collections, a spreadsheet was used to track fields in those collections, and to ensure they were in order (Arminda Fields - "Spreadsheet field checks" Tab) Spreadsheet uses equations pulling from Collection Master Spreadsheets to track field changes.

Metadata Content Remediation

Once the field structure was settled, those changes were made to the spreadsheets with the collection metadata, and in some cases combined, moved, or eliminated older fields. In addition to making changes to the fields, the team also edited the metadata for all of their objects in Islandora to ensure that it complied with the new standards.

This involved editing content into controlled vocabularies, enhancing descriptive information where needed, and populating new fields that they added in (i. e. genre and extent). For our descriptive work, they were guided by the "Archives for Black Lives in Philadelphia's Anti-Racist Description Resources." The team was able to make some of their metadata changes via batch functions in OpenRefine and Google Sheets, but many of the changes had to be done manually. Several student workers were trained to assist with this project.

RDP Mapping Documentation

The third attachment in the Trello card, the RDF mapping template (IMI_template_RDFMapping.docx), was used to compare various RDF mappings and discuss them as a group. The metadata librarian used Protege to make the crosswalk. To see the process in Protege, download Protege (https://protege. stanford.edu/) and open the ARMINDAMapping03-26-20 using Protege. At the time, the metadata librarian utilized all schema that seemed to work with RDF, including Bibframe, MADS and PREMIS, and it became messy very quickly.

A "Catch-all" Spreadsheet (ARMINDA Crosswalk) was used to continue building the mapping, comparing to MIG suggestions, resulting in two documents in addition to Protege and the Catch-all to create and remediate the RDF mapping.

- Whitman Islandora 8 Mapping

To better track the IMI to MODS to RDF process, HTML and Github were used to gain a different view, and to track the process, and to provide a public view of the mapping choices.