

2020-10-08 Whitman College Metadata Meeting

Attendees

- Alan Stanley
- Amy Blau
- Dana Bronson
- David Wilcox
- Kun Lin
- Noah Smith
- Paige Morfitt

Agenda

Topic
Goals and deliverables <ul style="list-style-type: none">1. What are our goals with regard to metadata remediation and mapping?2. What deliverables do we need to produce in order to accomplish these goals?
Requirements, blockers, and outstanding questions <ul style="list-style-type: none">1. What do we need in order to accomplish our goals?2. Are there any blockers that need to be resolved in order to proceed?3. Are there outstanding questions that still need to be resolved?
Timeline <ul style="list-style-type: none">1. What are the decision points and when do they need to happen?2. When are the deliverables due?
Wrap-up and next steps

Resources

- [Project Plan \(Draft\)](#)
- [ARMINDA Metadata Mappings](#) (draft)

Notes

Goals and deliverables

1. Whitman perspective
 - a. Metadata remediation
 - i. Bring use of fields and field names into conformance with new collections standard before moving into the new system
 1. In process now and almost complete
 - a. Complete in the sense of having identified which fields to remediate - changes have not yet been implemented
 - b. Working in spreadsheets
 - c. Use IMI remediation to make changes in Islandora 7
 - d. Planning on re-ingesting spreadsheets before migration
 2. Got rid of over 100 fields that were no longer needed
 3. Renamed fields to be clearer
 4. Only added a couple new fields
 - ii. Making sure all dates, names, etc. are standardized
 - iii. Looking at about a month to complete
 - b. Metadata mapping
 - i. All fields have a logical equivalent in RDF
 - ii. Any fields that still need to be removed could be removed during the mapping
 - iii. Mapping is about 90% done
 - c. Newspapers
 - i. Need to understand which fields will be required in a paged content model that might not already exist
 - ii. Sort by date
 - d. Working on TWIG templates
 - i. Translates flat CSV
2. Deliverables needed

- a. Alan: Would like to see spreadsheets detailing metadata to go into new system

Requirements, blockers, outstanding questions

1. Blockers
 - a. Repeating metadata fields are a problem in RDF
 - i. Some of this is being helped by remediation work
 - b. Staging server need to be updated
 - i. This is a potential blocker if things go poorly
2. Outstanding questions
 - a. Content models
 - i. Some photographs are not in the content model Whitman would like
 1. Basic vs. large image
 - ii. Content models are tied to file formats and viewers
 1. We can bring over the original files during the migration, create derivatives as appropriate, and associate the right viewer
 - iii. JPG2000
 1. Not a good native format for Windows machines
 2. JP2s are not a good archival format, kind of a black box
 3. TIFFs and JPGs are better for downloads as well
 4. Whitman has original TIFFs, which can be supplied along with a spreadsheet connecting them to PIDs
 - b. Entities
 - i. Adding URI-based entities in metadata would be useful to add sooner rather than later
 - ii. Need to work out fields where entities would be useful
 - iii. Need to schedule another meeting on entities

Timelines

1. Remediation: November
2. Spinning up new production server: January
3. Add metadata spreadsheets to Google Drive: October
4. Metadata and content mapping completion: December
5. Content freeze: January

Action Items

- Alan: Send static IP info to Kun for SSH access
- Amy: Add metadata spreadsheets to shared Google Drive folder