

## Arca Admins Meeting Notes

Date: Tuesday, July 19, 2022, 10-11:30 AM (Pacific)

Note-taker: Michel Castagné

discoverygarden (DG) has delivered a first draft of their Islandora 2 migration report. When complete, it will go to the Advisory Committee. Some topics we discussed:

### Metadata fields

- Metadata is more streamlined in Islandora 2, using Drupal fields instead of MODS XML.
- All Arca sites will need to use the same set of fields (i.e., we cannot add/modify fields ourselves), although we can suppress visibility of a field so it won't appear on our forms. There will be a single ingest form for all types of content.
- DGI has analyzed our MODS elements and some Arca sites have customizations that stand out from everyone else's. We need to review the less common MODS elements and clean up/standardize. If one of these less common elements is unique and important, we'll need to discuss further, such as how to map it.
- **ACTION:** Brandon will send the list of MODS elements currently being used, with questions and instructions. We will need to review and address less common elements to enable smooth migration.

### Metadata sharing

- Because metadata is in Drupal fields, we can use entity and taxonomy references for our metadata – that is, instead of entering text every time, we can look up existing taxonomies and objects and create references to those items. This allows for functional controlled vocabularies and linked data. We have three options for how to leverage this capability:
  1. Each site maintains its own local taxonomies and named entities. We would only be able to access and use own metadata; this could lead to inconsistency between sites.
  2. All taxonomies and named entities reside on a single central host. Each site makes references to the same set of metadata values. There is value in having shared metadata across institutions; downside is that if a site makes a mistake when adding new values, it becomes available to everyone, and everyone has access to sometimes irrelevant information
  3. Hybrid: some things are shared (e.g., genre terms), some things are local (e.g., names)
- Discussion: All in attendance preferred Option 2, as there is value in even named entities being shared between sites. Some things to think about:
  - Rules for differentiating different entities with similar names
  - How to manage named entities with Scholar attributes (see below)
  - Standards for adding new values
- The Advisory Committee will discuss further.

### Person object model

The "Person" object model will be gone in Islandora 2, replaced by a "Person" taxonomy. The objects will need to be converted by crosswalking the existing MADS metadata to new fields on the Person taxonomy. Connections to objects that reference the Persons will be preserved. To display a list of people, we can set up a Drupal view filtered on institutional affiliation or other criteria.

### Books and Newspapers

- The Internet Archive BookReader module will not be available in Islandora 2. Islandora community is discussing using Mirador (<https://projectmirador.org/>), but this is not implemented yet.
- In Islandora Legacy, the Book and Newspaper Issue content models generate individual TIFF files for each page from the PDFs that we ingest.
- PDFs can be ~90% smaller than complete Paged Content objects.
- **Question:** Are we OK with migrating only PDFs to Islandora 2, or do we want to keep these generated TIFF images?
  - Benefits: Save storage, reliably readable interface
  - Drawbacks: No individual Page objects; TIFF pages would be lost if you aren't keeping them elsewhere as backups.
- **Discussion:** Mostly leaning towards PDF-only. We would need to discuss interface details later.
- **ACTION (all):** Review your content and determine whether you would prefer to migrate the PDFs, or keep the Paged Content structure.

### Fedora

In Islandora 2, Fedora is optional. It does add costs, both labour and infrastructure. Is it useful? DGI recommends not using it and Brandon is inclined to agree. During discussion, Arca Admins are leaning towards no longer using it as well.

### Some features not available in Islandora 2 yet

Some in-house modules are not available in Islandora 2 yet, like Islandra Badges (Arca Office will plan to develop a replacement), Islandora Multi-Importer (replaced by new Spreadsheet Ingest), and Islandora WestVault (needs to be entirely rewritten). Access control will no longer be based on the XACML model. Alternative modules/implementations for access control are still under discussion, but will be user- and role-based, meaning it should be possible to migrate XACML policies. Islandora community is still discussing best approaches.

### Theming

- DGI has proposed using their base theme (compliant with WCAG 2 Level AA), with customizations that suit our needs. We would continue to be able to use our own colours, custom blocks, etc.
- **ACTION:**
  - Review the example sites using the DGI base theme:
    - <https://digitallibrary.ul.ie/index.php/>
    - <https://fsu.i8.dgicloud.com/>

- <https://discoveryvirginia.i8.dgicloud.com/>
  - <https://pc.i8.dgicloud.com/>
- Send questions/concerns/likes/dislikes about the base theme to Brandon, it will inform discussions with DGI.

### Custom datastreams (e.g., licences)

- Have you created custom datastreams; for instance, licences attached to objects?
- **ACTION:** Let Brandon know about this, including the specific DSIDs, how many you have, how you're using them, and whether they're important to migrate.

### Other concerns

Usage stats / metrics will likely reset after migration; will need to discuss further with DGI.

### New architecture

- Current architecture: one Drupal codebase, multiple subdirectories. New architecture: separate VM for each site with some shared components (e.g. taxonomies, some shared Solr info), with an aggregated search/browse feature performed by a separate Solr core on arcabc.ca. This parent site would be display-only and we'd no longer be viewing objects on arcabc.ca; instead it would point directly to the local organization's site.
- Briefly discussed the purpose of the parent site, including the possibility of putting together exhibitions, virtual collections etc. Significant interest in creating virtual exhibitions collecting objects from multiple member repositories.
- **ACTION:** All Admins will need to delete or publish inactive/orphaned objects to enable a smooth migration. Use the "Simple Workflow inactive objects" tool (under Islandora Utility Modules submenu) to view these objects.

### Research Data Management using Arca

- Michel led a short discussion on whether any sites are currently using or considering Arca for hosting research data. Join the #western-canada channel on Islandora Slack (<https://islandora.slack.com/>) to discuss further (including whether interested people would prefer a different forum).

### Other questions

Some questions were asked in the chat or after the meeting that were not addressed live These are answered here.

- Paige: Will the "published" function replace xacml / object policies?
  - No – "published/unpublished" replaces the Simple Workflow "active/inactive" status. This should be used for any objects that are being worked on and not yet ready for public consumption.
- Renee: How will MODS subelements be treated?
  - The same way as parent elements. There should be a Drupal field mapped for every MODS element.

- Richard: I see the metadata crosswalk spreadsheet on the agenda, is there another document for the DiscoveryGarden report where they flag the "problematic" MODS elements and/or content models for the migration or is that it?
  - Yes; this will be shared along with other information when ready.
- Bronwen: what about views that have been created for specific groups of objects (i.e a static page that uses a field to populate the page)?
  - Custom-made views will have to be recreated by local admins based on the new data.
- Gretchen: Will embargoes be migrated and how will we embargo in new form?
  - Embargoes are a form of access control; this is still being worked on.