Digital Archive Committee Minutes
December 17, 2019
1:00 PM

What is next for the archive discussion

- Brandon had asked the group to do a thought experiment at the end of the October meeting, but the group did not end up talking about this topic during the November meeting.
- Alysa asked to have this topic brought up again for the December meeting.
- The discussion was about what steps would the group take if Islandora were no longer available.
- Brandon let the group know that they could still use Islandora as long it runs on the current operating system.
- This discussion originally started in October when there was a question about formats and how the current content would be converted in the future.
- Alysa mentioned that this conversation could also be about what do you want to see next for the digital archives that our current system is not capable of handling.
- Alysa asked how do we incorporate a real magazine into the archives
  - When Bud did their magazine collection from the 80's it was one article after another, so it was easy to scan the magazine via article and describe each article and collate it into one magazine.
  - Bud has another magazine they would like to add. However, with this magazine, the article might start on page 10, and it leaps over to page 24, and page 55. Being able to accurately display at the article level those three pages is a big part of what they would want Islandora to do, as well as being able to browse the entire magazine in page order.
  - Alysa thinks that this might just be a field that needs to be added to the form (i.e. article continues on page …) and somehow Islandora compiles that article into those pages.
  - Jo mentioned that they have a huge backlog of magazines that they are ignoring for now, but would eventually love to be able to scan magazine issues and have them be in the correct order. This would mean sorted by page and the publication month.
  - Alysa mentioned that right now the way Islandora works is that it fully has the functionality to scan a magazine and it will display in direct order. What is missing is the ability to bring the article together and have the appropriate pages be collected under an article level.
  - Alysa pointed out that currently in Islandora each scanned page for a magazine is an object. There is an object for the magazine which is a compound object and each article is its own object as well. You end up describing the article and every article is different. We have to have the ability to describe each article level as well as the magazine issue itself.
  - Brandon asked if this would be a subcollection within a magazine collection. You could almost collect all the pages within an article into a subcollection that would display all the pages for an article.
  - Alysa commented on Brandon's observation to say that in this particular example that it is a very specific object to Islandora, a compound object, and so it does not use the subcollection concept, but that is essentially what is going on. This particular compound object requires us to upload the pages in the exact order you want them to be displayed, and that is not how a modern magazine works. To solve this problem would be great as far as the future of the archives.
- Brandon shared with the group that he, Pascal, and Ashley will be attending the Islandora Camp in February of 2020. They will learn more about the next iteration of Islandora. Brandon will make sure to look out for any updates on how articles will be added to Islandora.
- Alysa remembers that the current version of Islandora was set up to accommodate linked data, so she knows that the upgrade is a big deal.
- Pascal's understanding of the new official version of Islandora is that it is a major overhaul with the underlying components that have linked data as a primary element that will make this like a major migration to upgrade our Islandora.
- Here is the [demo video](#) that Brandon mentioned during the meeting and shared by email about the next iteration of Islandora.
- Alysa also mentioned that Bud Werner has some small born-digital projects. These were web pages on a website that no longer exists. The content was put on that site, and that is the only place that content lives.   They would like the ability to migrate that collection into the digital archives to preserve that work. A born-digital website is important to Bud Werner.
- Alysa mentioned that there is a content type for this type of collection. It was just not on the Marmot project development list because of the timing of things happening at Marmot. She wants to make sure this type of content stays at the forefront because she is not sure how long she will have access to the web content.
- Brandon made a note that he needs to look into the **Islandora Web ARChive Solution Pack**.

- John mentioned that they are actively updating and managing the genealogy and cemetery database which seems redundant because they have a superior digital access platform for those types of people entities. He wondered if there was a way to move the records from the genealogy and cemetery database into the digital archives. Having people entities created, all the obituary information, and vital record data that they have put into one collection, so they have a one-stop-shop for their digital collection.
    - Brando remembers some discussion regarding the Marmot genealogy database, and whether or not it was possible for us to migrate any of the information over.
    - John mentioned that they link them manually, but it is not complete. It is only for people who warrant entity creation who happen to have a genealogy record for them as well. It would be great if it was a single record and a single collection that his staff could manage and update on a monthly basis.
    - Brandon mentioned that Pascal may have more information about the genealogy database and whether or not it is possible for us to do something automated with those records in the future.
    - Pascal responded that he would love to migrate all the genealogy data to the archive. The genealogy search in Pika is working, but they are not doing any more development on it. Pascal can do more with this content in the archives. If someone can take time to map specific fields from Pika to what their equivalent would be in Islandora then they could build a script that would run through all the existing entries and create Islandora objects for them.
    - Ashley asked what this would mean to move all the genealogy records into the archive? She likes the idea of it being married together but wondered what others think about it.
    - John understands Ashley's concern about ingesting 158,000 people entities into the archive.
    - Alysa mentioned that genealogy was on the punch list the day they built the digital archives. She knows that other libraries (Eagle and Mesa County) built that genealogy database before they had a digital archive. Alysa suggested that we might want to have a conversation with those libraries to make sure they agree that this information should all live in one place.
    - Brandon asked how people would picture the genealogy records linking to existing entities. Would you picture that being additional fields within an existing entity?
    - Ashley asked if it would be its own object that links to an entity?
    - John commented that he sees it in the way it kind of works now. When you connect a genealogy or cemetery record to a person entity, a lot of the vital information (age, date of death and birth, etc) is automatically populated into the person entity. Alysa commented that the information is only in the display but it is not put into the entity record. The only thing that John would like to see displayed, that is not in a person entity when it is connected to a genealogy record, is the obituary image. Alysa mentioned that we could put the URL to the genealogy obituary record into the entity form.
    - Alysa mentioned they were trained initially that if the genealogy records exist for an entity that you are going to create that you do not have to copy all the data in that genealogy record into Islandora. All you have to do is put the link at the bottom of the page and in Pika it will display both sets of data. This does make things problematic because you cannot tell where the patron resides unless you click on the link. If we populate all the fields in Islandora than there is the one-stop-shop for birth, death, marriage, sibling relationship, obituaries, notes, and connections to objects. It does seem best to take the obituary database and burial database and move that into the archive. We are essentially maintaining two identical things in two different places.
    - Brandon asked which libraries are currently using the genealogy database and keeping it up-to-date and adding new items to it, and Bud Werner, Eagle Valley, and Mesa County replied. Brandon asked how many people who are not currently using the genealogy database would start using this functionality if it existed in Islandora, and Gunnison replied.
    - John asked what would it do from a storage perspective to migrate 158,000 people entity records. Would that change the expense associated with storing them, or are objects the only things that incur expenses as far as storing them in Islandora?
    - Brandon's understanding is that objects themselves are the primary factor to size, and since entities are shared, he does not think that Marmot charges individually for entities. Members own their objects, so they pay for their objects.
    - **Action Item:** Brandon will talk with JB to get a breakdown of the actual data usage points in relationship to cost.
    - Brandon thinks the true difficulty would be taking anything that currently exists in the genealogy database without doing any mapping to see if an entity already exists in the digital archive. Another thought would be to have a really large scale project with everyone who participates in the archives manually put things into the correct fields for the ones that are matching, or have the potential to match. We could potentially just have a bunch of duplicates where you have the primary entity with all the information that was added before the migration, but you have the genealogy obituary piece and how we link those two is an important question.

- Pascal commented that for automating a migration from the genealogy table to Islandora, they would make the migration tool be very cautious. If any name matches the migration tool would not move the content until a person could take a look at the information. If this is something that people want to do, what would be helpful for the automation is to take how each piece of information would relate to whatever the desired structure that would be in Islandora. He is guessing that this would all be inside a person entity rather than having to create a new object for any of this data. It would be nice for the committee to walk through a specification of what a migration should look like, so the experts on the data could figure out the right thing to do, rather than having Pascal making an incorrect assumption.
    - Brandon does have data mapping of genealogy records as a future agenda topic.

Entity fields and objects discussion

- Alysa shared her screen to show an object in Islandora (**Minute 26:27**)
- Alysa has a book object that has all the typical fields to say that this particular object has a related person or organization, a pictured entity, and a described entity. This is the way we take this object and link it to a particular entity. We have a lot of different types like a related place, or a related event. If the particular entity that you want to relate to an object does not fit one of the general categories, there is one last field called Generic Related Entity.
- Alysa discovered that when you put something in the Generic Related Entity field the particular object did not display on the entity page.
- Aysa asked why the Generic Related Entity field does not show up on an entity field? She does not know the history of that field and wanted to bring this up for discussion. For her, if she is going to make a relationship to an entity on an object it should not matter which field she uses. They should all automatically show up on that entity's page.
- Brandon asked to clarify if Alsya was talking about this object displaying in Pika?
- Alysa confirmed she was referring to displaying in Pika.
- Ashley wonders if that field may not be indexed for searching.
- Pascal does not recall having the Generic Related Entity not showing on an entity page as something the Pika team did on purpose.
- Alysa asked if this would be a ticket to have that field added to the Pika display. Pascal agreed that it should be a ticket.
- Brandon asked if we should look at all the fields from the Marmot Master Form to make sure they all display in Pika.
- Alysa suggested that we should go through all the available forms because they ran into this issue as well with pictures. They had pictures in the same collection that had a generic related field that did not display in Pika.
- **Action Item**: Brandon will submit a ticket to have all the available forms check to make sure the generic related entity field displays in Pika.
- **Action Item:** Pascal will look at the [Digital Archive Cleanup ticket](#) to see if any issues can be resolved during the Islandora Camp. This list will also be reviewed at the January meeting to come up with five items for the camp. You will need to [sign in](#) or wait for the meeting to see the information form the tickets shared at that time.

Pika search engine logic for archive content discussion (Minute 40:03)

- Alysa displayed the example [old town hot springs memory project](#). Item 6 on the list does not have any reference to any of the words used in the search. Items 9, 10, and 12 have the same issue as item 6. Alysa just wondered how these other results are listed under the original search.
- Alysa wondered if the search logic should be one of the topics at the Islandora Camp?
- Ashley suggested putting the ["old town hot springs memory project"](#) in quotes to help narrow the search down to the collection. The original search without quotes might have been looking for the terms "hot springs" or "old town" or whatever combination of the phrase used.
- Alysa mentioned that using quotations makes sense to library staff, but patrons would probably not know to use quotes when searching.
- There was a question about why the search indexes or search types (title, keyword, start of title …) are not used for archive searching.
- Pascal does not remember why the search types were disabled for the archives. He knows there are a keyword and a title search type already.
- Alysa would like searching added to the development list as we move forward. She was not sure if the search engine was from Islandora or if Marmot is indexing it and running it through something else.
- Pascal confirmed that Islandora has its own index, and things get indexed there. Within Pika there is a configuration of the search query.

- Pascal showed how they configured the keyword search query behind the scenes. When a keyword term is entered into the search box, it gets filtered to a query field. Pascal turns on his search debugging, it shows the actual query that is made from Islandora which is a quoted search. In Pika, the search is done by best match which is by relevancy. The search engine takes the query and then calculates the score for everything in the index and sorts by that score. You want a query that captures everything so something is not missing from search results. When a search brings up more than the search terms this kind of issue is just an interesting relevancy case. Pascal was curious to find out why the relevancy ranking would put non-related interviews in with the collection, but he does not have the answer to that yet.
- Pascal commented that one of the Pika stability projects is search improvements where he is hoping to play with relevancy ranking to adjust the numbers.
- Ashley asked if the best match for the archive is using a different sort logic than normal catalog results?
- Pascal replied that the catalog is its own search engine so it has a different formula.
- Pascal asked if the group thought the collection should be more prominent than pieces or objects of the collection? Ashley added that this would mean that top-level collections would show up above individual objects that might meet the search criteria. Pascal added that what he was thinking, if it is possible, that in the formula they would give a higher ranking to things that match the collection title, a collection object's MOD match would have more weight than a regular object so that it would tend to appear higher up in the order. It would not necessarily appear first because if you have different keywords that better match some other object than that object should be first. Pascal wanted to know if the group would be interested in him pursuing this type of formula for the search query?
- Brandon commented that in order for that type of search query to work an object would have to have the words used for the collection somewhere in the object title for it to be ranked at all. Whether there is higher relevancy all rests on the importance of making sure that information is in the title of each object within that collection. He also commented that most libraries do not add that information to the object title.
- Pascal commented that he understood that most collections do not use the title for the objects.
- Ashley asked how the ranking would affect the serendipitous discovery for the Explore More bar?
- Pascal mentioned that it would all be the same search, but with additional processing.
- Pascal's takeaway is that it does not happen that often that the collection and objects would have the same titles, so maybe the ranking is not something to fiddle with when it comes to the object title.
- Brando suggested that if anything could be done with the search results he thinks a grouping by the specific collection it sits in would be helpful.
- Ashley suggested that it might be helpful to have the parent collection come to the top of the ranking above the objects, more so than grouping.
- Alysa agreed with Ashley because the home page for the collection really features a lot of unique content that they want patrons to discover above finding a single object that does not seem to connect to anything.
- Ashley asked that if we could wave a wand would you like the top-level collection to show up first, or find something different to display at the top?
- John said with wand-waving it would be great to have the collection show up first.
- Pascal commented that it is probably pretty easy to give preference to a specific format. He thinks they could make an adjustment to give a format boost to the collections.

Next Meeting is January 21, 2019, at 1 p.m.