Digital Archive Committee Minutes
May 21, 2019
1:00 PM

1. **Linking your objects to existing entities discussion (Lead by Brandon)**
   - Clarification on how to best link any objects to existing entities, and also enriching existing entities with additional information.
   - This issue was brought up because Carol wanted to put a person is the Digital Archives and discovered a record that was already created for this person.
     - Carol was unsure about how to proceed, since you do not get the PID number until you create the record.
       - This would have created two PID numbers for the same person.
       - If it was a video, a library can link with the PID number.
   - When you discover that someone has already created an entity, what is the policy or best practice on enriching the existing entity?
     - What is the policy for adding a biography, or other metadata information?
     - Entities are shared, so that objects can be linked to them.
   - Alysa commented that if you are absolutely sure that it is the same individual, you should go ahead and add content to that individual, about that individual, and make notes of your source in the MODs record.
     - We should endeavor to enrich all of our entity records when possible.
     - Any additional information about the person can also be added to the Notes field.
     - This was one of the topics that was discussed in the Metadata Standards group.
     - Records should be enhanced, if it helps everyone make a good choice.
   - Carol has doing links to newspaper obituaries, and articles on the people, along with biographical information. Do you think that will fit in with what is being discussed?
     - Alysa agreed that what Carol has doing will definitely fit into what is being discussed about enriching existing entities.
   - Jill shared that when she adds biographical information to a person's record, she puts parentheses right after the information at the end to include the source like a pamphlet, brochure,or newspaper clipping.
     - A source example: Gunnison County Times article published on mm-dd- yyyy.
     - She would also include the PIDs in the related objects fields to link them.
     - An example of a person who was updated is "**Person 5771".**
   - Brandon mentioned that if you know the person you are adding in has a child and you have their PID number, you would add it to the person's entity record. The more fields that can be filled out the better.
   - Alysa mentioned that in the original metadata plan that the minimum there had to be a name.
     - It really helps if it is a name and a place.
     - The "Has Address" will have a very general place entity, so we can reuse them as much as possible.
     - An example of a person that Alysa updated is "**Person 13941".**
   - Brandon mentioned that how the source is written in each record can be discussed in the Metadata Group.
     - The group agreed to make sure everything is consistent.

2. **Demo of Pika Google Analytics tracking of archive object views (Lead by Chris)**
   - This will track how many times an object owned by a specific library is viewed.
     - This is still in the testing phase, so Chris wanted some feedback.
   - Chris updated the Google Analytics for the archive pages.
     - An archive page is anything that returns a specific archive object.
   - In the past, custom variables were used that are very difficult to manage.

- The new Google Analytics javascripts specifically offers a way to create either a custom metric, or a custom dimension.
  - Using the new analytics, Chris created a custom dimension that will allow Marmot to track how many times an object has been viewed from a specific owning library.
  - With custom dimensions everything falls in with all the other dimensions.
  - Chris create an Object Owner dimension that when selected, will show all the clicks for each object in the Google Analytic chart.
    - This creates a column that shows the object owning library information.
    - If entities are shared, the object owner column displays the word organization or person.
    - A filter in the Google Analytics can be used for each library to see only their Object Owner information.
    - The Object Owner can see how many total pageviews for their objects, as well as individual pageviews for each owned object.
  - These pageviews are recorded in the Marmot Digital Archive Google Analytic property account
  - Chris has set up the Marmot account to send pageviews to each library's Google Analytic account.
    - It is sending specific data over to the Marmot Islandora property account.
    - Each library will get all of the analytics for the archive pageviews in either the view they are using from Marmot, or their own Google Analytics account.
- The next step in creating better analytics for the archives is knowing what everyone would like to see.
  - These custom metrics can be assigned and more can be created as long as Marmot has access to the information.
  - Chris can get the pageview information into the Islandora account, so Brandon can share with individual libraries, or add in a Tableau report.
  - There are only a handful of entity types that might end up in the report that will show the generic organization and person as the Object Owner.
  - This would go live with the normal Pika release, so it will not be live immediately.
- Alysa mentioned that if someone is accessing a shared resource from their URL, she would count it as someone using their archive.
- Chris shared that all the entity and object information will be sent to your Google Analytics account, where you can filter it by URL to see how many hits you received, when patrons view it though your Pika site.
  - The ideas behind this is to see how many people view objects that you own across all the Pika archive team.
  - This idea came up because a report was needed to show archive use. Our archive is shared among Discovery Partners, and they are another source of views that can be collected.
  - This new Google Analytic property for Marmot Islandora allows us to collect from all those sites.
- Brandon thinks what would work best is to have the hits on the archive from every Pika instance.
  - For example: if you are Bud Werner, show me the hits for all the archive content regardless of who it comes from, show me what all the Bud Werner patrons are viewing. Also show me on a separate instance who is looking at Bud Werner's objects. Tracking entities in the larger search is not as important. He thinks entity tracking is more important on the individual instance. This would track what entities a library's patron is viewing. On a higher level, being able to see who is looking at a library's content. He would like to be able to track those stats, and share those stats with the members from both perspectives. These reports would show what a library's patron is viewing, and how much traffic all your content is getting currently. He think's Chris' new Google Analytic report will get those stats for archive members.

- Pascal wanted everyone to know that there will be a demo at the next Discovery Committee meeting on June 4th, for this new custom variable, because he needs to know who is still using any of the old custom variables.
  - If you realize there is another metric about archive objects you would like to track, please email pika-at-marmot.og, and they will try to get that done. Pascal would like to make this live by the next Pika release, so they can start tracking statistics for next year.

3. **DPLA feed improvements (Lead by Pascal)**
   a. Entities as DPLA subjects
      - Marmot requested to have object related entities added into the subject field for the DPLA feed.
      - Pascal finished this work and showed it during the meeting to get everyone's feedback.
      - Currently, only the actual subjects of your objects show up as subjects right now in the DPLA feed, plus the owning collection name.
      - Now, subjects have been expanded to include most of the related entities for the DPLA feed.
        - First displayed are the subjects
        - Next the owning library
        - Followed by related people who are not the publisher.
          - The publisher is excluded because is it another field in the DPLA feed.
        - Followed by related organizations which is included, unless it is the publisher.
        - Sponsors are included in the subject
          - Pascal asked if sponsors should be included?
        - Followed by related events as a subject
          - The related places will not be included, because there is a place field in the feed.
      - Brandon asked if this change would satisfy what Bud Werner requested about content added to the DPLA feed.
        - Alysa at first admitted this change would not work, because they way they chose to build their articles using the described entities field as opposed to place or organization.
        - Pascal and Alysa verified that the new feed will include the described entities field and display subjects that are necessary for her records.
        - Pika organizes from the namespace of the entity, the described fields are included.
      - Homework for everyone is to look at data in the feed on your library Pika test site (libraryname2.marmot.org) .
        - Look at your objects that you are really familiar with, and search them using the feed link from the DPLA Ingest process documentation.
        - Make sure to set your namespace.
      - This will be in place for the next ingest in late July.

   b. Hub ingest from a file for the Plains to Peak Collective (PPC) DPLA Feed
      - The process needs to change, because direct calls to the API gathers information page by page.
        - This is very labor intensive for the PPC people.
      - Pascal is building a process that for them with one giant JSON file for them to download.
      - Our hub will download this file directly and process it.
      - Pascal will have the file update a couple times a week, so it will be fresh for any ingest.

   c. Fix problem objects that would break feed
      - Couple objects were breaking the feed pages, so halting the DPLA process.
      - Ashley uncovered that these objects were duplicates.
        - Brandon cleared out the duplicates.

4. Metadata Sub group (Recording minute 42:17;26):
    a. Outline the needs for the best practices document
        i. Best practices for enriching entities
            1. Updating thumbnails images
        ii. Reduction of duplicates entities
        iii. Outline when to use an entity and when to use an object
        iv. Formatting of data entry (consistencies for patrons)
            1. Archive data optimization for DPLA and Pika (how data shows in each)
            2. Formatting dates (multiple date formats)
        v. Understanding the roles of entities
        vi. Subjects (Local or LC)
        vii. Creation of a metadata plan (pre-planning)
        viii. Best practices document for each entity type
            1. Person
            2. Place
            3. Organization
            4. Event

- Brandon will share the Google doc for the sub group.
    ○ The group can make the decision about how they want to organize the documentation
    ○ Brandon suggested that the group us the Google doc as the outline, and use the meeting time to start building the documentation.
    ○ Brandon asked if the group wanted to fill out the information as a group, or individually in their spare time?
        ■ Elizabeth thinks it's best to fill it out as a group.
        ■ Alysa agreed with Elizabeth.
- **Action Item**: The Metadata Sub group should think about the priority order for the Google doc list for the next meeting.

Next meeting is Tuesday, June 17th at 1:00 p.m.