2019-09-12 - Fedora Tech Meeting

Time/Place

This meeting is a hybrid teleconference and slack chat. Anyone is welcome to join...here's the info:

- Time: 11:00am Eastern Daylight Time US (UTC-4)
- Audio/Video Conference Link: https://lyrasis.zoom.us/my/fedora
  - Dial-in:
    +1 408 638 0968
    +1 646 876 9923
    +1 669 900 6833
    Meeting ID:
    812 835 3771
- Join fedora-project.slack.com on the "tech" channel

Attendees

Part 1:

1. Peter Winckles
2. Jared Whiklo
3. Bethany Seeger ⭐
4. Youn Noh
5. Thomas Bernhart
6. Daniel Lamb
7. Aaron Birkland
8. Ben Pennel
9. Steve Liu

Part 2:

- Cancelled for today

Agenda

1. OCFL and Fedora:
   a. inventory.json bloat and what to do about it.
   b. Is OCFL intended for a small number of versions?
   c. And if so, is that intention at odds with autoversioning in Fedora
   d. https://github.com/OCFL/spec/issues/367
   e. What about `squashing` versions?
   f. What about OCFL on S3 type filesystems?
2. Organizing Sprint work
   a. Review of Goals for Sprint 1
   b. Kick Off Meeting: Monday September 16 at 10am Eastern
   c. Major Areas of Work
      i. Design/Development
         1. Interface Definition
            a. Persistence API
               i. ?
            b. OCFL Client Development
               i. OCFL Java API
               ii. OCFL Java Client Implementation
            c. Transactions
      ii. Documentation
         1. Matrix of all the pages a la 5.x Documentation Updates
         2. Review of docs, flagging pages that will need to be changed, deleted, or added
      iii. Testing
         1. Performance Testing
      iv. Import/Export/Migration
         1. ?
3. Sprint Planning
   a. 6.0 Architecture Review
   b. Coming to consensus on:
   c. Transaction Sidecar Spec Update
4. Status
   a. API Test Suite PRs
      i. https://github.com/fcrepo/Fedora-API-Test-Suite/pulls
   b. Minimal 4 5 migration needs testing and code review:
      i. https://github.com/fcrepo4-exts/fcrepo-upgrade-utils/pull/17
5. Your topic here...

Tickets

1. In Review

type	key	summary	assignee	reporter	priority	status	resolution	created	updated	due

> Unable to locate Jira server for this macro. It may be due to Application Link configuration.

2. Please squash a bug!

key	summary	type	created	updated	due	assignee	reporter	priority	status	resolution

> Unable to locate Jira server for this macro. It may be due to Application Link configuration.

3. Tickets resolved this week:

key	summary	type	created	updated	due	assignee	reporter	priority	status	resolution

> Unable to locate Jira server for this macro. It may be due to Application Link configuration.

4. Tickets created this week:

key	summary	type	created	updated	due	assignee	reporter	priority	status	resolution

> Unable to locate Jira server for this macro. It may be due to Application Link configuration.

Notes

1. OCFL and Fedora:
2. From the OCFL meeting yesterday
   a. Potential of changing spec to allow mutable space
   b. Potential of changing spec to allow unversioned objects.
   c. Do they want to change the spec to accommodate the workflows around it (like Fedora's). We can offer use cases and experience.
   d. Are they willing lot look at in-motion objects and consider them more?
      i. If not, would we then rather have OCLF as a landing place and not the backend? Thinking about workflow and preservation - maybe we need to think about and manage both slightly differently.
   e. Anyone using OCFL - there was the group in Australia. Peter has done an implementation that's not in use yet.
   f. Maybe have 2 persistence stores – ie a db and OCFL.   Duplicate data space maybe, but possible to not.
   g. Overwriting V1 - How would we use the OCFL identifiers to find things from the fedora size. Moving is allowed in OCFL, but concerned about lots of moving of things.
      i. From Virginia Beach planning - fedora would be able to assign an id - mapping from fedora to OCFL.
      ii. Notion of someone creating an object - ie, archive group. Would want users to have the ability to give an ID to the object. It's not clear about what Fedora should do in terms of accessioning / deaccessioning of an OCFL object.
      iii. Confusion around no versions / versus versions discussion from yesterday's OCFL meeting. Seems like it'd have to be either one or the other for it to work successfully.
      iv. Is OCFL identifier part of the identity of the location w/i the object route? The identifier - it does make it easier if you can identify the root based on OCFL object path.   Still have a tombstone redirect on the OCFL objects in the case where OCFL path is

relevant to the object. There was something in the spec for creating object paths that need to be uniform / consistent. Can one do whatever they want for identifier path? https://ocfl.io/0.3/spec/#root-hierarchies

1. There are a few should there, but looks flexible.
2. Looks like one can do what we want with paths, but should try to be consistent.
3. Whatever fedora ends up doing it should be written down so a client can understand what fedora expects in terms of identifiers/paths.
4. Applies to any human read able format, not just OCFL.
5. Another option - id - > consistent location which may have any number of directories that are essentially versions of the objects. Then we pick the most recent version one.

v. What if we had a different OCLF layout – ie, we have a directory layout where we take you to a directory with OCFL objects. Ie, we keep objects and can keep multiple copies of the objects, so one move the object and make sure we don't lose anything.
1. Jared's attempt at a write-up. https://docs.google.com/document/d /1vWGPgp61IiRqEfUnDmlS_aEeAkGo_LSM4BIHuSO1k-4/edit

vi. Aaron Birkland talked about the design we discussed at Virginia Beach. About how the paths in fedora would relate to paths in OCFL.

vii. Was there a possible path for doing the layout in the way for reading directly from disk?

viii. Fedora will have an index of fedora uri OCFL uri, but it would be good for other clients to have a way to know to walk the data.
1. LDP path will be in the fedora data, as well as the hierarchy (children of containers, etc). The LDP hierarchy would be something fedora would construct based on the metadata. The fedora index would contain this data (after reading the data). Children pointing to parent, we don't have to worry about the parent being versioned every time a child is added. (if they are separate OCFL objects, this works). Is the child pointing to the OCFL parent or the Fedora path for parent. Is it the logical (LDP path) or physical (OCFL)?
2. In other words, it may not be easy, but it will be possible to recreate the information of what the layout is by not containing the physical layout w/i the object.
3. Modeshape limited us by pointing from parent to child (creating many members problem).   We will have a flat structure in OCFL.

ix. Squashing for every un-versioned changed might be challenging - it's really a copy and add new changes to object.   Hopefully there will be another option to do this.

x. There are some communities that will never use fedora versions

xi. Having a HEAD directory could provide a space that's mutable and we can keep stuff there, edit it, and then version it at some point. Might be easily ignorable by clients that don't want to know about it. Would a HEAD always exist? Advantage of having this there is for rebuild-ability.

xii. Having it outside the structure might have it's advantages as well.

xiii. A use case - for those that have warm storage where they can't delete anything, OCFL is nice and potentially very helpful.

h. Jared created a PR to start the flesh out the persistence API. - https://github.com/fcrepo4/fcrepo4/pull/1542

3. Organizing Sprint work

Actions

- Aaron Birkland to look explore notion of OCFL client with database as authoritative metadata source + asynchronous writing of the inventory.json file
- Peter Eichman   and maybe Ben Pennell to make recommendations re transaction side car specification.
- Andrew Woods will look into java 11 transition
- David Wilcox will review the NDSA matrix and pull out the concrete technical requirements that could be considered during the Fedora 6 development.
- Jared Whiklo will try to do some work on the PersistentStorage Interface. - https://github.com/fcrepo4/fcrepo4/pull/1542