

# 2019-09-05 - 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](https://fedora-project.slack.com) on the "tech" channel

## Attendees

Part 1:

1. [Danny Bernstein](#)
2. [Peter Winckles](#) (out)
3. [Andrew Woods](#) (out)
4. [David Wilcox](#)
5. [Peter Eichman](#) (out)
6. [Joshua Westgard](#) (out)
7. [Jared Whiklo](#) ★
8. [Bethany Seeger](#)
9. [Youn Noh](#)
10. [Thomas Bernhart](#)
11. [Ben Cail](#)
12. [Rosie Le Faive](#)
13. [Daniel Lamb](#)
14. [Aaron Birkland](#)
15. [Ben Pennell](#)

Part 2:

1. [Danny Bernstein](#) ★
2. [Peter Winckles](#) (out)
3. [Andrew Woods](#) (out)
4. [David Wilcox](#)
5. [Jared Whiklo](#)
6. [Bethany Seeger](#)
7. [Ben Cail](#)
8. [Aaron Birkland](#)
9. [Ben Pennell](#)

## Agenda

1. Announcements
  - a. Andrew, Danny and David will be out next week - volunteer to facilitate the this meeting?
2. OCFL and Fedora: inventory.json bloat and what to do about it. Is OCFL intended for a small number of versions? And if so, is that intention at odds with autoversioning in Fedora
3. Status on organizing a Fedora documentation review
4. Applying a digital preservation framework (e.g. [NDSA Levels of Digital Preservation](#)) to Fedora 6
5. Organizing Sprint work
  - a. Review of [Goals for Sprint 1](#)
  - b. [Kick Off Meeting: Monday September 16 at 10am Eastern](#)
  - c. Tentative plan for who will focus on what:
    - i. [Danny Bernstein](#)
    - ii. [Ben Pennell](#)
    - iii. [Peter Eichman](#)
    - iv. [Jared Whiklo](#)
    - v. [Bethany Seeger](#)
    - vi. [Aaron Birkland](#)
    - vii. [Anna Dabrowski](#)
    - viii. [Youn Noh](#)
    - ix. [Dan Field](#)
    - x. [Jenny A'Brook](#)
    - xi. [Mohamed Mohideen Abdul Rasheed](#)
    - xii. [Richard Williams](#)
    - xiii. [Michal Dulinski](#)

- xiv. [Remigiusz Malessa](#)
- d. 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. ?
- 6. [Sprint Planning](#)
  - a. [6.0 Architecture Review](#)
  - b. Coming to consensus on:
  - c. [Transaction Sidecar Spec](#) Update
- 7. 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>
- 8. 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

### Part I

1. [Jared Whiklo](#) will facilitate next week's meeting : [Bethany Seeger](#) will take notes.
2. Sprint kickoff meeting for 10:00ET
3. Documentation review: Call for people who are interested in participating. Have heard back from 4 people so far. Now to connect and look to organize the sprint. Anyone who is interested can connect with [David Wilcox](#).
4. This first pass is just to review the documentation with an eye towards the usability, correctness and understand-ability of the docs. A second pass to align the software with Fedora 6 can occur later.
5. Could we align Fedora with one or more preservation standards. These standards require a whole view towards preservation, but the software usually needs to do certain things and is the technical team considering these standards.
6. OCFL filesystem bloat:
  - a. It was known that in the case of a largish number of versions and a largeish number of OCFL objects. The inventory.json can become quite larger. Because there is a mapping between logical files and physical file paths. This must be performed during each version.
  - b. Writing/parsing is not a huge issue and is mitigated by caching of the parsed inventory.
  - c. The issue is the huge amount filesystem storage required by the inventory.json. There is a SHOULD in the spec which suggests that you store a version of the inventory.json in the version directory so end up storing several copies of these larger inventory files.
  - d. If Fedora creates an OCFL version for each change, this could cause quite large file storage requirements.
  - e. Things we could do to reduce/mitigate this file storage issue.
    - i. Not use SHA-512 (though recommended) and use SHA-256 instead for smaller hashes.
    - ii. Don't store additional inventory.json files in each version. This is seemingly more of a soft-requirement for the purposes of forensics in the case of a file-system error.
  - f. Are there any concrete actions we need to take or blockers for the upcoming sprint? None identified.
  - g. Should bring these issues to the OCFL community (via Slack or meeting) to ensure they are aware of our concerns. [Ben Cail](#) will take the previous issue to the OCFL community for discussion amongst that group.
7. [Jared Whiklo](#) feels that if the OCFL library is not ready to start, should we be working towards a simple DB + filesystem backend to get the core Fedora ITs working and spend time on those things.
8. Getting the PersistentStorage API fleshed out might help with some of the OCFL vs simple filesystem backend questions.

### Part II

Versioning: further discussion on [https://docs.google.com/document/d/1HaSFFesWcp\\_iThl-7\\_Whd3YCLiIDquBG1BcYyp9JtU/edit?ts=5d6834ce](https://docs.google.com/document/d/1HaSFFesWcp_iThl-7_Whd3YCLiIDquBG1BcYyp9JtU/edit?ts=5d6834ce)

Question 1: where to put the version tag file within OCFL?

1. One option is to put the version tag file in the OCFL Root:

lots of tag files in ocfl root. or large file - bad option

2. Another option : tag file lives in the Object's content directory. The tag file is updated before commit to OCFL is made. POSTing to an LDPCv will result in a new OCFL version since the tag file will need to be updated and pushed to OCFL. There seems to be consensus on this approach.

Question 2:

What is the format of the tag file?

- 1) list of ocfl version numbers
- 2) list of timestamps
- 3) a list of tuples with timestamp and OCFL version number

Discussion:

Re Option 1 [Aaron Birkland](#) raised the question fo whether having OCFL specific information in the tag file couples Fedora to OCFL in a way that is undesirable.

Option 2 is better, but Fedora will need a policy for resolving dates against versions since mutiple OCFL versions can have the same timestamp since the OCFL version timestamp, like a Memento's timestamp, offers 1 second precision.

[Ben Cail](#) suggested Option 3 as an option so that Fedora's handling of the versions tag file can be made more consistent across different backends file supporting more precise mapping afforded by OCFL.

## 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.