

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 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

T	Key	Summary	Assignee	Reporter	P	Status	Resolution	Created	Updated	Due
	FCREP O-3268	Rebuild does not succeed for Binary resources	Jared Whiklo	Andrew Woods		IN REVIEW	Ready for Test	Apr 02, 2020	Apr 04, 2020	
	FCREP O-3254	Replace String resource IDs with new FederalID class	Jared Whiklo	Andrew Woods		IN REVIEW	Ready for Test	Mar 19, 2020	Apr 04, 2020	
	FCREP O-3246	Cannot create binary resources with content types claimed by RDF sources	Jared Whiklo	Peter Winckles		IN REVIEW	Ready for Test	Mar 10, 2020	Apr 02, 2020	
	FCREP O-3035	Add RDFSource updates to OCFL impl	Andrew Woods	Andrew Woods		IN REVIEW	Ready for Test	Aug 01, 2019	Mar 02, 2020	
	FCREP O-3020	Fix Direct and Indirect Container Import	Danny Bernstein	Danny Bernstein		IN REVIEW	Fixed	Jul 08, 2019	Jul 12, 2019	
	FCREP O-3017	Align External Content Handling in import export tool for 5.x roundtripping.	Danny Bernstein	Danny Bernstein		IN REVIEW	Fixed	Jun 27, 2019	Jul 12, 2019	
	FCREP O-2740	Float value not updated when current value rounds to the new one	Yinlin Chen	Kevin Ford		IN REVIEW	Fixed	Mar 27, 2018	May 30, 2018	

7 issues

2. Please squash a bug!

Key	Summary	T	Created	Updated	Due	Assignee	Reporter	P	Status	Resolution
FCREP O-3269	NonRdfSource descriptions are missing triples		Apr 02, 2020	Apr 02, 2020		Unassigned	Jared Whiklo		OPEN	Unresolved
FCREP O-3261	404 should not include low-level response		Mar 26, 2020	Mar 26, 2020		Unassigned	Andrew Woods		OPEN	Unresolved
FCREP O-3257	PUT on existing container should be 409		Mar 20, 2020	Mar 26, 2020		Unassigned	Andrew Woods		OPEN	Unresolved
FCREP O-3256	Ghost nodes should be immutable		Mar 20, 2020	Apr 02, 2020		Unassigned	Andrew Woods		OPEN	Unresolved
FCREP O-3252	Disallow creating mementos within the same second		Mar 17, 2020	Mar 19, 2020		Unassigned	Peter Winckles		OPEN	Unresolved
FCREP O-3244	Response triples should reference URIs and not internal IDs		Mar 09, 2020	Mar 09, 2020		Unassigned	Peter Winckles		OPEN	Unresolved
FCREP O-3243	Fix concurrency issue when creating persistence sessions		Mar 09, 2020	Mar 09, 2020		Unassigned	Peter Winckles		OPEN	Unresolved
FCREP O-3238	Address naming collisions for .fcrepo files		Feb 20, 2020	Feb 21, 2020		Unassigned	Peter Winckles		OPEN	Unresolved
FCREP O-3234	PaxExam update needed to support Maven's HTTPS requirement		Feb 20, 2020	Feb 20, 2020		Unassigned	Jared Whiklo		OPEN	Unresolved
FCREP O-3145	Fix migration-utils inline XML migration to not transform XML en route		Nov 25, 2019	Mar 02, 2020		Andrew Woods	Scott Prater		OPEN	Unresolved
FCREP O-3140	Implement disambiguation pattern for OCFL file naming collisions		Nov 21, 2019	Mar 20, 2020		Unassigned	Jared Whiklo		OPEN	Unresolved

FCREP O-3139	Design a disambiguation pattern for the OCFL file naming		Nov 21, 2019	Apr 03, 2020	Unassigned	Jared Whiklo		OPEN	Unresolved
FCREP O-3138	Test HTTP -> Service layer		Nov 21, 2019	Nov 22, 2019	Unassigned	Jared Whiklo		RECEIVED	Unresolved
FCREP O-3114	Unauthorized users can delete child resources		Nov 07, 2019	Nov 07, 2019	Unassigned	Danny Bernstein		RECEIVED	Unresolved
FCREP O-3070	Creating external binary with bad URI and digest blows up		Sep 24, 2019	Sep 24, 2019	Unassigned	Ben Pennell		OPEN	Unresolved
FCREP O-3037	Prefer header 'omit' containing 'PreferMinimalContainer' URL not honored		Sep 06, 2019	Nov 22, 2019	Unassigned	Bethany Seeger		OPEN	Unresolved
FCREP O-3022	Intermittent build failure		Jul 15, 2019	Jul 16, 2019	Unassigned	Aaron Birkland		OPEN	Unresolved
FCREP O-3005	GET and PUT on resource/fcr: metadata/fcr: acl and resource/fcr: versions/fcr: acl return 400		Jun 06, 2019	Jun 06, 2019	Danny Bernstein	Danny Bernstein		OPEN	Unresolved
FCREP O-3003	JSON-LD representation of rdf:Lists is incorrect		Jun 03, 2019	Jun 03, 2019	Unassigned	Jared Whiklo		OPEN	Unresolved
FCREP O-2998	Potential issue with namespace registry		May 22, 2019	Jun 21, 2019	Bethany Seeger	Anna Dabrowski		OPEN	Unresolved

Showing 20 out of 66 issues

3. Tickets resolved this week:

Key	Summary	T	Created	Updated	Due	Assignee	Reporter	P	Status	Resolution
FCREP O-3268	Rebuild does not succeed for Binary resources		Apr 02, 2020	Apr 04, 2020		Jared Whiklo	Andrew Woods		IN REVIEW	Ready for Test
FCREP O-3263	NonRdfSource descriptions are missing headers		Apr 01, 2020	Apr 02, 2020		Jared Whiklo	Jared Whiklo		CLOSED	Fixed
FCREP O-3259	User provided RDF does not get the correct subject		Mar 26, 2020	Mar 31, 2020		Andrew Woods	Jared Whiklo		CLOSED	Fixed

FCREP O-3254	Replace String resource IDs with new FedoraID class		Mar 19, 2020	Apr 04, 2020	Jared Whiklo	Andrew Woods		IN REVIEW	Ready for Test
FCREP O-3246	Cannot create binary resources with content types claimed by RDF sources		Mar 10, 2020	Apr 02, 2020	Jared Whiklo	Peter Winckles		IN REVIEW	Ready for Test

5 issues

4. Tickets created this week:

Key	Summary	T	Created	Updated	Due	Assignee	Reporter	P	Status	Resolution
FCREP O-3269	NonRdfSource descriptions are missing triples		Apr 02, 2020	Apr 02, 2020		Unassigned	Jared Whiklo		OPEN	Unresolved
FCREP O-3268	Rebuild does not succeed for Binary resources		Apr 02, 2020	Apr 04, 2020		Jared Whiklo	Andrew Woods		IN REVIEW	Ready for Test
FCREP O-3267	Update SMT documentation for fcrepo6		Apr 02, 2020	Apr 02, 2020		Ben Pennell	Ben Pennell		OPEN	Unresolved
FCREP O-3266	Implement Transaction Expiration		Apr 01, 2020	Apr 01, 2020		Unassigned	Danny Bernstein		OPEN	Unresolved
FCREP O-3265	Implement cancel transaction on transaction endpoint		Apr 01, 2020	Apr 01, 2020		Unassigned	Danny Bernstein		OPEN	Unresolved
FCREP O-3264	Implement Start and Commit of multi-request spanning transactions		Apr 01, 2020	Apr 06, 2020		Ben Pennell	Danny Bernstein		IN PROGRESS	Unresolved
FCREP O-3263	NonRdfSource descriptions are missing headers		Apr 01, 2020	Apr 02, 2020		Jared Whiklo	Jared Whiklo		CLOSED	Fixed

7 issues

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_jThI-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.