

2017-08-17 - Fedora Tech Meeting

Time/Place

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

- Time: 11:00am Eastern Daylight Time US (UTC-4)
- Dial-in Number: (712) 775-7035
 - Participant Code: 479307#
 - International numbers: [Conference Call Information](#)
 - Web Access: <https://www.freeconferencecallhd.com/wp-content/themes/responsive/flashphone/flash-phone.php>
- IRC:
 - [Join the #fcrepo chat room via Freenode Web IRC](#) (enter a unique nick)
 - Or point your IRC client to #fcrepo on irc.freenode.net

Attendees

- [Danny Bernstein](#)
- [Carrick Rogers](#)
- [Ben Pennell](#)
- [Esmé Cowles](#)
- [Andrew Woods](#)
- [Kevin S. Clarke](#) ★
- [Nick Ruest](#)
- [Ilya Kreymer](#)
- [Yinlin Chen](#)
- [Daniel Lamb](#)
- [Aaron Birkland](#)

Agenda

1. Any pending issues from the last two weeks?
 2. Webrecorder Integration with Fedora:
 - Current proof-of-concept: <http://fedora.webrecorder.net/>
 - Webrecorder writing WARCs, reading from Fedora (no data model, just flat list so far)
 - Using Fedora's HTTP range request support
- Goals:
- Create PCDM data model for web archives
 - Store WARCs, as well as other web archiving objects created by Webrecorder
3. [4.7.4 release](#)
 - a. 4.7 LTS?
 - b. Next release: 5.0.0
 - c. Migration approaches from 4.7 to 5.0
 4. [Performance lessons from PREMIS events \(Ben Pennell\)](#)
 5. Volunteer for next week's tech meeting (8/24)?
 6. ...
 7. Status of "in-flight" tickets

Ticket Summaries

1. Please squash a bug!

Key	Summary	T	Created	Updated	Due	Assignee	Reporter	P	Status	Resolution
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 'PreferMinima Container' URL not honored		Sep 06, 2019	Sep 06, 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
FCREP O-2988	Creating versions of containers containing tombstones issue		Feb 08, 2019	Feb 13, 2019	Unassigned	Ben Pennell		RECEIVED	Unresolved
FCREP O-2841	Vulnerability on the commons-httpclient-3.1 dependency of fcrepo-integration-rdf		Aug 06, 2018	Aug 06, 2018	Unassigned	Peter Eichman		RECEIVED	Unresolved
FCREP O-2800	Issues running Fedora multi-threaded		May 14, 2018	May 14, 2018	Unassigned	Harsha Ummerpillai		RECEIVED	Unresolved
FCREP O-2790	Using a hash URI as a ldp:membershipResource doesn't work		May 04, 2018	May 04, 2018	Unassigned	Esmé Cowles		OPEN	Unresolved
FCREP O-2780	GET inbound refs incomplete response		Apr 26, 2018	Apr 30, 2018	Unassigned	Kevin Ford		OPEN	Unresolved
FCREP O-2746	Failed to reset the input stream when not accessing s3 directly or s3 compatible gateway		Apr 06, 2018	Apr 06, 2018	Unassigned	Thomas Schneider		OPEN	Unresolved

FCREP O-2739	Server returns an error after inserting a particular triple syntax		Mar 17, 2018	Mar 19, 2018	Unassigned	Stefano Cossu		OPEN	Unresolved
FCREP O-2712	GET with Prefer=InboundRefs on /fcr:metadata does not include InboundRefs		Mar 09, 2018	Mar 09, 2018	Unassigned	Kevin Ford		RECEIVED	Unresolved
FCREP O-2670	baseURI incorrect for audit JMS messages		Feb 02, 2018	Feb 05, 2018	Unassigned	Bethany Seeger		OPEN	Unresolved
FCREP O-2666	HTML UI does not update properly after binary resource reversion		Jan 23, 2018	Feb 01, 2018	Unassigned	Danny Bernstein		OPEN	Unresolved
FCREP O-2665	Integration tests flap on non-virtualized windows 10 machines.		Jan 23, 2018	Jan 24, 2018	Unassigned	Danny Bernstein		OPEN	Unresolved
FCREP O-2663	MOVE - Cannot delete Solr index		Jan 16, 2018	Jan 17, 2018	Unassigned	Ferdinando Terranova		OPEN	Unresolved
FCREP O-2662	Sparql Delete works inconsistently in transaction		Jan 03, 2018	Jan 04, 2018	Unassigned	Ben Pennell		OPEN	Unresolved

Showing 20 out of 53 issues

2. Tickets resolved this week:

Key	Summary	T	Created	Updated	Due	Assignee	Reporter	P	Status	Resolution
FCREP O-3121	Migration-utils: Remove GoLang OCFL client		Nov 11, 2019	Nov 12, 2019		Andrew Woods	Andrew Woods		CLOSED	Fixed
FCREP O-3118	Migration-utils: Set RDF format as n-triples		Nov 11, 2019	Nov 11, 2019		Andrew Woods	Andrew Woods		IN REVIEW	Ready for Test
FCREP O-3116	Migration-utils: Add ability to run the tool from user-provided pid-list		Nov 08, 2019	Nov 11, 2019		Unassigned	Andrew Woods		CLOSED	Duplicate
FCREP O-3113	Migration-utils: remove unnecessary OCFL version		Nov 07, 2019	Nov 08, 2019		Andrew Woods	Ben Cail		CLOSED	Answered

FCREP O-3112	Migration-utils: Migrate from a list of PIDs		Nov 07, 2019	Nov 11, 2019	Andrew Woods	Andrew Woods		IN REVIEW	Ready for Test
FCREP O-3111	Migration-utils: Enable resume /restart of migration		Nov 07, 2019	Nov 11, 2019	Andrew Woods	Andrew Woods		IN REVIEW	Ready for Test
FCREP O-3109	Advanced command line options for migration-utils		Nov 07, 2019	Nov 07, 2019	Remigiusz Malessa	Remigiusz Malessa		CLOSED	Duplicate
FCREP O-3108	Add support for snapshots to the parent pom.		Nov 06, 2019	Nov 07, 2019	Danny Bernstein	Danny Bernstein		CLOSED	Fixed
FCREP O-3103	Replace a binary resource (HTTP->PersistentStorageSession)		Nov 04, 2019	Nov 12, 2019	Mohamed Mohideen Abdul Rasheed	Danny Bernstein		IN REVIEW	Fixed
FCREP O-3095	Implement OCFLObjectSession		Nov 01, 2019	Nov 12, 2019	Ben Pennell	Danny Bernstein		CLOSED	Fixed
FCREP O-3094	Implement OCFLPersistentStorageSession.persist (CreateRDFSourceOperation)		Nov 01, 2019	Nov 12, 2019	Danny Bernstein	Danny Bernstein		CLOSED	Fixed
FCREP O-3081	Minimally implement OCFLPersistentStorageSession		Sep 25, 2019	Nov 08, 2019	Danny Bernstein	Danny Bernstein		CLOSED	Fixed

12 issues

3. Tickets created this week:

Key	Summary	T	Created	Updated	Due	Assignee	Reporter	P	Status	Resolution
FCREP O-3123	Perform locking on ocfl objects during commit		Nov 12, 2019	Nov 12, 2019		Unassigned	Ben Pennell		RECEIVED	Unresolved
FCREP O-3122	Create constant logger throughout project		Nov 12, 2019	Nov 12, 2019		Unassigned	Bethany Seeger		RECEIVED	Unresolved
FCREP O-3121	Migration-utils: Remove Go-Lang OCFL client		Nov 11, 2019	Nov 12, 2019		Andrew Woods	Andrew Woods		CLOSED	Fixed
FCREP O-3120	Migration-utils: Add fedora: hasParent relationship to datastream metadata		Nov 11, 2019	Nov 11, 2019		Unassigned	Andrew Woods		OPEN	Unresolved

FCREP O-3119	Implement OCFLPersistentStorageSession. getTriples() and OCFLPersistentStorageSession. getManagedProperties()		Nov 11, 2019	Nov 12, 2019	Danny Bernstein	Danny Bernstein		IN PROGRESS	Unresolved
FCREP O-3118	Migration-utils: Set RDF format as n-triples		Nov 11, 2019	Nov 11, 2019	Andrew Woods	Andrew Woods		IN REVIEW	Ready for Test
FCREP O-3117	Making staging directory configurable in OCFLPersistentStorageSession		Nov 08, 2019	Nov 08, 2019	Unassigned	Danny Bernstein		RECEIVED	Unresolved
FCREP O-3116	Migration-utils: Add ability to run the tool from user-provided pid-list		Nov 08, 2019	Nov 11, 2019	Unassigned	Andrew Woods		CLOSED	Duplicate
FCREP O-3115	Unauthorized users may not delete child resources		Nov 07, 2019	Nov 07, 2019	Unassigned	Danny Bernstein		RECEIVED	Unresolved
FCREP O-3114	Unauthorized users can delete child resources		Nov 07, 2019	Nov 07, 2019	Unassigned	Danny Bernstein		RECEIVED	Unresolved
FCREP O-3113	Migration-utils: remove unnecessary OCFL version		Nov 07, 2019	Nov 08, 2019	Andrew Woods	Ben Cail		CLOSED	Answered
FCREP O-3112	Migration-utils: Migrate from a list of PIDs		Nov 07, 2019	Nov 11, 2019	Andrew Woods	Andrew Woods		IN REVIEW	Ready for Test
FCREP O-3111	Migration-utils: Enable resume /restart of migration		Nov 07, 2019	Nov 11, 2019	Andrew Woods	Andrew Woods		IN REVIEW	Ready for Test
FCREP O-3110	Migration-utils: Expose option for storage layout		Nov 07, 2019	Nov 11, 2019	Andrew Woods	Andrew Woods		IN PROGRESS	Unresolved
FCREP O-3109	Advanced command line options for migration-utils		Nov 07, 2019	Nov 07, 2019	Remigiusz Malessa	Remigiusz Malessa		CLOSED	Duplicate
FCREP O-3108	Add support for snapshots to the parent pom.		Nov 06, 2019	Nov 07, 2019	Danny Bernstein	Danny Bernstein		CLOSED	Fixed



17 issues

Minutes

1. Any pending issues from the last two weeks?

There is a PR for the import/export tool.

2. Webrecorder Integration with Fedora

Webrecorder project (Ilya and others) have been looking into having Fedora as a backend. Webrecorder is an interactive web archiving tool. Anyone can use it to record sites. They'd like to add a preservation backend. One of the things they'd like to do in the future is to have a standardized way to preserve web recordings. This is an area that's currently lacking as far as web archives are concerned. The current prototype, linked in the schedule, was a very quick weekend project, intended as a proof of concept (but, as Andrew noted, works beautifully). There is currently no tool that provides preservation and access. This integration that they're proposing addresses that.

The Webrecorder folks are interested in a discussion on the data model for web archives. The PCDM discussion group (pcdm@googlegroups.com) would be the best place to ask these questions. They're interested in these sorts of discussions and that would be the best way to move the conversation forward. The Webrecorder folks have started brainstorming about the data modeling in a Google doc: <https://docs.google.com/document/d/1RiZnX4g3u1ydwX9odu5Y1s2ajquhIkqYema5TZp5UOQ/edit?ts=596fe0a4>.

Web archives are large objects so they're interested in learning how well Fedora handles this type of material. Andrew reports Fedora (Esme) has tested up to a TB file and that the tests have been successful.

State of S3 backend storage and clustering? Clustering, as far as Andrew is aware, has not been exercised very much (or at all). It's sort of a feature Fedora gets from Modeshape for free. There have been issues at the Modeshape level that have driven their recent work. So there is clustering in Fedora, but there are sprinkled caveats all around it. As for S3, with the most recent release, there is official support for S3 as a backend. Danny Bernstein has done some testing of the performance of S3 as a backend. Its performance is more or less in line with a local installation. S3 may not be deployed anywhere in production since it's really new. It is going into the Hyku deployment though so it will be pushed on more. Several Samvera people report that there is also S3 integration at the Samvera level (though this is different from Fedora's integration).

S3 support would be a higher priority for the Webrecorder groups because currently they store everything on S3. Fedora's S3 support is really undocumented at this point so maybe the Webrecorder folks working through this might be a good way to get some documentation. Or maybe some back and forth between Fedora and Webrecorder folks would be a good way to generate some documentation.

Related: there is work going on around specifying the formal API of Fedora, which will probably be slightly different from the one Webrecorder is currently using. Just as a note.

3. 4.7.4 release

4.7.4 release is out now. Fedora will be targeting a 5.0 release next that will have some breaking changes as the Fedora specification is finalized. The idea of Fedora having a long term support (LTS) version that the Fedora community would support for a period of years was discussed. This would mean that patches would continue to be applied to 4.7.x. Fedora 5 is only notional thing at this point and it will be quite some time before people migrate to it. Discussion agreed that having an LTS release is a good idea. What types of fixes can folks expect? Security fixes are definitely in. The group ought to articulate what will and won't be done. Another example is the project's dependencies (Java, itself, and otherwise); should underlying versions be upgraded over time? Java, definitely. Maven dependencies might be upgraded on a case by case decision(?)

4. Performance lessons from PREMIS events (Ben Pennell)

Ben has tested different ways of storing PREMIS events (objects vs. RDF logs (serialized RDF in a binary)) There are some graphs in the Google Groups message. Storing events as objects, of course, results in more objects in the repository and Ben did find performance implications for this. At around 50k objects, the performance was getting significantly slower for creating events as resources/objects. Ben and his group's conclusion was that it didn't seem like it would be a good idea to keep events as objects in Fedora. As an alternative, for any object in the repository, there would be an RDF log where PREMIS events would be stored.

5. Volunteer for next week's tech meeting (8/24)?

Someone willing to host next week's call? Andrew will be at a Fedora users group meeting in Texas. Aaron volunteered.