2015-10-29 - Fedora Tech Meeting

Time/Place

This meeting is a hybrid teleconference and IRC chat. Anyone is welcome to join...here's the info:
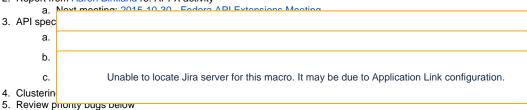
- Time: 11:00am Eastern Daylight Time US (UTC-4)
- Dial-in Number: (712) 775-7035
  - Participant Code: 479307#
  - International numbers: Conference Call Information
  - Web Access: https://www.freeconferencecallhd.com/wp-content/themes/responsive/flashphone/flash-phone.php
- IRC:
  - Join the #fcrepo chat room via Freenode Web IRC (enter a unique nick)
  - Or point your IRC client to #fcrepo on irc.freenode.net

Attendees

- Unknown User (acoburn)
- Bethany Seeger
- Jared Whiklo ⭐
- Yinlin Chen
- Esmé Cowles
- Aaron Birkland
- A. Soroka
- Andrew Woods
- Nick Ruest
- Unknown User (daniel-dgi)

Agenda

1. Report from Andrew Woods re: performance testing meeting

   a. Next meeting: 2015-11-09 Performance - Scale Meeting
2. Report from Aaron Birkland re: API-X activity
   a. Next meeting: 2015-10-30 - Fedora API Extensions Meeting
3. API spec
   a.
   b.

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

4. Clustering
5. Review priority bugs below
6. Close (or justify) old pull requests https://github.com/fcrepo4/fcrepo4/pulls, especially those older than ~6 months
7. ...

Ticket Summaries
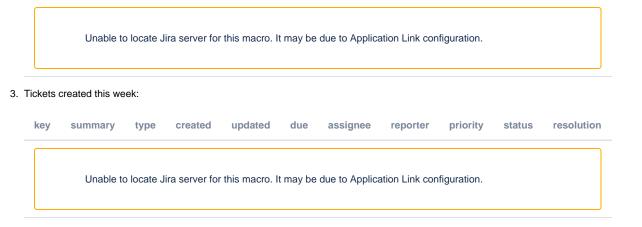
1. Please squash a bug!

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. Tickets resolved this week:

key	summary	type	created	updated	due	assignee	reporter	priority	status	resolution

3. Tickets created this week:

key	summary	type	created	updated	due	assignee	reporter	priority	status	resolution

Minutes

1 ------ Report from Andrew Woods re: performance testing meeting

The initial conversation was focused on surfacing the salient requirements and concerns in the areas of performance and scalability. It was recognized that a reproducible benchmarking suite, collaboratively defined and agreed upon by the community, would serve the purpose of clarifying the current state of Fedora as well as identifying changes with subsequent Fedora releases.

The benchmarking items are being collected on the wiki (https://wiki.duraspace.org/display/FF/Performance+and+Scalability). You are encouraged to review and update the content of that page.

Additionally, Fedora performance testing has been executed and documented in the past. Links to this work are also referenced in the above wiki page. As a first step in designing a benchmarking approach, this prior art must be reviewed.

A scripted set of tests that runs over F4, and document in a systematic way over a set of community identified goals?
What are the benchmarks?
What are the tools?
Where we are lacking, adding tickets to make improvements.

Very large video files, and LDP interactions (Hydra) Questions where current LDP interactions can be improved on both client and server side.

Batch migration and ingests with fixity on the way in. Sharing of fixtures to have common tools.

Doron Shalvi has started to do some scalability testing on the migration front. Run into some OOM in migration-tools side, not on F4 side. Andrew Woods suggests what is most helpful is to create a Jira ticket with all relevant information so it can be properly addressed. Michael Durbin has been waiting for some real feedback before investing time.

Discussion around knowledge of large scale migration tests in the community.

Meetings every two weeks for performance and scalability. Next meeting on Monday Nov. 9, 2015 @ 1:00pm ET

2 ------ Report from Aaron Birkland re: API-X activity
Collected 30 use cases from various institutions.
Agreed to move on to distilling requirements.
Criteria for evaluation agreed upon.
Started working through the evaluation of use cases.
Meeting tomorrow Friday Oct. 30, 2015 @ 2:00pm ET. Meeting bi-weekly.
Currently no formal roles or time commitments, are they ready to formalize.

The async storage effort had the idea to work with API-X effort. The use case from async storage will be sent to API-X to see which parts might or might not be applicable.
Some synergy between these two efforts.

Aaron Birkland is now working for Johns Hopkins. To be able to invest more time in the API-X architecture.

3 ------ API specification: what triples are we committing to on which resources
Fedora is moving toward a tightly specified API. One part is that when you make a request on a RDF source you get specific triples. However in some instances the triples returned are not always the expected or desired ones. This clean-up is part of moving Fedora towards a formalized API as one of our technical priorities https://wiki.duraspace.org/display/FF/2015+-+2016+Technical+Priorities
.
Better to have an assessment of the triples Fedora provides and their value to the community instead of a bunch of tickets singling out one or more problematic ones.
Esmé Cowles feels this is important and would like to help in the effort, I.e. to write up some of this specification.

A. Soroka - are we referring to the contents of the system managed triples "box". Those that are not defined by LDP, but extra triples defined by Fedora?
We should be excluding from this talk the LDP defined triples and restrict ourselves only to the extra server-managed triples Fedora provides.

Aaron Birkland: What about optional LDP features that Fedora supports and the specific way in which Fedora supports them?
A. Soroka: This is specifically Fedora server managed triples. Or the reverse: is there any LDP defined triples the community can think of that should be part of this process?
Andrew Woods: I think the Fedora API is a contract between Fedora and it's community and what each expects to receive from the other.
A. Soroka ~~~~~~ed for and therefor whether we need

them. (ie.   Unable to locate Jira server for this macro. It may be due to Application Link configuration.   - rdf:type versus fedora:mixinTypes)

Strip unused triples and document any that will continue to be presented.

A. Soroka feels we will be on the prune side more than document side. Unknown User (daniel-dgi) agrees.

Is this a high-level ticket
A. Soroka feels that all triples that remain should be mapped to RDF. Unknown User (daniel-dgi) possibly dcterms??

Esmé Cowles and A. Soroka both volunteered for this effort.
One ticket to perform a review which will result in more tickets or subsuming some other tickets.
Related tickets:

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 -------- Clustering activity and opportunities
avmich and p_ire are interested in clustering and are reaching out for assistance.
Are we reaching critical mass on a group that can help move clustering forward and improving our documentation.
https://github.com/infinispan/infinispan/wiki/Design-For-Cross-Site-Replication

Encouraging a more concerted effort around clustering.
Unknown User (daniel-dgi) is interested in joining the effort.

Should we try to get the community to articulate some specific goals/objectives around clustering in Fedora and provide a forum to work on these.

Should the first call out be to just figure out if there is a common understanding of what "clustering" means.
General Fedora configuration and scripts around the setup/testing of clustering would be available.
Unknown User (daniel-dgi) to make the call out on the fedora-tech list and initial wiki-page.

5 ------ Review priority bugs below

Some high priority b

Esmé ~~~~~~uration.   is a high priority. Possibly

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

A. Sor~~~~~~~~~~
There is a discussion to be had and is being had on e-mail list.

Unable to locate Jira server for this macro. It may be due to Application Link configuration.	- Esmé Cowles feels this is straight up and
	- follow-up to single subject work.
	- is same as above
Unable to locate Jira server for this macro. It may be due to Application Link configuration.	- Andrew Woods feels this is low-hanging

6 ----- Close (or justify) old pull requests, especially those older than ~6 months

Tabled to next meeting