Washington D.C. Area Fedora User Group Meeting and Hackathon: 17-18 December 2018

When

17-18 December 2018
10:00am - 4:00pm each day

Remote Participation

https://duraspace.zoom.us/my/fedora

Where

Johns Hopkins University, Baltimore, MD

Milton S Eisenhower Library M Level Digital Research and Curation Center (DRCC) Conference Room

Precise location on google maps - head for the security desk, which is on the northern side end of the floor. The DRCC is located in staff areas (through the hallway to the right as you face the circulation desk) behind the security desk. Tell security guards you are here for the Fedora Users Group meeting at the DRCC

Visitors will need to provide a valid photo ID

Arriving by Metro/Bus

Homewood Shuttle

Purple Charm City Circulator

Arriving by Car

Campus Visitor Parking

Lodging

Closest hotel is the Inn at the Colonnade

Meals

Morning coffee and snacks will be provided.

There are many wonderful lunch options around campus,

Meeting Notes

Google Docs

What to Bring

A laptop for hackathon activities

Attendees

If you plan to attend the meeting, please add your name below.

- Aaron Birkland Johns Hopkins University
- Joshua Westgard, University of Maryland
- Doron Shalvi , National Library of Medicine

- Ben Wallberg , University of Maryland (Monday only)
- Jennifer Gilbert, National Library of Medicine (Monday only)
- Hanh Vu Johns Hopkins University
- Jim Martino Johns Hopkins university
- Lindsay Franz National Library of Medicine
- Mark Patton, Johns Hopkins university
- Karen Hanson, Johns Hopkins university
- Elliot Metsger, Johns Hopkins university
- Peter Eichman, University of Maryland
- Aseem Dhakal. NASA Goddard Library
- Mohamed Abdul Rasheed, University of Maryland

Agenda/Presentations

The theme of this Fedora user's group meeting is improving Fedora based on experience in the field. With more institutions having experience with Fedora 4+ in production, or interested in adapting fedora 4+, now is an excellent opportunity to translate this experience into improvements in Fedora. In that context, the format of this meeting will be a series of presentations, updates, and/or discussions, followed by a hackathon with the goal of producing code, design, or documentation.

Possible discussion/presentation/update topics

(please add items here. Feel free to comment or add your name if a topic is of particular interest)

- The Public Access Submission System (PASS) JHU's experience and pain points
  - Link is to a google doc with various notes from our experience. This would be the basis of a presentation and one of many inputs for possible hackathon topics
- Fedora from a sysadmin perspective - configuring, deploying, or maintenance (e.g. backup/restore).
- Performance - experience reports, workarounds, and wins.
- Roundtable discussion: Making the case for Fedora. Despite some past efforts to do this, it cannot hurt to reconsider the project's goals and why decisions have been made. If potential adopters are not on board with the rationale behind, for example, the adoption of RDF and LDP, it will continue to be difficult to convince them that the extra overhead those choices entail is worth it.

Hackathon tasks

TBD. This will be determined by a discussion at the end of the first day, based on shared interests, experience, and practical consideration. The goal at the end of the first day will be to produce a concrete set of tasks for individuals or teams to accomplish on the second day. Ideally, these tasks will be inclusive of the talent in the room. Possible areas include software development, design, or documentation updates.

Schedule

(proposed. Feel free to add a presentation to the calendar, or ad it to the discussion/presentation topic list above)

Day 1

Time	Topic	Presenter
10:00 - 10:30	Welcome and Introductions	
10:30 - 11:15	Update from UMD	Joshua Westgard Peter Eichman
11:15-12:00	Update from NLM	TBD
12;00-1:00	Lunch Break	
1:15 - 2:00	Update from JHU	Aaron Birkland Hanh Vu
2:00- 2:15	Fedora 5 Update	
2:15 - 2:30	Break	
2:30-4:00	Roundtable discussion	

Day 2

Time	Topic	Presenter
10:00 - 11:30	Hackathon planning	

12:00 - 1:15	Lunch Break	
1:15 - 4:00	Hackathon	

Summary and Outcomes

- Use usage of Fedora by UMD, NLM, and JHU were significantly different from one another
  - NLM: Fedora feeds mostly static content to Solr via transformation (gsearch). Solr powers user-facing Blacklight
  - UMD: Repository-centric architecture relying heavily on async messaging. UIs provided by HippoCMS, and potentially a Samvera head to suit different use cases.
  - JHU: Single-page app in browser makes request directly to Fedora (protected by Shibboleth, ACLs), and elasticsearch. Fedora supports interactive workflows and async services, is not used as a preservation repository
- While use cases were diverse, users group members are interested in tackling infrastructure challenges together; recognized the community of practice around Fedora as valuable.
- There was disagreement about the utility of LDP. Perspectives ranged from "essential to the way the application works" to "we have not moved to Fedora 4 in part due to LDP-centricity"
  - See also Stanford evaluation of Fedora, which cites the LDP API as a reason to discount Fedora as an option for their repository.
- The prospect of an OCFL-based Fedora was compelling to all attendees, but for different reasons
  - Use cases around using OCFL to organize content on disk via a separate process, have Fedora able to expose it in a way suitable for SOLR indexing
  - Attracted to transparency of filesystem content
  - See an advantage of OCFL as far as a plausible path to migration, where migration tools produce content in OCFL that Fedora can then index and serve
    - Compelling Fedora 3 Fedora OCFL migration path
    - Migration has been a kind of second-class citizen in fedora 34, 45 OCFL offers more/simpler(?) potential migration paths, which lends credibility to the notion of a successful migration
- Hackathon activities focused around taking objects modeled in each institution's repository, and attempting to model in OCFL.
  - Two areas of mapping Fedora resources to OCFL were explored
    - 1:1 mapping, where each Fedora resource corresponds to an OCFL object
      - Worked for JHU, UMD cases. Probably not acceptable for NLM
      - Most difficult to navigate just by looking at objects in OCFL. Physical model is flat, software would need to build a logical model based on contents of the OCFL objects
        - Fedora could store the RDF it needs to maintain an ldp hierarchy in a special area of the OCFL objects, maybe a .fcrepo directory, or something like it
        - In the case of UMD, they do not use containment to build a logical model, but instead use PCDM. The notion of segregating Fedora's server-managed triples from user-managed triples as separate files in OCFL was seen as a good approach
    - "tree" mapping", where multiple Fedora resources are encapsulated in an OCFL object
      - 1:1 mapping, which worked in JHU and UMD use cases, is a special case of this,
      - Tree mapping most strongly aligns with NLM, and probably many Fedora 3 users
      - Would want Fedora to "by default" provide a basic mapping of files and directories in an OCFL object to resources exposed by the API. i.e. absent any fedora-specific metadata or triples, the object can still be exposed in a useful way by Fedora API
        - This would work well for NLM's use case of populating an OCFL directory as a method of ingest, or as the result of a content migration
      - Mapping of an OCFL object to an LDP container, where LDP children are artifacts in an OCFL object seems doable, but API might need amending. e.g. add an "OCFL object" interaction model to a container
      - Problem for Fedora's API comes in managing a sequence of updates to such an object. Almost need transactions to encapsulate a series of changes as a single version
      - Since ocfl objects are whole-object versioned, mementos for individual resources would be straightforward, all resources in a version would be in lock-step with one another

Additional Reading

As we think about next steps, it would be helpful to compile a reading list of the published evaluations of the current stack and other publications of relevance to the alternative approaches being explored in 2019. Please feel free to add anything you think might be useful to consider as part of this environmental scan. This page should be conisdered a temporary parking space for this information until there's a better location.

Linked Data: Usability and Adoption

- Ruben Verborgh, "Designing a Linked Data developer experience: Making decentralized Web app development fun"

Assessments of Fedora on Modeshape

- Fedora Leaders meeting at CNI 2018
- Fedora Performance and Scalabilty working group
- Fedora Assessment for TACO Truck
- Jonathan Rochkind, "On the present and future of samvera technical architectures"
- Adam Wead, "Performance Metrics with Valkyrie and Hyrax" and "Benchmark Testing with Valkyrie and Hyrax"
- There was another blog post recently that I've been unable to locate (if anyone knows please add – it was a self-described brain dump from somebody who had worked on Samvera)

OCFL and Potential Advantages of the Filesystem

- OCFL specification, v0.1
- Alex Garnett, Mike Winter, and Justin Simpson, "Checksums on Modern Filesystems, or: On the Virtuous Consumption of CPU Cycles"
- Roy Tennant, "A Paean to the Filesystem"