

# Washington D.C. Area Fedora User Group Meeting and OCFL Hackathon: 7-8 May 2019

## When

7-8 May 2019  
10:00am - 4:00pm each day

## Where

National Library of Medicine  
8600 Rockville Pike  
Bethesda, MD 20894

## Arriving by Metro

The NIH campus is very accessible via [Metro](#). Take the **Red Line** towards Shady Grove to the **Medical Center** station. The station's escalators come out on the NIH campus.

You'll then need to get a visitor badge from the [NIH Visitor Gateway Center](#), which will be facing you as you come up the Metro escalator. You will be required to show one form of identification (a government-issued photo ID-driver's license, passport, green card, etc.) and to state the purpose of your visit.

On exit from the Gateway Center, you will head up a path which leads right to NLM - bear left up the hill, walking past the Natcher Building (45) on your right before crossing the street to NLM. The meeting is in the Lindberg Room, on the 2nd Floor of the main library, which is the smaller building (38) in front of the taller building (38A). Walk left past the row of parked cars, then bear right, up to the main entrance, which faces Rockville Pike. Take the stairs or elevator, on the right, up one floor. See the [NIH Campus Map](#).

## Arriving by Car

See the [Driving Directions to the NIH Main Campus](#).

Park in the nearby MLP-11 parking lot, which is just outside the NIH campus. You'll then need to get a visitor badge from the nearby [NIH Visitor Gateway Center](#). You will be required to show one form of identification (a government-issued photo ID-driver's license, passport, green card, etc.) and to state the purpose of your visit. See [the map](#) showing the MLP-11 lot and NIH Visitor Gateway Center (66) in relation to Rockville Pike (Rte 355).

On exit from the Gateway Center, you will head up a path which leads right to NLM - bear left up the hill, walking past the Natcher Building (45) on your right before crossing the street to NLM. The meeting is in the Lindberg Room, on the 2nd Floor of the main library, which is the smaller building (38) in front of the taller building (38A). Walk left past the row of parked cars, then bear right, up to the main entrance, which faces Rockville Pike. Take the stairs or elevator, on the right, up one floor. See the [NIH Campus Map](#).

## Remote Participation

Meeting link: <https://nih.webex.com/nih/j.php?MTID=m7a4060243ceec1421591bcb6ca204c3d>

Meeting number: 628 339 679

Password: hZmAZCD2

Join by phone: 1-650-479-3208 Call-in toll number (US/Canada); Access code: 628 339 679

## Refreshments

We are looking for a sponsor for light refreshments; coffee / pastries.

## Meeting Notes

1. Fedora 6 open questions and design issues
  - a. **OCFL only.** Should Fedora 6 support multiple storage layer technologies, or just the OCFL specification? Some feel that Fedora 6 should not be strictly tied to only OCFL. Also being able to work with cloud storage is a major concern (see below, "Cloud").
  - b. **Implicit OCFL.** Many people see value in a filesystem storage layer, such as specified in OCFL, for Fedora; however, do not want to tie Fedora 6 strictly to OCFL. Is there value in indicating that the Fedora 6 design will provide filesystem-based storage without explicitly stating that the storage spec is OCFL, and just use OCFL under the hood?
  - c. **Other filesystem approaches.** A filesystem-based storage approach seems like a useful option for Fedora 6 in any case. If Fedora 6 were to use a filesystem-based storage approach that was not OCFL, what would it use? What would be the value proposition of that storage approach over OCFL?

- d. **Pluggable storage.** There seems to be a desire for a diversity of storage options. If Fedora 6 were to support multiple storage layer technologies (e.g., OCFL, PostgreSQL), would these storage options need to be available via a single Fedora distribution (internal pluggable components), possibly the default community implementation? Or should these be wholly separate Fedora implementations of the Fedora spec?
  - e. **Transparency.** The group feels that regardless of the storage technology used, the technology should offer file transparency. It should be obvious to humans how to access, navigate and interpret the files stored by Fedora. It may be useful for Fedora to cache items in a non-transparent manner for performance, but by default Fedora should offer file transparency.
    - i. An interesting thought exercise is to imagine a version of Fedora (4, 5) that uses Modeshape but did not suffer from the "many members" (or other?) performance limitations. In Modeshape files are not transparent and not easily human-understandable. Would such a performant Fedora-Modeshape have broad community appeal, even if it lacks file transparency?
    - ii. Some participants felt that a Fedora implementation that resolved the performance issues inherent in the current implementation (specifically "many members"), and that could both read from and write to OCFL (via import/export tooling), but which did not use OCFL as its native storage layer, could be a sufficient solution.
    - iii. It could be useful to explore using the current export tool to persist resources to disk and then use the prototype OCFL clients to create OCFL objects from these exported resources. Modifying the existing import tool to be able to re-ingest an OCFL-ized export seems like low-hanging fruit.
  - f. **Data preparation.** OCFL seems to make great sense for the use case of organizing files in advance, and then presenting Fedora with this organization. What about the use case when Fedora is presented with the files via the API, and must store the files? Should the files be stored within an OCFL structure as well, or something else?
  - g. **Maturity.** OCFL is still an untested pre-beta specification. Is it mature enough to use as the only storage technology for Fedora 6? It seems similar to the level of maturity of WebAC when Fedora started with it.
  - h. **Performance.** File transparency likely comes at a cost in performance. Internal caches may improve performance but could introduce problems if there is delay in synchronising the authoritative store with the cache. Apparently in the past, some organizations have indicated that a one second delay in such synchronisation was unacceptable. How big of a concern is performance in this respect? The group generally feels that the majority of Fedora use cases do not have such stringent performance requirements, and would greatly benefit from file transparency. Fedora users are generally not concerned with high performance computing, and this use case is perhaps better met by a specialised Fedora implementation or another application.
  - i. **Cloud.** Many organisations are moving to the cloud and it seems reasonable to expect to be able to run Fedora 6 in the cloud, and in particular, on AWS S3. Recent OCFL community calls have implied that OCFL is geared towards POSIX filesystems. Will use of OCFL introduce limitations to Fedora in a cloud environment? Will EBS be required to run Fedora 6 within AWS?
2. Fedora 6 development
    - a. In order to obtain institutional commitments to use Fedora 6, it would be helpful to have a basic design in place that answers many of the above questions. This would allow institutions to determine the feasibility of using the product.
    - b. A working prototype may also enable institutions to better evaluate their potential use of Fedora 6. However, prototype development in advance of detailed design (above) may not be a good use of resources.
    - c. To drive adoption, Fedora 6 should be reasonably simple to use, offer file transparency, and offer a migration path for the existing broad Fedora 3 user base.
    - d. The group notes that additional support is needed to develop Fedora 6, in addition to the current technical team.

## What to Bring

A laptop for hackathon activities

## Attendees

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

- [Doron Shalvi](#) , National Library of Medicine
- [Joshua Westgard](#) , University of Maryland
- [Peter Eichman](#) , University of Maryland
- [Ursula Pieper](#) , National Agricultural Library, USDA
- [Ben Wallberg](#), University of Maryland (Day 1 only)
- [Jennifer Gilbert](#) National Library of Medicine
- [Lindsay Franz](#) National Library of Medicine
- [Bethany Seeger](#) , Amherst College
- [Aaron Birkland](#) , Johns Hopkins (remote)
- [Greg Jansen](#), University of Maryland
- Steve Liu, National Library of Medicine
- Xin Wang, National Transportation Library, USDOT
- [Yanru Bi](#), National Library of Medicine
- [Calvin Xu](#) National Library of Medicine (remote)
- [Elliot Metsger](#) Johns Hopkins (remote)
- [TA Nguyen](#), National Library of Medicine (remote)

## Agenda/Presentations

With the recent release of Fedora 5.0, the theme of this Fedora user's group meeting is to discuss the current and upcoming versions of Fedora, and how best to meet institutional needs. 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.

### Day 1 - Presentations / Discussions

Possible topics include the following. Please feel free to suggest topics. If you would like to present at the meeting, please add your name to the schedule below (note: times are subject to change).

- DuraSpace / Lyris merger
- Fedora roadmap
  - requirements and use-cases for the next version of the community implementation of the Fedora API
  - status update on alternative implementations of the API (DRAS-TIC? Lakesuperior? Trilpy? Others?)
- Institutional updates

| Time          | Topic                                                               | Presenter                                                              |
|---------------|---------------------------------------------------------------------|------------------------------------------------------------------------|
| 10:00 - 10:20 | Welcome and Introductions                                           | All                                                                    |
| 10:20 - 10:40 | <a href="#">DuraSpace update</a>                                    | <a href="#">David Wilcox</a>                                           |
| 10:40 - 11:00 | Fedora update                                                       | <a href="#">Bethany Seeger</a>                                         |
| 11:00 - 11:30 | <a href="#">UMD Update</a>                                          | <a href="#">Joshua Westgard</a>                                        |
| 12:00 - 1:10  | Lunch Break                                                         |                                                                        |
| 1:10 - 1:40   | <a href="#">Modern Manuscripts and UI enhancements</a>              | <a href="#">Jennifer Gilbert</a> , <a href="#">Lindsay Franz</a> , NLM |
| 1:40 - 2:10   | <a href="#">Drastic Fedora</a> / Trellis LDP / Cassandra            | <a href="#">Greg Jansen</a>                                            |
| 2:10 - 2:30   | Amherst College Repo. Update                                        | <a href="#">Bethany Seeger</a>                                         |
| 2:30 - 2:40   | Break                                                               |                                                                        |
| 2:40 - 3:00   | <a href="#">Repository Ecosystem Architecture and ArchivesSpace</a> | Steve Liu, <a href="#">Doron Shalvi</a> , NLM                          |
| 3:00 - 3:30   | <a href="#">Fedora 6 / OCFL Design Implications</a>                 | Virginia Beach Stowaways                                               |
| 3:30 - 4:00   | Tour of History of Medicine Division                                | NLM                                                                    |

## Day 2 - Hackathon

The structure of the hackathon is TBD and 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. Possible topics include the following; please add additional topics.

- Whiteboarding how a particular use case would be handled in OCFL (continuation of [December 2018 hack-a-thon](#)).
- Developing / using experimental software to interface with an OCFL structure.

| Time          | Topic                                   | Presenter |
|---------------|-----------------------------------------|-----------|
| 10:00 - 11:00 | OCFL Design Structures for Repositories |           |
| 11:00 - 12:00 | OCFL Hackathon                          |           |
| 12:00 - 1:15  | Lunch Break                             |           |
| 1:15 - 4:00   | Hackathon (continued)                   |           |

## Resources

- [Washington D.C. Area Fedora User Group Meeting and Hackathon: 17-18 December 2018](#)
- [2019-02 Fedora Design Summary](#)
- [OCFL Draft Spec \(29 April 2019\)](#)
- [Aaron Birkland OCFL client](#)
- [Simeon Warner OCFL implementation](#)