

2020-09 Online Fedora Users Group Meeting - European Timezones

This meeting is open to anyone who would like to attend; it will provide an opportunity for members of the Fedora community to connect, share information, and provide updates on local projects and initiatives. This meeting will be based around European timezones, but anyone is welcome to attend.

When

September 9, 2020

Where

Online via Zoom

Agenda/Presentations

All times are CEST.

Time	Topic	Presenter
13:00 - 13:10	Welcome and Introductions	
13:10 - 13:30	Fedora Program and Community Update	David Wilcox, LYRISIS
13:30 - 13:50	Islandora Updates	Melissa Anez, Islandora Foundation
13:50 - 14:10	Phaidra - University of Vienna	Raman Ganguly, University of Vienna
14:10 - 14:30	Break	
14:30 - 14:50	Memobase 2020: Using Fedora as central repository to store linked data and digital objects	Thomas Bernhart, Docuteam
14:50 - 15:10	A field report from the relaunch of an infrastructure: A FEDORA based framework for long-term storage and dissemination of research data	Johannes Stigler, Centre for Information Modelling, University Graz
15:10 - 15:30	Break	
15:30 - 16:20	Lightning Talks: <ol style="list-style-type: none">1. Fedora Commons in the CLARIN-D infrastructure2. eHumanities3. Past problems - Fine Future4. Unversity library "Svetozar Markovi"	<ol style="list-style-type: none">1. Thomas Eckhart, Leipzig University2. Jaime Penagos, University Library Ludwig-Maximilian University Munich3. Oliver Schöner, Berlin State Library4. Dragana Stoli, Unversity library "Svetozar Markovi"
16:20 - 16:50	Wrap-up and Discussion	

Notes

13:10 - 13:30 Fedora Program and Community Update David Wilcox, LYRISIS

Questions:

- What is the timeline for release of the Fedora 6 beta?

- What will the DB cache be used to store?
 - All of Fedora's content and metadata are stored on disk in an OCFL storage root. For performance reasons, Fedora maintains a cache of system and user metadata in a rebuildable database: <https://wiki.lyrasis.org/display/FEDORA6x/Configuring+JDBC+Object+Store>

13:30 - 13:50 Islandora Updates Melissa Anez, Islandora Foundation

1. Islandora: Drupal + Fedora
 - a. Islandora 7: Vertical stack (cheeseburger)
 - b. Islandora 8: Distributed architecture (bento box)
2. Using Fedora for smart storage, fixity, versioning (memento), and OCFL (in version 6)
 - a. Only putting preservation copies in Fedora; access copies are stored in Drupal
3. Flysystem
 - a. Fedora is treated as the Drupal filesystem from the user perspective
4. Islandora community
 - a. 320+ installations
 - b. Backed by the Islandora Foundation, which is member supported
5. Examples:
 - a. Archives Central probably the first digital archive using the Records in Context ontology in production
6. Questions
 - a. Does Islandora make assumptions about data model used to describe the collection default form is based on MODS, but Islandora is very flexible regarding the data model (use Drupal forms and map those to Fedora properties)

13:50 - 14:10 Phaidra - University of Vienna

1. Phaidra - repository system based on Fedora
2. Started in 2007
3. Both the repository itself and the services around the repository
4. Community: Austria, Italy, Western Balkans
5. Architecture
 - a. Changed from monolithic to modular
 - b. Fedora at the centre, surrounded by Phaidra Core
 - c. System of plug-ins and a rich API
 - d. Multiple user interfaces and workflows
6. Workflows
 - a. 4 phased workflow model
 - b. Pre-ingest, ingest, management, re-use
7. Focused on long tail of research data

14:30 - 14:50 Memobase 2020 - Thomas Bernhart, Docuteam

1. Using Fedora to store linked data and digital objects
2. Memobase: audio/video collection of Swiss heritage
 - a. <http://memobase.ch>
 - b. Collaboration between University of Basel, Outermidia, and Docuteam
 - c. Built on Fedora 5
3. Data model
 - a. Transition from XML to RDF
4. Architecture
 - a. SFTP server where users can upload data
 - b. Transformed in import pipeline and stored in Fedora
 - c. Processed via messages (Kafka cluster)
5. Fedora integration
 - a. Fedora ingest service
 - b. Messages (Activity Streams) to event handler
 - c. Metadata extractor
6. Challenges
 - a. Mapping external URIs to internal URIs
 - b. Handling references between resources (cannot create relationships to resources that don't exist yet - this is no longer an issue in Fedora 6)
 - c. When to use different LDP container types

14:50 - 15:10 A field report from the relaunch of an infrastructure: A FEDORA based framework for long-term storage and dissemination of research data - Johannes Stigler, Centre for Information Modelling, University Graz

1. Migrated Fedora 3 to 4.7
2. Using Fedora since version 2
3. New versions of Fedora represent a paradigm shift
4. Infrastructure
 - a. Fedora is a storage layer in a complex system that stores research data generated by a variety of digital humanities projects
 - b. Based on XML structure
 - c. 114,000 digital objects in Fedora 3, many of which are complex
 - d. Services are triggered by user uploads
5. Content models
 - a. Structural definitions for object types
 - b. Manuscript transcripts based on TEI
 - c. Content models trigger workflows
 - d. Content models support specific XML schemas
6. Migration project
 - a. Need to migrate Fedora 3 REST API functionality to Fedora 6
7. Java based thin desktop client to handle ingest workflows
8. Using Docker and Kubernetes
 - a. Advantage: isolation
 - b. Disadvantages: complexity, difficulty identifying errors, need to learn new practices
 - c. Using Kubernetes for clustering and orchestration

Questions:

1. From Gunter Vasold: What software is used for the Media Server?
Response: I think it is a custom implementation from the University Basel

15:30 - 16:20 Lightning Talks:

Fedora Commons in the CLARIN-D infrastructure - Thomas Eckhart, Leipzig University

1. CLARIN - Common Language Resources and Technology Infrastructure
 - a. CLARIN-D - National consortium in Germany
2. Distributed infrastructure with standardized interfaces
3. Using OAI-PMH for metadata exchange - most important endpoint
4. CLARIN-D (10 centres): majority Fedora 3.x + ProAI
5. Migration
 - a. Issues migrating to Fedora 4.x
 - b. Majority still using Fedora 3.x
 - c. Created a requirements document

eHumanities - Jaime Penagos, University Library Ludwig-Maximilian University Munich

1. Research data from the digital humanities
2. Lexical information from the Alpine region
 - a. Approximately 500,000 datasets
 - b. Many relations between datasets
3. Infrastructure
 - a. Used to transform XML to specific file formats
 - a. Fedora 5, Solr, Blacklight
 - b. Apache Camel
 - c. Storing CSVs
4. Challenges
 - a. Scalability: issues with messaging queue
 - b. Many concurrent operations can lead to performance degradation
5. Questions
 - a. Cluster vs. standalone?
 - b. Impacts of eventual migration to Fedora 6?

Past problems - Fine Future - Oliver Schöner, Berlin State Library

1. Many Fedora 4.x repositories at the Berlin State Library
 - a. ITR (East Asian department): Many millions of objects (mostly pages)
 - b. Stabirep: Born digital materials
 - c. RGZ: Prussian judgments
 - d. DC: Bibliographic materials (no binaries)
2. Fedora 4.5.1 server shutdown due to power outage
 - a. One file in the Infinispan layer had a bad OS owner
 - a. Fedora restarted but Tomcat did not
 - b. Problem was difficult to discover and solve
 - c. If migrations were easier the data could have been transferred to a new instance with an upgraded Fedora
3. Believe that OCFL is the future

Phaidra / Islandora - Dragana Stoli, University library "Svetozar Markovi"

1. Using Phaidra for almost a decade
2. Started using Islandora 2 years ago
3. Website: <https://phaidrabg.bg.ac.rs>
 - a. Started in 2011
 - b. PhD theses
 - c. Permanent archiving
 - d. Looking for a more adjustable tool
4. Serbian Literary Criticism
 - a. Using Islandora 7
 - b. Self-archiving of literary criticism texts
5. Phaidra is very user-friendly and simple
6. Islandora is more adjustable
7. Archiving ethnographic data
 - a. Ethnographic data is very compatible with Fedora / Phaidra
 - a. Which repository to choose?
 - b. Decided to return to Phaidra when the time comes (probably next year)

Wrap-up and Discussion

1. Nice to have an event scheduled around European timezones to hear from the European Fedora community
2. Issue with Fedora: too many major version upgrades, hard to absorb
3. Many smaller institutions have a hard time installing and setting up Fedora. Not out-of-the-box enough
 - a. ISLE provides a Docker deployment of Islandora
 - b. Phaidra also looking into Docker deployment
 - c. Some institutions choose Fedora over an out-of-the-box solution in order to achieve a lot of flexibility
4. How often to have users group meetings?
 - a. An open forum to ask questions
 - a. Meetings should occur at least once per year, or perhaps once every 6 months
 - b. More discussion on problems and open issues, difficulties
 - c. Another call after some initial experiences with migrations to discuss

Resources