

Fedora 4 for Research Data

David Wilcox
Fedora Product Manager
DuraSpace

What is Fedora?

- Flexible Extensible Durable Object Repository Architecture
- Open Source digital repository software
- Community-driven design and implementation
- Fedora 4 is the new, revitalized version

2014 Fedora Members

(52)

- Arizona State University Libraries
- Brown University Library
- Case Western Reserve University Libraries
- Charles Darwin University
- Colorado Alliance of Research Libraries (CARL)
- Columbia University Library
- Cornell University
- Durham University
- FIZ Karlsruhe
- Ghent University Library
- Gothenburg University Library
- Indiana University
- ICPSR
- Johns Hopkins University Libraries
- La Trobe University
- London School of Economics & Political Science
- LYRISIS
- Macquarie University
- National Library of Medicine
- National Library of Wales / Llyfrgell Genedlaethol Cymru
- Northeastern University Libraries
- Northwestern University Libraries
- Ohio State
- Oregon State
- Pennsylvania State University
- Rutgers University Libraries
- Smithsonian Institution, Office of Research Information Services
- Stanford University
- State and University Library of Denmark
- The Art Institute of Chicago
- Tufts University
- University of Alberta
- University of California, Los Angeles
- University of California, Santa Barbara
- University of Cincinnati
- University of Connecticut Libraries
- University of Hull
- University of Lausanne
- University of Manitoba
- University of Massachusetts Amherst Libraries
- University of New South Wales
- University of Notre Dame
- University of North Carolina
- University of Pittsburgh
- University of Oxford
- University of Prince Edward Island
- University of Rochester Libraries
- University of Texas Libraries Austin
- University of Toronto
- University of Virginia
- University of Wisconsin
- Yale University

Hydra and Islandora

- Popular Fedora front-end applications
- Both will support Fedora 4
 - PSU Beta pilot is Hydra-based
 - Currently organizing code sprints for Islandora integration

Research data requirements

- Diverse formats
- Complex relationships
- Large file sizes
- Many files
- Long-term preservation and access
- Others...

Flexible storage

- Federate Fedora over an external file system
 - Use Fedora's management and preservation features
 - No need to ingest large datasets
- Pluggable support for other back-end systems
- Support for asynchronous storage is on the roadmap

Fixity and versioning

- Verify and maintain the fixity of repository objects
 - Checksums can be calculated and verified on ingest
 - Checksums can be recalculated at any time
- Support for flexible versioning
 - Turn on versioning for the entire repository
 - Or selectively create new versions using a REST API call

Backup and long-term access

- The repository can be backed up and restored at any time
- Particular objects or groups of objects can be exported and imported
- The repository can also be exported in JCR/XML format for long-term transparency



- Entering 4th year of production
- Integrated with Fedora 3.x, Planning Fedora 4 integration
 - Also integrated with DSpace 3.x and higher, Dspacedirect, Archive-it
- TDL running platform locally in production (first case)
- In beta testing Archivematica-DuraCloud integration, production early 2015
- Integrated with Chronopolis to be DPN first node (in pilot now)

Policy-driven storage

- Different types of files may need to be stored in different back-end filesystems
 - Policies can be configured to route files to different locations on ingest
 - These policies can define any number of ingest rules

Transactions

- Writing data to disk is slow
- Transactions bundle multiple actions together
 - Results: 30-60% performance increase
- Transactions can be rolled back

Clustering

- Fedora 4.0 supports clustering for high-availability
 - Multiple Fedora instances in replication mode
 - Load balancer distributes server requests evenly
 - Failover if one instance goes down

Scalability

- A number of scalability tests have been run:
 - Uploaded a 1 TB file via REST API
 - 16 million objects via federation
 - 10 million objects via REST API

How to get involved

- Platforms for research data BoF at RDA
4th plenary in Amsterdam
- Contribute use cases
- Complete acceptance tests
- Contribute developer effort

Resources

- Fedora 4 wiki
 - <https://wiki.duraspace.org/display/FF/>
- Fedora 4.0 Features
 - <https://wiki.duraspace.org/display/FF/Fedora+4.0+Feature+Set>
- Mailing lists
 - <https://wiki.duraspace.org/display/FF/Mailing+Lists+etc>

Questions?