Acknowledging and engaging with the Problem

Most Fedora installations are unsupported (3.x or earlier)

There are many risks to running unsupported software

Migrations to Fedora 4.x and newer have been challenging
Assessing Barriers to Migration

IMLS funded grant to investigate barriers to migration

Focused on assessment and recommendations

Completed in August, 2019
Outcomes

- Identify most common barriers to migration
- Determine where best to focus effort
- Apply for follow-up grant funding to pursue solutions
Survey Results: highlights and takeaways
Demographics

111 total responses

Most institutions Academic (80.9%) or Research (17.3%)

Most respondents located in the United States (72.7%)
Content Types

Cultural heritage and special collections data (85.6%)
Archival materials (77.5%)
Scholarly publications (56.8%)
Research data (46.8%)
Off the top of your head, roughly how many objects are in your repository?

111 responses

- 100,000 objects or less: 45 (40.5%)
- 500,000 objects or less: 29 (26.1%)
- 1,000,000 objects or less: 13 (11.7%)
- 5,000,000 objects or less: 13 (11.7%)
- 10,000,000 objects or less: 6 (5.4%)
- Over 10,000,000 objects: 2 (1.8%)
- I don't know: 3 (2.7%)
- Not Applicable (N/A): 0 (0%)
Issues Impacting Major Upgrades

Availability of the resources (staff or funding) (68%)

Maturity of the software (61.2%)

Architecture and Design (35%)
To the best of your knowledge, what version of Fedora repository software are you using currently? Skip this question if not applicable.

102 responses

- Fedora 1.x: 0 (0%)
- Fedora 2.x: 1 (1%)
- Fedora 3.x: 81 (79.4%)
- Fedora 4.x: 20 (19.6%)
- Fedora 5.x: 1 (1%)
- I don't know: 5 (4.9%)
Favourite Fedora 3.x Features

Flexible content models (47.5%)
Preservation Capabilities (40.6%)
File System Layout (29.7%)
Favourite Fedora 4.x+ Features

Linked Open Data Support (32.7%)
External Component Integrations (25.5%)
Focus on Existing Standards (23.5%)
“I don’t know” (23.5%)
Things That Would Help

Content Migration Tools (70.5%)
Metadata Migration Tools (61.4%)
Documentation (52.3%)
What Fedora alternatives have you assessed or are you assessing?

51 responses

- 9.8% DSpace
- 1% EPrints
- 1% Invenio
- 1% Digital Commons
- 1% Dataverse
- 1% Esploro
- 1% TIND
- 1% TACO

1/6
Reasons for Assessing Alternatives to Fedora 4+

Because it (Fedora) doesn't meet our needs (~55%)
Because it's a completely different platform (~17%)
Because of concerns over long term support (~6%)
Have you decided to migrate away from Fedora?

54 responses

63% Undecided
16.7% No
20.4% Yes
Survey Consultation and Recommendations
63 survey respondents provided contact information

11 US-based respondents participated in focus group call

Questions provided in advance
Focus Group Summary

Need for community best practices for data models

Labour-intensive migration is a decision point - need compelling reasons

OCFL-backed Fedora 6 is compelling, makes migration easier

Security risks of staying on Fedora 3 are a real concern
Recommendations

Focus on effort and value

1. Develop and deliver Fedora 6 with OCFL
2. Develop migration tools
3. Test Fedora 6 at scale
4. Document case studies, data models, and best practices
Fedora 6
Technical barriers to migration (from F3)

- Fedora 4,5 doesn't have the same concept of "object" as F3
- Need to use different URLs.
  - e.g. info:fedora/my:pid -> http://localhost:8080/fcrepo/rest/ab/cd/ef/abcederg
- Performance and stability of Fedora 4 and 5
  - The "many members" problem
- Data in Modeshape (underpinning Fedora 4,5) is opaque
  - Nobody fully understands the native format
  - Simple questions are hard "How many objects are in my repository"
## Design for migration

<table>
<thead>
<tr>
<th>Feature</th>
<th>Fedora 3</th>
<th>Fedora 4,5</th>
<th>Fedora 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDP API</td>
<td></td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Transparent storage</td>
<td>✔</td>
<td></td>
<td>✔</td>
</tr>
<tr>
<td>Structure</td>
<td>Flat</td>
<td>Hierarchical</td>
<td>Either</td>
</tr>
<tr>
<td>Cohesive Objects</td>
<td>✔</td>
<td></td>
<td>✔ (opt-in)</td>
</tr>
<tr>
<td>URL identifiers</td>
<td>info:fedora/</td>
<td><a href="http://localhost:8080">http://localhost:8080</a>...</td>
<td>Configurable</td>
</tr>
<tr>
<td>Basic search API</td>
<td>✔</td>
<td></td>
<td>✔</td>
</tr>
</tbody>
</table>
OCFL: A stable foundation

- What is OCFL?
  - A filesystem layout
  - Emerging standard for long-term preservation (http://ocfl.io)
    - Imagine it as "if bags supported versioning"

- What does OCFL provide?
  - Transparency (easily readable on disk)
  - Robustness (checksums, redundancy)
  - Versioning (capture how content changes through time)

- What does Fedora get out of it?
  - Re-gain robustness/transparency/rebuildability from Fedora 3 in a standardized way
Migration to 6

- Filesystem-to-filesystem migration (FOXML to OCFL) possible!
  - Faster than via Fedora's APIs
  - Inspect/verify afterwards
- Transparent, minimal migration from Fedora 3 is now possible
  - 1:1 mapping between Fedora 3 objects and OCFL objects
  - Don't need to transform RDF/URIs
- Don't forget about migrations from Fedora 4 and 5!
  - Leverage existing import/export tooling, add option to export to OCFL
Pilots
Migration Challenges from Fedora 3.x

- Insufficient Time/Staff/Resources (48.5%)
- Lack of Compatibility with Front End Applications (44.7%)
- Change in Metadata Standards for Description (39.8%)
- Issues with Performance and Scale (33%)
Fedora 6 Pilot - Summary

Limited term (Sept. 2019-Aug. 2020)

Participating institutions
- Docuteam
- National Library of Medicine
- University of Wisconsin-Madison

Fall 2019 Sprints: 1 (Sept. 16-27), 2 (Nov. 4-15)
Sprints planned Spring and Summer 2020
Fedora 6 Pilot - Goals

Validate the design and development of Fedora 6

- Scale, performance
- Ingest
- Migration of Fedora 3 content into Fedora 6 using OCFL
- Transparent file persistence (preservation)
Fedora 6 Pilot - Activities

Pilot institutions:
- Provide sample data for modeling
- Provide use cases and requirements (performance, scale, migrations, ingest)
- Provide feedback on deliverables throughout pilot

Project team:
- Model sample data in Fedora 6
- Review and prioritize pilot use cases and requirements
- Solicit feedback on deliverables
Option to migrate all versions of an object
Validation that each object migrated successfully
Ability to resume a migration that is interrupted
Robust documentation that communicates Fedora’s value and provides implementation guidance
Current Status

Testable migration tool: available now!

Early testable Fedora 6 application: late 2019

Next sprints: First half of 2020

Initial release: mid-to-late 2020
More information: https://wiki.duraspace.org/display/FF/Designing+a+Migration+Path

Contact: david.wilcox@lyrasis.org

Oxford Common File Layout (OCFL) Specification: oclf.io/0.3/spec

OCFL: A Community Developed Approach to Digital Preservation (CNI Spring meeting presentation.)

Fedora 6 Technical Roadmap: https://wiki.duraspace.org/display/FF/Roadmap