LYRASIS

Fedora Migration Paths and Tools: A Pilot Project

The DuraSpace Community Supported Programs division of LYRASIS proposes an 18-month project to develop, pilot, and document migration tools and paths to upgrade the repository software Fedora 3, which is widely used but no longer supported, to Fedora 6. The proposed project is the result of a year-long planning effort funded by IMLS National Leadership Grant LG-72-18-0204-18, which investigated barriers to upgrading at libraries and archives running older, unsupported versions of Fedora. Hundreds of libraries and archives in the U.S. use Fedora repository software to deliver scholarly, scientific, and cultural heritage resources and services to patrons. Continued community reliance on Fedora 3 puts the stability, security, accessibility, and functionality of these repositories at risk. LYRASIS requests $248,500 to produce tools, documentation, and case studies that can support and improve the upgrade process for Fedora 3 users and serve as models for developing resources to assist with future upgrades. The project supports IMLS's goal to "Build Capacity" by strengthening the ability of libraries and archives to sustain repositories that are critical to providing access to digital content and collections.

Statement of Need: Fedora is a flexible, modular, open source repository system for the management and dissemination of content through digital libraries and archives, including specialized access to large and complex digital collections of historic and cultural materials as well as scientific data. Fedora's architecture is built on the principle that interoperability and extensibility is best achieved by providing a limited set of stable, standards-based repository services and common patterns for integrating with other best-practice systems and applications. These services are provided via RESTful APIs. Fedora provides a foundation upon which many types of repository frameworks can be built, including the Islandora and Samvera repositories.

Fedora was first released in 2003. The currently supported versions, 4 and 5, were released in 2015 and 2018 respectively. However, more than two-thirds of installations (approximately 240 institutions) are still using Fedora 3, which was released in 2008. Fedora 4/5 is a completely re-written application, designed to address performance and scale issues, that shares no code with Fedora 3. It impacts underlying technologies, data models, standards for description, and functionality, meaning that an upgrade requires re-modeling and migrating data. Fedora 6 is under development and will be released prior to the start date of this project (9/1/2020), so the proposed plan of work is designed for an upgrade to version 6. The Fedora 6 roadmap addresses community needs for query services and preservation persistence, the latter through implementation of the Oxford Common File Layout. Most Fedora 3 repositories emphasize digital preservation and library-managed, at-risk collections. Fedora 3 is used by a variety of sizes and types of organizations. including liberal arts colleges, library consortia, special libraries, historical organizations, and moderate or higher research universities. About one-third of Fedora 3 repositories are at highest (R1) research institutions. A single repository can store millions of objects and hundreds of terabytes of unique content.

The 2018 IMLS planning project (LG-72-18-0204-18) included an environmental scan, review of sample Fedora 3 institutional profiles, assessment of relevant technologies, a survey of the Fedora community that closed with 111 responses, and an invitational focus group with 11 participants to explore survey results. Findings are documented in a Final Report. Variance and customization among front-end applications were found to have a large impact on the resources required to conduct a migration. Of existing open source migration tools, migration-utils and migrate_7x_claw showed the most promise, but required additional work to serve Fedora 3 repositories. The Fedora API was found to suffer from performance issues with large migrations. The most pressing barriers to upgrading repository platforms were identified as availability of staff and funding (68% of survey respondents), maturity of the software (61%), and architecture and design of the application (35%). Among survey participants who had attempted an upgrade from Fedora 3, the most common barrier continued to be lack of resources (49%), but others included lack of compatibility with front-end applications (45%), changes in metadata standards for description (40%), and issues with performance and scale (33%). When asked what would help to upgrade their version of Fedora, the top three survey responses were content migration tools (71%), metadata migration tools (61%), and documentation (52%). Two important themes emerged during the planning project: effort and value. Any software upgrade/migration that requires significant effort (labor, time, and cost) needs to provide enough additional value to justify the cost. This insight has

informed design of Fedora 6. In the meantime, tools, documentation, case studies, and defined paths will provide resources to support a more effective and less costly upgrade and migration process.

Project Design: The pilot will be implemented with two Fedora 3 institutions representing the majority of front-end application types, Islandora (54% of Fedora 3 users) and custom (38%). The sites would be selected for inclusion in a full grant proposal if invited by IMLS. David Wilcox, Fedora Product Manager, will serve as the principal investigator (PI). In addition to a collaborative team that includes a part-time developer, data migration expert, and representatives from pilot institutions, the Fedora community will be invited to review and test tools and documentation.

The PI will gather information from pilot sites to inform a group meeting to start the project. The meeting will identify areas of overlap and distinction among the repositories, define the desired end state for the pilots, map a process for the initial pilot, and determine the desired documentation deliverables. The pilot upgrades will be done iteratively so that there is time to evaluate results and draft documentation after the first before moving to the second. In general, the desired end states for both pilots are: all data migrated from Fedora 3 to Fedora 6, including mapping of data models and metadata; front-end applications updated to work with the new Fedora API; and adequate performance results for institutions and their repository users. For each pilot:

- The project and repository teams will define a detailed end state, project plan, milestones, and timeline for the upgrade. For data migration, this includes metadata clean-up and improvements as agreed to and scoped by all pilot stakeholders. For the front-end, this includes code changes to comply with the Fedora 6 API, interface updates, and workflow improvements.
- For data migration, the project staff will review sample data, model it in Fedora 6, and define a metadata map. Applicable migration tools will be developed or, if existing tools might work, improved. For example, the migration-utils tool can be updated to work with Fedora 6, developed to work effectively with Fedora 3, and documented with templates. Once tools have been developed, the project team will work with the repository team to test and iterate until migration is complete.
- For the front-end upgrade, project staff will review existing front-end applications and workflow tools. Updates will then be defined, made, tested, and implemented to comply with Fedora 6.
- The fully updated/migrated system will be tested under performance and scale scenarios, such as ingest performance, large files, many users, cloud storage, and horizontal scale. The team also will work with pilot sites to ensure the new system meets needs, including assessments by library staff and end-users (e.g., students, faculty). This will not involve an in-depth usability test, but it will be important to make sure the new system meets the needs of users at least as well as the old system.
- The Fedora project team will document each pilot, produce templates and instructions that could be useful to other sites, and detail the process through a case study.

Each pilot will include one site visit by the Fedora project team at an appropriate point; this could be early during planning, while the upgrade is in process, or during user testing. The second pilot site will apply tools and documentation used at the preceding site to map similarities and note differences. Additional tools and new documentation may be produced at each pilot. It is expected that some tools and documentation would apply to all Fedora 3 users independent of their front-end application, and others would be unique to a specific front-end application. Tools and documentation for all pilots will be provided under open source or Creative Commons attribution licenses as appropriate. They will be disseminated through the Fedora GitHub and Confluence wiki spaces, and sustained and supported as part of the Fedora community. Project results will be shared via 2 conference presentations, 3 free webinars, and announcements to Fedora communication channels.

National Impact: Bridging the gap between the widely used, but unsupported, Fedora 3 and the new Fedora 6 is essential to safeguarding the digital heritage entrusted to the Fedora community. The deliverables from this project will provide resources and support for hundreds of libraries and archives using Fedora to deliver scholarly publications, research data, and cultural heritage content and services to patrons.

Budget: The total budget and amount requested from IMLS is $248,500. Funds will support salaries for a 50% PI, 50% developer, and 25% data migration expert for 18 months (77%, $190,685); fringe benefits (19%, $47,675); and travel for the planning meeting, 2 site visits and 2 conference presentation (4%, $10,140).