

Fedora and Fedora Commons

Fedora Project (Start Date: 2001)

Fedora Commons, Inc. (Start Date: May 2007)

URL: www.fedora-commons.org

Project Goals:

The Fedora Project has been a joint project of Cornell University and the University of Virginia since 2001. In December 2007, the final phase of the joint project was completed with the delivery of the beta release of Fedora 3.0. The Fedora Project provides open source software for digital repositories and related services. In 2008, the Fedora Project continues at Cornell University with support from the Mellon Foundation. The University of Virginia continues to contribute to Fedora, now in the role of a Fedora community member. The overall Fedora effort is now under the governance of the new Fedora Commons non-profit organization (which was incorporated in May 2007 and received startup funding from the Gordon and Betty Moore Foundation in July 2007).

The Fedora software has become an attractive open-source option for those who have embarked on building service-oriented platforms for scholarly communication, e-research, advanced digital libraries, and more. With funding from Mellon, Cornell University is currently evolving the original Fedora software to achieve “enterprise-grade.” The goal of this work is to provide an improved repository service and to enable new service integrations in accordance with the principles of service-oriented architecture. This work is motivated by use cases that suggest new models of scientific and scholarly communication.

From the Web perspective, the Fedora repository service has been shown to be an effective underpinning for web applications (including traditional web apps, Web 2.0 style apps, and increasing Semantic Web style apps). The new Fedora Commons non-profit organization has an over-arching goal of providing key technologies in open source to enable the building of information systems built to provide access to “durable, enduring, and re-usable” digital content. The goal is to achieve this by creating open source solutions that straddle both the enterprise system and Web paradigms.

Under the Fedora Commons umbrella, software development has been divided into multiple projects in order to provide better management and tracking. This is an evolving process as the organization and its software engineering processes evolve. The Fedora Repository Project will continue development of the Fedora distribution which includes the Fedora Repository Service and several closely related components, services, and libraries. Over time this project will be divided into smaller projects to improve integration options, simplify feature development and facilitate management flexibility.

The Fedora Repository and related components/services will evolve in conjunction with requirements from a key set of strategic partners that include:

- Johns Hopkins University and Cornell University focusing on data curation and integration of datasets with publications
- Topaz and the Public Library of Science focusing on open access publishing and semantic technologies
- Sun Microsystems, BNF/Atos, and others focused on preservation and archiving
- NSDL focusing on Web 2.0 and semantic digital libraries
- OAI Object Reuse and Exchange focusing on re-use and interoperability
- University of Virginia and FIZ Karlsruhe focusing on scholarly “workbenches”
- Missouri Botanical Gardens and Biodiversity Heritage Library focusing on digital asset management and historic journals in support of scholarly/scientific communication

Augmenting the development of the core Fedora software services/components, the Fedora Enterprise Middleware Project will provide reference integrations of the Fedora Repository with middleware products (provided in open source) which are suitable for use in solutions that include Fedora software. Middleware products for consideration and use include messaging products (i.e., JMS providers), Enterprise Service Bus products, Workflow/Business Process Execution Engines, and Distributed Transaction Managers. No middleware products will be built as part of this project though a select group of best-of-breed integrations will be demonstrated with the Fedora Repository Service. Both simple, lightweight approaches and enterprise-level approaches will be included as part the project. Also, approaches to management of business semantics, message formats, and governance issues will be considered.

Leading Institutions:

Cornell University

Fedora Repository Project

Fedora Enterprise Middleware Project

Sandy Payette, Principal Investigator

Dan Davis, Enterprise Architect

Aaron Birkland, Software Developer (50% FTE)

Jim Blake, Software Developer (50% FTE)

Fedora Commons, Inc.

Non-Profit Organization ; 501(c)(3)

Sandy Payette, Executive Director

Thornton Staples, Director Community Strategy

Carol Minton Morris, Director Communication

Dan Davis, Chief Software Architect

Chris Wilper, Lead Software Developer

Bill Branan, Sr. Software Developer

Eddie Shin, Sr. Software Developer

Carissa Smith, Exec Assistant and Website

Milestones and Deliverables:

Last 12 months

1. Start-up of new Fedora Commons non-profit organization.
2. Fedora Repository Project
 - Fedora Repository Service (3.0 Beta)
 - Content Model Architecture – An initial implementation that provides a uniform approach to identifying classes of digital objects and for building model-managed content systems. The CMA replaces the previous Fedora dissemination architecture.
 - Migration Utilities – Implements a set of utilities to aid in upgrading collections to Fedora 3.0. The utilities support converting old disseminators to use the CMA and upgrading object to FOXML 1.1.
 - Fedora REST API – An experimental interface that provides the ability to ingest and modify content using a Web interface without use of SOAP. Contributed by MediaShelf.
 - Mulgara Support – Provides support for using the Mulgara RDF triplestore for deployment of the Resource Index (graph-based index of Fedora repository exposing object-to-object relationships and other semantics).
 - Codebase Simplification – With the introduction of CMA, the most complex of Fedora’s relational database tables are no longer necessary. Essentially, the only use of a relational database in the Fedora repository is for the registration of digital objects. This reduces the complexity of the code while also reducing the amount of time Fedora needs to add, change, and remove information
 - Relationships API – The Fedora Management Interface has been extended to support the per-object assertion and retrieval of user-defined relationships in the RELS-EXT datastream. The new relationships API methods are addRelationship, getRelationships, and purgeRelationship. The Resource Index does not need to be enabled in order to use these methods.
 - Support for Sun ST5800 Storage – Sun has contributed a Fedora Low Level Storage (LLStore) plug-in to provide immediate compatibility for the Sun ST5800 archival storage system. Engineering work has begun to design updates and refinements that will be developed later to take advantage of the ST5800 capabilities.
 - Many other enhancements and new features (see release notes for Fedora 3.0)
 - Multiple Owners Per Digital Object – By specifying a comma-separated list, you can now specify multiple owners as the “ownerID” of Fedora objects. The list of owners can be provided as input to XACML policies to support more flexible policy enforcement.
3. Fedora Enterprise Middleware Project

- Participating member of Mellon sponsored Enterprise Service Bus (ESB) Evaluation Project
- Simple Repository JMS Messaging (Alpha) – Introduces JMS messaging capability to Fedora for the first time. This implementation consists of using Apache ActiveMQ as the reference JMS provided though coding will utilize strict JMS standard message types which should work with any provider. For this first use, the Fedora Repository service will act as a publisher of significant events including “object ingest,” “object modified,” and “object purged.” Remote services may subscribe to the messages (see JMS Message Interface for GSearch).
- Standardized Transaction Support in Fedora Repository Service: This work is being done to ensure that Fedora’s approach to transaction support is standards-based and improve how the repository fits within larger systems that require robust transaction support.
 - a. Completed analysis on technical requirements for making the Fedora repository service compliant with XA and the Java Transaction Architecture (JTA).
 - b. Prototype developed for a new Fedora Low Level Storage (LLStore) plug-in providing a transactional file system (JBlobStore). Used as input for the new Akubra storage project described below.

Next 12 months

1. Fedora Repository Project:

- Fedora 3.0 Final (final production version of all Fedora 3.0 Beta features)
- Pluggable, Multiplexing Storage: As a result of the analysis and prototyping completed in the last 12 months for standardized transaction support, the Fedora development team kicked off the new Akubra project (a new sub-project of Fedora Commons). Akubra will create a new plug-in architecture for the Fedora Repository’s low-level storage layer (LLStore). This work will enable plug-in of multiple, heterogeneous storage systems with Fedora and also to improve transactional capabilities across all configured datastores. This is very significant in terms of enabling multiple stores to be configured with Fedora (with different stores optimized for different types of content, including datasets and very large images). Among the initial stores we intend to demonstrate in the next 12 months using the new storage plug-in approach are: Sun Honeycomb, Internet Archive Petabox, and the Akubra basic transactional file system.
- Digital object validation modules (pluggable)
- Features and enhancements driven by 2008 strategic partner use cases:
 - to support data curation and data archives
 - to support OAI-ORE compliance
 - to support preservation and archiving
 - to support more powerful use of semantic technologies

2. Fedora Enterprise Middleware Project

- Fedora Repository as JMS Provider (Final production release in Fedora 3.0).
- JMS Message Interface for GSearch – A JMS message subscriber will be implemented to permit GSearch to receive notification of events published by the Fedora Repository. If the events match criteria, GSearch will take appropriate action to inform the search engine (e.g. index content). The GSearch message subscriber will provide a reference implementation that can be adapted for other uses.
- Select open source workflow engine and open source ESB for reference middleware integrations
- Provide reference workflows for simple and complex ingests related to:
 - Digital asset management (with Cornell and Missouri Botanical)
 - scholarly/scientific communication
 - data curation
- Extended Scalability Testing – Perform vertical scalability testing (single instance) on the Fedora repository beginning with duplicating previous testing methods as a control. Sun and Fiz Karlsruhe will be contributing extended scalability testing and contributing to a bottleneck analysis. Improvements will be made to the code from the bottleneck analysis to improve vertical scalability.

Community:

The uptake of the Fedora open source software is clearly accelerating, with a wide variety of types of institutions reporting installations of Fedora. With the startup of Fedora Commons, we have begun a systematic analysis of our existing user community.

At this writing, we have confirmed 104 installations of Fedora, the complete list of which can be seen at http://fedora.info/wiki/index.php/Fedora_Commons_Community_Registry. We are confident that the actual list of installations is much longer, given that we have had over 19,000 downloads of the software in the last 12 months. The Outreach and Communication effort led by Thornton Staples and Carol Minton Morris will continue to pursue leads and gather new information about who is using Fedora. The data we have gathered already provides an interesting picture of the community.

The current *reported* installations of Fedora fall into the following categories (types of institutions and the number in each category):

University Libraries and Archives: 45
 National Libraries and Archives: 13
 Corporations: 9
 Research Groups and Projects: 8
 Consortia: 5
 Publishing: 4
 Medical Centers and Medical Libraries: 4
 Semantic and Virtual Library Projects: 4
 University Information Technology: 3
 Museums and Botanic Gardens: 2
 Professional Societies: 2

Government agencies: 1

Our series of community-organized user group meetings continues and the number of meetings and locations is increasing. In the last year we had the following user group meetings:

02/07: Fedora User Group at Open Repositories 2007, San Antonio, TX

11/07: UK User Group, London, UK

The lineup for 2008 User Group meetings:

02/08: Australian User Group, Monash University

04/08: Fedora User Group at Open Repositories 2008, Southampton, UK

07/08: Fedora User Group at Swedish National Library, Stockholm

08/08: NE United States Fedora User Group at Yale University, New Haven

08/08: UK-Ireland Users' Group, Dublin

In his new role of Director of Community Strategy and Outreach with Fedora Commons, Thornton Staples will begin organizing a series of community councils in 2008 whose mission will be to define new requirements that can inform the Fedora Commons software development process.

Sustainability:

The new Fedora Commons non-profit organization was incorporated on May 22, 2007 and received its advanced ruling as a 501(c)(3) charitable organization in August 2007. Fedora Commons is the new sustainability organization for the Fedora Project. It will coordinate all development of the core Fedora software, and also incubate other complementary or synergistic open source projects. Fedora Commons has both a community outreach focus and a software development focus.

Fedora Commons secured startup funding from the Gordon and Betty Moore Foundation in the form of a four-year grant totaling \$4.9 million. This grant provides essential funding to build up the new non-profit organization and make it self-sustaining beyond the influx of Moore startup funding. Fedora Commons has begun work on developing a multi-pronged business strategy to ensure long-term sustainability to ensure sources of funding and strong community participation *beyond* the four-year startup period.

Currently, both the Executive Director and the Director of Community Strategy are working with the Fedora Commons Board of Directors and the Ithaka Foundation to explore new revenue-generating strategies. The challenge is to identify strategies that will be viable for open source software organizations that support the special types of domains we serve (e.g., scholarly/scientific communication, open access, preservation/archiving, higher education). We look at larger open source organizations with providing core infrastructure such as Apache, and we note that we are both similar and different. What will be right for our type of organization is likely to require some fresh ideas and require some creativity around new opportunities.

To review a few on the initial plans for supporting sustainability, we note the following. The Moore grant was intended to fund innovative development of Fedora Commons technologies with the goal of attracting new users of Fedora. Extending the uptake the Fedora software is an important prong in the sustainability strategy. Evolving the software to fit extremely well as a “sweet spot” between enterprise systems and the Web is an important niche to fill, and we believe our positioning here will strengthen the Fedora Commons brand. Aligning with other complementary and synergistic open source projects is another prong in the Fedora Commons strategy. For example, Fedora Commons is currently incubating the Topaz Project and is financially supporting a developer on the Mulgara Project. The Topaz and Mulgara projects are described later.

In addition to innovation in software development, another important part of the Fedora Commons strategy is focusing on organizing and enabling the community of developers and users growing around the Fedora Commons mission. The new Moore grant funds key roles in the Fedora Commons organization that will be focused on community outreach, marketing, and communication. Another aspect of our sustainability strategy is building strong synergistic relationships with companies (vendors, system integrations, consulting firms). For example, we are currently building new partnerships with Sun Microsystems (around preservation and archiving) and Atos Origin (around system integrations with Fedora for national libraries, initially in the context of the new BNF digital archive system).

During 2008 we will provide the community with more information on our strategic business plan.

Marketing/Evangelism:

Fedora Commons supports two employees who are focused out communication/outreach and marketing. The Director of Community Strategy/Outreach and the Director of Communications are taking a very active role contacting members of the Fedora community to understand current perceptions of the Fedora “value proposition.” Another dimension of this is working with consultants at the Ithaka Foundation to explore new market potential for Fedora Commons. Overall this is part of a larger effort to understand the market segmentation of Fedora Commons, and to build our outreach plan accordingly.

Also, Fedora Commons has taken a number of actions over the past 12 months aimed at improving communication about Fedora Commons. We have completed the following in the past 12 months:

- Completely new website for Fedora Commons
- Promotional video for Fedora Commons
- Brochures for Fedora Commons
- New logos for Fedora Commons
- Regular “News” headlines on email and website
- Branding campaign
- Fedora Commons newsletter (premiering March 4th 2008)

- Organizing of Fedora User group meetings
- Executive Director on the road giving presentations on Fedora Commons vision and plans

We are currently completing our plan for 2008 in which we will continue our outreach and marketing. This plan is being informed by the market studies that we are currently concluding.

Synergy with other projects:

OAI Object Reuse and Exchange

Fedora Commons has a close relationship with the OAI-ORE effort, with Sandy Payette being a member of the ORE Advisory Committee and Liaison Group. In 2008, Fedora Commons plans on developing components to facilitate exposure of Fedora repositories using ORE data model and protocols.

Kuali Rice

Kuali Rice provides core enterprise middleware components in open source. We have been meeting with Kuali architects to plan new integrations of the Fedora Repository Service with Rice components to support deliverables of the Fedora Enterprise Middleware project. Kuali at Cornell already has done initial work in integrating the Fedora repository service to support asset management (particularly document storage).

Topaz Project

The Topaz Project (<http://www.topazproject.org>) is an independently managed software development effort producing technologies used to construct the PLoS ONE Open Access Publishing application. The Topaz project is being incubated by Fedora Commons. Topaz has its own software base which uses the Fedora Repository but is otherwise independent. Fedora Commons is working with the Topaz Project to develop new interfaces and plug-in modules to connect XA/JTA-compliant data stores into both the Fedora Repository Service and the Topaz application platform. Topaz also provides components such as the OTM (Object to Triple Mapper) that is analogous to Hibernate for relational databases (where OTM maps triples in a triplestore to java business objects). In 2008, the Fedora/Topaz collaboration will focus on developing open source software components that enable integration of heterogeneous data/content stores and better integration of semantic triplestores with these data/content stores.

Mulgara Project

Mulgara (<http://www.mulgara.org>) is one of the premier RDF databases (semantic triple-stores) available in open-source. Fedora Commons is supporting one full-time software architect/developer on the Mulgara project and hosting their project resources in conjunction with the Topaz project. Practical introduction of semantic technologies is key to realizing the full potential of the Fedora Commons mission and ensuring that a production-ready, supported triple-store is available as free, open-source software is a required component for many applications we support.

NSDL NCore

NCore is an open source architecture and software platform for creating flexible, collaborative digital libraries. NCore was developed by the National Science Digital Library (NSDL) project, and it serves as the central technical infrastructure for NSDL. NCore consists of a central Fedora-based digital repository, a specific data model, an API, and a set of backend services and frontend tools that create a new model for collaborative, contributory digital libraries.

VUE (Visual Understanding Environment)

The Visual Understanding Environment was integrated with Fedora several years ago, and is credited with developing a Fedora-to-OKI adaptor in open source. As the VUE project continues to evolve, we look for more opportunities work together to provide new solutions to support scholarly communication and teaching/learning.