

# Fedora

## Digital Repositories and Digital Curation

Digital repositories play a vital role in the curation of digital materials and offer a convenient way to store, manage, reuse and curate a variety of digital materials. The term 'digital repository' can be applied to a number of different digital storage initiatives, which are often also referred to as 'institutional repositories', 'digital archives', or 'digital libraries' although in practice these each have slightly different functionality and underlying philosophies.

A growing number of repository models and systems are available and used by a variety of communities. They can take many forms and carry out many different functions. This technology watch paper provides an introduction to the features and functionality of the Fedora digital object repository management system.

## Fedora

Fedora stands for Flexible Extensible Digital Object and Repository Architecture. It is a general-purpose open source digital object repository management system for managing and delivering digital content. Jointly developed by Cornell University and the University of Virginia since 1999, Fedora can manage multiple digital object types within a single implementation, including complex, compound and dynamic objects. It is used in a range of repository scenarios, from digital asset management and institutional repositories to digital libraries.

Fedora runs as a service within an Apache Web Server with Tomcat. The Fedora download package contains all of the required software components, although Sun's Java Software Development Kit must be available on the machine running the Fedora server. The server is backed in part by a relational database; alternatively, the system can be configured to work with MySQL, v3.23.x, MySQL 4.x, or Oracle 9i. Several additional interface applications are available that have been developed either by the Fedora community or commercial organisations. The system is highly scalable and can provide support for upwards of 10 million objects. Fedora is written in Java and will run on either Unix, Windows, or Mac OS. Installation and management documentary support is available on the Fedora website, wiki and mailing list, supplemented by resources emerging from the Fedora user community.

The Fedora repository system is available under the Educational Community License. This permits widespread use and re-use of the software for any purpose, on the provision that changes are clearly documented in any new distribution. The Fedora repository system should not be confused with the Fedora Core operating system sponsored by RedHat.

---

## Functionality

One of the main principles behind Fedora is that interoperability and extensibility are best achieved through a distinct and modular separation of data, interfaces, and applications. Comprehensive, extensible, and flexible data storage and basic management is thus the focus of the core Fedora distribution, which functions as a generic repository foundation. Different client and end user interface applications can be installed and integrated with the core distribution to provide enhanced functionality and user services. Functionality of any given implementation will therefore depend on the additional functionality provided by the client applications. A single application can provide multiple functions across different aspects of the repository, including for example, ingest, workflow, and user access.

Development of the software is ongoing. This paper refers to Fedora repository software V 2.1.1.

**Ingest.** The Fedora core distribution (also sometimes referred to as 'out-of-the-box Fedora') includes the necessary basic software tools for administrators to ingest objects in either the Fedora Object XML (FOXML) format or the Metadata Encoding and Transmission Standard (METS). Specific details of the ingest process using front-end user interface applications depend on which application is installed. Applications currently listed on the Fedora Tools web pages include:

- VALET for ETD's, with configurable metadata entry, staged submission processes, and authorisation of edit, delete, or approval for submitted content prior to ingest;
- FEZ, with highly configurable web interface screens, automatic preservation metadata extraction, and controlled vocabularies;
- ELATED, a lightweight application that supports, amongst other things, Dublin Core metadata entry and search;
- NSDL CMS, a sample PHP application for uploading content into a Fedora repository with metadata entry templates. After further development, this application will be re-launched under the name FIRE.

**Workflow.** Workflow support is not integrated into the basic repository system and requires a separate application service. Workflow functionality is addressed by some of the user interfaces described above, including FEZ and ELATED.

**Storage.** Uploaded content is stored as part of a digital object, which can combine any number and variety of data stored as datastreams. Each digital object also includes metadata about the object and contents, as well as metadata to identify relationships between objects. Objects include an audit trail to record changes to the contents. The exact composition of a digital object varies according to the type of data stored and can be complex; readers are referred to the project website for more information on digital object composition. Non-standardised unique identifiers are allocated by the system in the core distribution. Client applications such as VTLS OSC can be used to create standard handles.

**Preservation.** Fedora incorporates a number of features that support preservation, including use of XML and open standards such as SOAP and METS. In conjunction with use of standards, the separation of data from interfaces and applications in the system architecture facilitates independent migration of the data objects over time. As the system is extensible and flexible, it will cater for the addition or inclusion of any designated metadata, including preservation metadata. The system architecture can easily be mapped against OAIS information package types, and a technology watch to inform of obsolete formats is planned. Other preservation services can be added using client applications.

**Access.** Fedora supports the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Client applications provide end user access over the Internet. Search functionality depends on the application, although basic search functionality is available in the core distribution.

## Selected Implementations

Digital Case [<http://library.case.edu/digitalcase/>], Case Western Reserve University — an institutional repository providing access to scholarly outputs from the University.

National Science Digital Library [<http://nsdl.org/>] — US national online library for education and research in science, technology, engineering and mathematics.

National Library of Wales/Llyfrgell Genedlaethol Cymru — Digital Asset Management System.

Encyclopaedia of Chicago [<http://www.encyclopedia.chicagohistory.org/>], Chicago Historical Society and Northwestern University — repository objects disseminated via the Chicago History website.

Digital Peer Publishing (DiPP) [<http://www.dipp.nrw.de/>] — an Open Access initiative for eJournals.

## Additional Resources

- Fedora website & wiki [<http://www.fedora.info/>]
- Fedora UK and Ireland user group mailing list: FEDORA-UKI@jiscmail.ac.uk
- Commercial supported version of FEDORA with built-in workflow for adding and discovery of objects — VITAL [<http://www.vtls.com/Products/vital.shtml>] from VTLS Inc.
- Open Society Institute: A Guide to Institutional Repository Software (2004) [<http://www.soros.org/openaccess/software/>]
- Johnston, Leslie. Development and Assessment of a Public Discovery and Delivery Interface for a Fedora Repository (2005) [<http://www.dlib.org/dlib/october05/johnston/10johnston.html>]
- CPIT Technical Evaluation of selected Open Source Repository Solutions (2006) [<https://eduforge.org/docman/view.php/131/1062/Repository%20Evaluation%20Document.pdf>]
- Payette, Sandra & Staples, Thornton. The Mellon Fedora Project — Digital Library Architecture Meets XML and Web Services (2000)
- RepoMMan project automating metadata and workflow for Fedora [<http://www.hull.ac.uk/esig/repomman/>]
- Fedora and the Preservation of University Records Project [<http://dca.tufts.edu/features/nhprc/index.html>]