

# Fedora and the Preservation of University Records Project

## 1.4 Glossary

**Version**

1.0

**Date**

September 2006

**Digital Collections and Archives, Tufts University  
Manuscripts & Archives, Yale University**

An Electronic Record Research Grant funded by the  
National Historical Publications and Records Commission

Digital Collections and Archives  
Tisch Library Building  
Tufts University  
Medford, Massachusetts 02155  
<http://dca.tufts.edu>

Manuscripts and Archives  
Yale University Library  
Yale University  
P.O. Box 208240  
New Haven, Connecticut 06520-8240  
<http://www.library.yale.edu/mssa/>

© 2006 Tufts University and Yale University

Co-Principle Investigators  
Kevin Glick, Yale University  
Eliot Wilczek, Tufts University

Project Analyst  
Robert Dockins, Tufts University

This document is available online at  
[http://dl.tufts.edu/view\\_pdf.jsp?urn=tufts:central:dca:UA069:UA069.004.001.00004](http://dl.tufts.edu/view_pdf.jsp?urn=tufts:central:dca:UA069:UA069.004.001.00004)  
(September 2006)

Fedora and the Preservation of University Records Project Website at  
<http://dca.tufts.edu/features/nhprc/index.html>

Funded by the  
National Historical Publications and Records Commission  
Grant Number 2004-083

# Fedora and the Preservation of University Records Project

## **PART ONE: INTRODUCTION**

- 1.1 Project Overview
- 1.2 System Model
- 1.3 Concerns

## **1.4 Glossary**

- 1.5 Requirements for Trustworthy Recordkeeping Systems and the Preservation of Electronic Records in a University Setting

## **PART TWO: INGEST**

- 2.1 Ingest Guide
- 2.2 Ingest Projects
- 2.3 Ingest Tools

## **PART THREE: MAINTAIN**

- 3.1 Maintain Guide
- 3.2 Checklist of Fedora's Ability to Support Maintain Activities

## **PART FOUR: FINDINGS**

- 4.1 Analysis of Fedora's Ability to Support Preservation Activities
- 4.2 Conclusions and Future Directions



**TABLE OF CONTENTS**

**Glossary Guide ..... 1**  
**Terms ..... 2**



---

## GLOSSARY GUIDE

This glossary has definitions of terms used throughout the project documents. To the degree possible, the project team has tried not create its own definitions of key terms when sufficient definitions could be found in an existing glossary or dictionary. Because there is a wide range of stakeholders involved in electronic records preservation, the project team has utilized the terminology from the OIAS Reference Model as much as possible. The Reference Model is aimed at a range of different stakeholder communities that are involved in the preservation of information and records in digital form. The OAIS terms do not exactly match all of those terms familiar to any particular discipline (e.g., traditional archives, digital libraries, science data centers). Instead, OAIS uses terms that are not already overloaded with meaning to reduce conveying unintended meanings across stakeholder communities.

Terms citing the Tufts-Yale project as its source are developed by the project team.

Each term in this glossary has four elements:

**Term:** the term used in the project documents.

**Definition:** the standard definition utilized by the project team.

**Source:** the source of the standard definition.

**Project Reports:** the primary reports that use the term.

The primary sources for the definitions are:

ISO 14721:2003, Space data and information transfer systems -- Open Archival Information System -- Reference model. <<http://public.ccsds.org/publications/archive/650x0b1.pdf>>.

Cited as: *ISO 14721:2003*

InterPARES2 Terminology Database <[http://www.interpares.org/ip2/ip2\\_terminology\\_db.cfm](http://www.interpares.org/ip2/ip2_terminology_db.cfm)>.

Cited as: *InterPARES2*

Richard Pearce-Moses, *A Glossary of Archival and Records Terminology* (Archival Fundamentals Series II) Chicago: Society of American Archivists, 2005.

<<http://www.archivists.org/glossary/>>

Cited as: *SAA Glossary*

Oxford Univeristy Press, *Oxford English Dictionary*, 3<sup>rd</sup> edition, 2006. <<http://www.oed.com>>

Wikipedia: The Free Encyclopedia <[http://en.wikipedia.org/wiki/Main\\_Page](http://en.wikipedia.org/wiki/Main_Page)>

Cited as: *Wikipedia*

Computer Desktop Encyclopedia. Computer Language Company Inc., 2006. Answers.com.

<<http://www.answers.com/topic/infrastructure>>

Cited as: *Computer Encyclopedia*

### TERMS

#### **Access**

The OAIS entity that provides the services and functions that support Consumers in determining the existence, description, location and availability of information stored in the OAIS, and allows Consumers to request and receive information products. Access functions include communicating with Consumers to receive requests, applying controls to limit access to specially protected information, coordinating the execution of requests to successful completion, generating responses (Dissemination Information Packages, result sets, reports) and delivering the responses to Consumers.

*ISO 14721:2003*

Requirements for Recordkeeping and Preservation

#### **Administration**

The OAIS entity that provides the services and functions for the overall operation of the archive system. Administration functions include soliciting and negotiating submission agreements with Producers, auditing submissions to ensure that they meet archive standards, and maintaining configuration management of system hardware and software. It also provides system engineering functions to monitor and improve Archive operations, and to inventory, report on, and migrate/update the contents of the Archive. It is also responsible for establishing and maintaining archive standards and policies, providing customer support, and activating stored requests.

*ISO 14721:2003*

Maintain Guide

#### **Administration Metadata Store**

A logical storage area where AIPs and PDI are stored. This is generally a hard disk in the Preservation Application Hardware Environment, and represents a fairly high per-unit storage cost.

*ISO 14721:2003*

Maintain Guide

#### **Application**

(see Recordkeeping Application and Preservation Application)

#### **Archival Information Package (AIP)**

An aggregation of records components and metadata, or information package, consisting of the Content Information and the associated Preservation Description Information (PDI), which is preserved within an Archive.

*ISO 14721:2003*

Maintain Guide

#### **Archival Storage**

The OAIS entity that contains the services and functions used for the storage and retrieval of Archival Information Packages (AIP). Archival Storage functions include receiving AIPs from Ingest and adding them to permanent storage, managing the storage hierarchy, refreshing the

media on which archive holdings are stored, performing routine and special error checking, providing disaster recovery capabilities, and providing AIPs to Access to fulfill orders.

*ISO 14721:2003*

Maintain Guide

### **Archive**

An organization of people and systems that has accepted the responsibility to preserve information and make it available for a Designated Community. It meets a set of responsibilities defined in ISO 14721:2003. Such an Archive may be distinguished from other uses of the term archive or archives.

*ISO 14721:2003*

Maintain Guide

### **Authenticity**

Authenticity is the trustworthiness of the record as a record—that the record is what it purports to be and has not been tampered with or corrupted in essential respects. A person cannot automatically presume the authenticity of an electronic record; he or she must weigh the evidence that the record either is or is not what it purports to be and either has or has not been modified or corrupted in essential respects—and then judge whether the record is authentic or not.

*Based in part on InterPARES2*

Concerns

### **Checksums**

A checksum is a form of redundancy check, a very simple measure for protecting the integrity of data by detecting errors in data that is sent through space (telecommunications) or time (storage). It works by adding up the basic components of a message, typically the asserted bits, and storing the resulting value. Later, anyone can perform the same operation on the data, compare the result to the checksum, and (assuming that the sums match) conclude that the message was probably not corrupted. There are more sophisticated types of integrity or redundancy checks, including Fletcher's checksum, Adler-32, cyclic redundancy checks (CRCs), and cryptographic hash functions, such as SHA-256 are designed to address these weaknesses by considering not only the value of each byte but also its position. However, these integrity checks are commonly referred to as checksum in the reports of this project.

*Wikipedia*

Maintain Guide

### **Consumers**

The role played by those people, or client systems, who interact with the services of the Archive to find preserved records of interest and to access those records in detail. This can include other Archives, as well as people or systems within the same institution.

*ISO 14721:2003*

Requirements for Recordkeeping and Preservation; Maintain Guide

### **Cryptographic Checksum or Cryptographic Hash Function**

A hash function with certain additional security properties to make it suitable for use as a primitive in various information security applications, such as authentication and message integrity. A hash function takes a long string (or message) of any length as input and produces a fixed length string as output, sometimes termed a message digest or a digital fingerprint. It is less vulnerable to attack than Checksums and cyclic redundancy checks.

*Wikipedia*

Maintain Guide

### **Cyclic Redundancy Check**

A cyclic redundancy check (CRC) is a type of hash function used to produce a CRC checksum against a block of data, such as a packet of network traffic or a block of a computer file.

*Wikipedia*

Maintain Guide

### **Data Management**

The OAI entity that contains the services and functions for populating, maintaining, and accessing a wide variety of information. Some examples of this information are catalogs and inventories on what may be retrieved from Archival Storage, processing algorithms that may be run on retrieved data, Consumer access statistics, security controls, and schedules, policies, and procedures.

*ISO 14721:2003*

Maintain Guide

### **Designated Community**

An identified group of potential Consumers who should be able to understand a particular set of information. The Designated Community may be composed of multiple user communities.

*ISO 14721:2003*

Maintain Guide

### **Digital Object**

A unit of digital information that includes properties of the object and may also include methods of performing operations on the object. This is the basic unit for information aggregation in Fedora. In Fedora, a digital object has consist of at least two elements: (1) an identifier or PID, that provides the key by which the digital object is accessed from the repository, and (2) Dublin Core metadata that provides a basic description of the digital object.

*InterPARES2, Tufts-Yale*

Maintain Guide

### **Dissemination Information Packages (DIP)**

The aggregation of records components and metadata, or information package, derived from one or more AIPs, received by the Consumer in response to a request to the Archive.

*ISO 14721:2003*

Maintain Guide

**Hot Spare**

An extra disk drive in a RAID (redundant array of independent disks) configuration that is ready and waiting to be put into action automatically when another drive fails. Using the RAID algorithms, the missing data from the faulty drive is reconstructed and written to the hot spare. When the bad drive is replaced, it then becomes the hot spare. If a hot spare is not used, then the faulty drive must be manually removed and replaced with a new one.

*Wikipedia*

Maintain Guide

**Infrastructure**

The fundamental structure of a system or organization. The basic, fundamental architecture of any system (electronic, mechanical, social, political, etc.) determines how it functions and how flexible it is to meet future requirements; refers to system and development programs in contrast to applications. A computer system's infrastructure would include the operating system, database management system, communications protocols, compilers and other development tools.

*Computer Encyclopedia*

Requirements for Recordkeeping or Preservation

**Ingest**

The OAIS entity that provides the services and functions to accept Submission Information Packages (SIPs) from Producers (or from internal elements under Administration control) and prepare the contents for storage and management within the archive. Ingest functions include receiving SIPs, performing quality assurance on SIPs, generating an Archival Information Package (AIP) which complies with the Archive's data formatting and documentation standards, extracting Descriptive Information from the AIPs for inclusion in the archive database, and coordinating updates to Archival Storage and Data Management.

*ISO 14721:2003*

Maintain Guide

**Institution**

An establishment, organization, or association, instituted for the promotion of some object, e.g. a church, school, college, hospital, asylum, reformatory, mission, or the like; an established organization. An Institution creates and uses records to conduct its business. For this document it refers specifically to colleges and universities, although institutions in other industries may adopt this document.

*Tufts-Yale*

Requirements for Recordkeeping or Preservation

**Juridical System**

A collectivity organized on the basis of a system of rules; a social group that is organized on the basis of a system of rules and that includes three components: the social group, the organizational principle of the social group, and the system of binding rules recognized by the social group.

*InterPARES2*

Requirements for Recordkeeping and Preservation

### **Juridical Person**

An entity having the capacity or the potential to act legally and constituted either by a position (a succession) or collection (an organization) of natural persons; role(s) taken by natural people; e.g. university archivist or university archives.

*InterPARES2*

Requirements for Recordkeeping and Preservation

### **Knowledge Base**

A set of information, incorporated by a person or system, that allows that person or system to understand received information.

*ISO 14721:2003*

Maintain Guide

### **Natural Person**

An individual human being, as distinguished from a corporate body, representative, or juridical person.

*Tufts-Yale*

Requirements for Recordkeeping and Preservation

### **Policy**

An articulation of the goals and aims of the institution; a principle or course of action adopted or proposed as desirable, advantageous, or expedient; esp. one formally advocated by an institution or government. Policies exist as documents.

*Oxford English Dictionary*

Requirements for Recordkeeping and Preservation, Ingest Guide

### **Preservation**

The act of physically and intellectually protecting and technically stabilizing the transmission of the content and context of electronic records across space and time, in order to produce copies of those records that people can reasonably judge to be authentic.

*Based in part on SAA Glossary*

All documents

### **Preservation Application**

The software package (or collection of software) which an Archive uses as part of its preservation system. This is intended to cover only the software actively involved in the management of the preservation system and not utility or operating system software.

*Tufts-Yale*

Maintain Guide

### **Preservation Application Hardware Environment**

The hardware platform upon which the preservation application runs. This may be a single computer or a group of cooperating computers. Because of the high need for reliability, this hardware should be dedicated to the preservation application and should not participate in other functions. This hardware likely contains the Administrative Data Store primary media, probably in the form of hard disks. It may also contain the Records Components Store primary media.

*Tufts-Yale*  
Maintain Guide

### **Preservation Description Information**

The information which is necessary for adequate preservation of the Content Information and which can be categorized as Provenance, Reference, Fixity, and Context information.

*ISO 14721:2003*

Maintain Guide

### **Preservation Planning**

The OAIS entity that provides the services and functions for monitoring the environment of the OAIS and providing recommendations to ensure that the information stored in the OAIS remains accessible to the Designated User Community over the long term, even if the original computing environment becomes obsolete. Preservation Planning functions include evaluating the contents of the archive and periodically recommending archival information updates to migrate current archive holdings, developing recommendations for archive standards and policies, and monitoring changes in the technology environment and in the Designated Community's service requirements and Knowledge Base. Preservation Planning also designs information package templates and provides design assistance and review to specialize these templates into SIPs and AIPs for specific submissions. Preservation Planning also develops detailed Migration plans, software prototypes and test plans to enable implementation of Administration migration goals.

*ISO 14721:2003*

Maintain Guide

### **Procedure**

Procedures articulate the actions required to successfully complete a task. Procedures also describe how to execute those actions. Procedures articulate how a Juridical System will fulfill its policy requirements. Procedures exist as documents.

*Tufts-Yale*

Requirements for Recordkeeping and Preservation

### **Producer**

The role played by the people, corporate bodies, administrative units, families, or client systems, who provide the records to be preserved; equivalent to donor, often, but not always the records creator (the physical or juridical person who makes, receives, and/or accumulates records by reason of its mandate/mission, functions or activities).

*ISO 14721:2003*

Maintain Guide

### **Recordkeeping Application**

Any electronic program that creates and/or imports and maintains, stores, and distributes electronic records.

*Tufts-Yale*

Requirements for Recordkeeping and Preservation

### **Recordkeeping Component**

The people, institutions, applications, infrastructure, and procedures necessary for records to be created, collected, organized, and categorized to facilitate the records' preservation, retrieval, use, and disposition.

*InterPARES2*

Ingest Guide, Maintain Guide

### **Recordkeeping Infrastructure**

(See Infrastructure)

### **Recordkeeping Institution**

(See Institution)

### **Recordkeeping Natural Person**

(See Natural Person)

### **Recordkeeping Policy**

(See Policy)

### **Recordkeeping System**

A manual or automated system in which records are created, collected, organized, and categorized to facilitate their preservation, retrieval, use, and disposition. Recordkeeping Components compose a recordkeeping system.

*Tufts-Yale*

Requirements for Recordkeeping and Preservation

### **Records Component**

A digital object that is part of one or more electronic records, including any metadata necessary to order, structure, or manifest the content, requiring a given preservation action.

*InterPARES2*

Maintain Guide

### **Records Component Store**

A logical storage area where records components are stored. This would probably be a dedicated storage environment separate from the Preservation Application Hardware Environment, and may represent a lower per-unit storage cost than the Administration Metadata Store.

*Tufts-Yale*

Maintain Guide

### **Stasis**

Setting the Preservation Application in a state that allows only system administrators to make changes application settings and data.

*Tufts-Yale*

Maintain Guide

**Storage Hardware Environment**

The hardware platform that is responsible for the storage functions of the Administrative Metadata Store or the Records Components Store. This can be a network appliance, a storage virtualization network, a tape robot, or some other storage subsystem. The Storage Hardware Environment also includes any networking or software components needed for the storage function to operate. The Records Components Store is usually implemented using a Storage Hardware Environment separate from the Preservation Application Hardware Environment, whereas the Administrative Metadata Store may or may not be.

*ISO 14721:2003*

Maintain Guide

**Submission Information Packages**

The aggregation of records components and metadata, or information package that is delivered by the Producer to the OAIS for use in the construction of one or more AIPs.

*ISO 14721:2003*

Ingest Guide

**Trustworthiness**

The quality of being dependable and reliable.

*SAA Glossary*

Requirements for Recordkeeping and Preservation

**Trustworthy Electronic Recordkeeping System**

The combination of all the recordkeeping components (people, institutions, applications, infrastructure, and processes) necessary for records to be created, collected, organized, and categorized to facilitate their preservation, retrieval, use, and disposition in a manner that provides a circumstantial probability of the authenticity of those records.

*InterPARES2*

Requirements for Recordkeeping and Preservation