

Fedora and the Preservation of University Records  
Interim Narrative Report

NHPRC Grant #2004-083  
August 24, 2005

Submitted by:



---

Eliot Wilczek  
Digital Collections and Archives  
Tufts University



---

Kevin Glick  
Manuscripts & Archives  
Yale University

The Digital Collections and Archives (DCA), Tufts University, in conjunction with Manuscripts and Archives (MSSA) of Yale University Library, is working on a National Historical Publications and Records Commission (NHPRC) electronic records research grant project (2004-083) to synthesize electronic records preservation research with digital library repository research in an effort to develop systems capable of preserving university electronic records at both institutions. This project is testing the potential of Fedora (the Flexible Extensible Digital Object and Repository Architecture) <<http://www.fedora.info>> to serve as the architecture for such an electronic records preservation system. The project's original grant proposal narrative and the revised plan of work are on the project's website at: <<http://dca.tufts.edu/features/nhprc/reports/index.html>>.

Since January 2005 the project staff has continued to make significant progress on this research grant. The work during this calendar year featured the following activities:

- Writing the Step 1 document, "Requirements of a Trustworthy Recordkeeping System in a College or University Setting"
- Writing the Step 2 document, "Requirements of a Trustworthy Preservation System in a College or University Setting"
- Writing the Step 3-1 document, "Ingest Guide"
- Developing of an ingest system
- Developing of simple extraction tools used for ingest
- Undertaking ingests projects

The Tufts and Yale grant project teams also met to have two face-to-face meetings this year on March 29<sup>th</sup> and July 25<sup>th</sup>.

As we noted in our January 31, 2005 progress report, the work on the Steps 1 and 2 reports have taken longer than we had originally anticipated. For both steps we synthesized requirements from previous work on recordkeeping and preservation system requirements. Because the literature on preservation systems is not nearly as extensive as the literature on recordkeeping systems our Step 2 document became much more dependent on the Step 1 document than we had at first expected. We have completed drafts of the Steps 1 and 2 requirements and posted them on the project website at <[http://dca.tufts.edu/features/nhprc/reports/1\\_1\\_draftpublic.pdf](http://dca.tufts.edu/features/nhprc/reports/1_1_draftpublic.pdf)> for Step 1 and <[http://dca.tufts.edu/features/nhprc/reports/2\\_1\\_draftpublic.pdf](http://dca.tufts.edu/features/nhprc/reports/2_1_draftpublic.pdf)> for Step 2. In addition to

synthesizing recordkeeping requirements appropriate for a college or university setting, the Step 1 document also carefully defines “recordkeeping system” and describes its relationship to “records controls.” Because so many of the issues concerning recordkeeping and preservation systems overlap, the project staff is considering consolidating the Step 1 and 2 reports into a single report with separate sections for recordkeeping system requirements and preservation system requirements.

The project staff has made significant progress on the Step 3-1 Ingest Guide since January, recently posting a draft of the Guide on the project website at [http://dca.tufts.edu/features/nhprc/reports/3\\_1\\_draftpublic.pdf](http://dca.tufts.edu/features/nhprc/reports/3_1_draftpublic.pdf). Building primarily on the work of the Consultative Committee for Space Data System’s *Producer-Archive Interface Methodology Abstract Standard*, the Ingest Guide describes the actions needed for a trustworthy ingest process. This process enables an Archive and Producer to move records from a recordkeeping system to a preservation system in a manner that allows a presumption of authenticity.

This Guide contains two main sections. Section A, Negotiate Submission Agreement, deals with the Producer and the Archive creating and agreeing to a Submission Agreement that defines the terms and conditions of the transfer of records from a Producer to an Archive, detailing the scope of the records along with the nature of their validation and transformation. Section B, Transfer and Validation, deals with the actual transfer, validation, and transformation of records. Section A currently contains eight parts and Section B contains six parts. Each part includes a narrative summary, a flowchart illustrating all of its steps and a description of each step. Each description includes an overview, a list of Components, Resources, Products, and Documentation each step utilizes and/or produces, and crosswalks to the *Producer-Archive Interface Methodology Abstract Standard*. The Ingest Guide also includes a separate section on Components, Resources, Products, and Documentation that describes each of their roles in the ingest process and refers to the steps that use and produce them. The Ingest Guide also has a Submission Agreement section that explains the Agreement in further detail and includes a sample Agreement.

The project staff is currently making revisions to the Ingest Guide that will reorganize the order and composition of some of the parts in Section A and redefine some of the Components, Resources, Products, and Documentation. The staff plans to post the second draft of the Guide on its website by September 2005.

The Ingest Guide requires individual institutions to make a number of decisions about how to implement the procedures it describes. In order to gain experience with the Guide and validate it with real-world situations we have developed a prototype application, which conforms to Section B of the Guide. This application, which we call the Ingest System, is a Java program with a database backend and a web frontend. It defines a concrete format for SIPs and for Submission Agreements using XML and performs validation and transformation via scripts called object profiles. The Ingest System is able to accept SIPs, validate their content, accession the records therein, and finally submit them to a Fedora repository. We intend to use the Ingest System when performing the test Ingest projects of Step 3-3. Furthermore, we intend to release the Ingest System as an open source project so that others in the Fedora community can learn from and perhaps use our work.

Project staff are also developing a number of simple extraction tools that we plan to use during our test ingest projects. These tools are used for extracting records from the following environments:

- SharePoint Team Services
- LiveLink
- Eudora Mail
- CDs
- HTML webpages

The project team has had to make some adjustments to the Ingest projects it planned to run in Step 3-3 of the grant. Tufts is now undertaking three projects which include desktop application files stored in an instance of SharePoint Team Services; a website composed of html files, image files, and PDF files; and desktop application files stored on CDs. Yale is undertaking three projects which include Eudora email stored on LiveLink; desktop applicaton files stored on CDs;

and CAD files. To date, both Yale and Tufts have focused on the mechanics of actually getting these records out of their recordkeeping environments so they can be turned into SIPs. Once the team completes its current revision of the Ingest Guide we will undertake actual ingest projects with these six sets of records, putting the Guide through its paces.

The project staff has also begun work on Step 4-1, developing a guide for maintaining records in a preservation system. Once we have completed the Maintain Guide and the ingest projects we will use to the same set of records from the ingest projects to undertake six maintain projects to test the validity of the Maintain Guide.

This maintain work will turn out to be the part of the grant that truly focuses on Fedora's ability to act as a preservation system for electronic university or college records of archival value. The requirements for recordkeeping and preservation system are application-neutral, not applying specially to Fedora or any other recordkeeping or preservation application. The Ingest Guide is also application-neutral. In addition, when the project staff undertakes its six projects to ingest records into Fedora, most of this ingest work occurs outside of Fedora (in its current instantiation). Only the creation of the AIPs at the end of the ingest process is a Fedora-specific task. The project staff will undertake the six maintain projects, however, entirely within a Fedora environment. Based on this work, the staff will be able to measure Fedora's ability to maintain records against the requirements of the Maintain Guide.

Thornton Staples of the University of Virginia will consult with the project team at a September 14-15, 2005 meeting at Tufts University on Fedora's ability to maintain records. He will also talk with the team about how the project can contribute to Fedora's development and other Fedora-related projects.

There have been a number of staffing changes at Yale brought on by unforeseen circumstances. In the original grant proposal Yale promised 18 months of 35% of Kevin Glick's time and 25% of Stephen Yearl's time as shared costs. Instead, from July 1, 2004 to March 31, 2005 Mr. Glick devoted 35% time to the project, and Mr. Yearl and Raman Prasad each devoted 12.5% time to

the project. From April 1, 2005 to June 30, 2005 Mr. Glick devoted 60% time to the project. From July 1, 2005 to the end of the project Mr. Glick will be devoting 60% time to the project.

The project staff has given the following presentations on the project during this calendar year:

- Kevin Glick, Robert Dockins, and Eliot Wilczek, “Researching Fedora’s Ability to Serve as a Preservation System for Electronic University Records,” Fedora Users’ Conference, Rutgers University, May 14, 2005. The presentation focused on Requirements of a Trustworthy Electronic Recordkeeping System and Preservation System and the Ingest Guide.
- Kevin Glick and Eliot Wilczek, “Fedora and the Preservation of University Records,” ECURE: Preservation and Access for Digital College and University Resources, Arizona State University, March 1, 2005. The presentation focused on Requirements of a Trustworthy Electronic Recordkeeping System and Preservation System.
- Kevin Glick, “Fedora and Institutional Repositories,” Meeting of the Information Technology Interest Group and Conservation Interest Group of the ACRL New England Chapter, March 18, 2005. The presentation briefly described the grant project.
- Eliot Wilczek, “Helping Jumbo Remember: Fedora and the Tufts University Institutional Repository,” Institutional Repositories: Models and Approaches, a NELINET Seminar, January 27, 2005. The presentation briefly described the grant project.
- On September 7, 2005 Eliot Wilczek is scheduled to give a presentation on the grant project at the International Council on Archives, Section on University and Research Institution Archives Seminar at Michigan State University.