

Fedora and the Preservation of University Records
Interim Narrative Report

NHPRC Grant #2004-083
February 27, 2006

Submitted by:



Eliot Wilczek
Digital Collections and Archives
Tufts University



Kevin Glick
Manuscripts & Archives
Yale University

Project Activities July 1, 2005 – December 31, 2005

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 analyzing the potential of Fedora (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>>.

From July 1, 2005 to the end of 2005 the project staff has continued to make significant progress on this research grant. The work during this time period featured the following activities:

- Starting work on combining the Step 1 document, "Requirements of a Trustworthy Recordkeeping System in a College or University Setting" and the Step 2 document, "Requirements of a Trustworthy Preservation System in a College or University Setting" after seeing significant overlap with the two documents.
- Completing a draft for public comment version of the Step 3-1 document, "Ingest Guide." Available at: <http://dca.tufts.edu/features/nhprc/reports/3_1_draftpublic2.pdf>.
- Completing ingest projects
- Completing an ingest system framework
- Writing the Step 4-1 document, "Maintain Guide"

The Tufts and Yale grant project teams also met to have two face-to-face meetings on July 25th and on September 14th-15th. The September 14th-15th meeting included the consulting visit of Thornton Staples of the University of Virginia.

As we have noted in our January 31, 2005 and August 24, 2005 interim reports, the work on the Steps 1 and 2 reports have taken longer than we had originally anticipated. We continue to make revisions to both documents, particularly the Step 2 report. We also noted in our August 24th report that we were considering combining the Step 1 and 2 reports into a single document with

separate sections for recordkeeping system requirements and preservation system requirements. We have decided to undertake that course of action and have begun to undertake that work of consolidation.

The project team posted a draft of the Step 3-1 Ingest Guide on the project website at http://dca.tufts.edu/features/nhprc/reports/3_1_draftpublic.pdf in time for the Society of American Archivists annual conference held from August 14th-21st, 2005. The project staff subsequently made significant revisions to the Ingest Guide that culminated in a September 30, 2005 draft for public comment version. The project staff asked for comments on the draft guide on the Archives & Archivists, Electronic Records, and Fedora Users' listservs. We have received a small number of comments and we are now incorporating those suggestions into our final version of the Guide. Despite the major revisions we made in August and September, the Guide is still organized into two main sections. Section A. Negotiate Submission Agreement illustrates 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 and 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.

During the Fall we conducted six ingest projects testing the Ingest Guide, which is Step 3-3 of the grant. Tufts undertook 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 finishing three projects which include multiple preservation strategies, including a project to migrate aggregations of records created and stored in proprietary file formats like Qualcomm's Eudora email format (.mbx), to a non-proprietary format stored in a records management application (LiveLink), so that the records can be accessioned into a records preservation system (in this case Fedora).

The project team has completed a draft of the Step 4-1 Maintain Guide and is currently engaged in making revisions. The Maintain Guide is a prescriptive guide for an Archive to conduct a Trustworthy Maintain Electronic Records process. The Guide is composed of two parts, Scheduled Events and Irregular Events. Scheduled events may occur according to a pre-

determined schedule. Irregular events occur according to some external stimulus or as the result of the actions of a scheduled event.

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 project walkthroughs to test the validity of the Maintain Guide. Originally the project staff intended to undertake six maintain projects. However, the requirements we produced for the Maintain Guide are much more rigorous than we anticipated when we developed the Plan of Work. In order to follow the Guide, we would have to buy a substantial amount of hardware that we simply did not expect to need and in turn did not budget for. Rather than run actual maintain projects, we propose undertaking conceptual maintain walk-throughs. The main portion of this conceptual work will be a comparison of the Maintain Guide requirements with the features of Fedora 2.1—the current version of Fedora as of October 2005. We feel this work will still produce a valid gap analysis of Fedora’s capabilities against maintain requirements.

Thornton Staples of the University of Virginia visited Tufts University on September 14-15, 2005 to meet with the project team and consult on Fedora’s role in preserving electronic records with archival value. He submitted a report to the project team articulating how current and future developments of Fedora could meet the needs and requirements of electronic records preservation that the project has articulated. Highlights of his report include:

- The potential of using eXtensible Access Control Markup Language (XACML) to address concerns about access restrictions raised in the Ingest Guide.
- The potential of expressing the Ingest Guide in Business Process Execution Language (BPEL).
- Using content models—a concept in Fedora Community that defines the number and type of datastreams and the list of disseminators that are in a data object—to define Record Type Records, a resource defined in the Ingest Guide.
- Ways the grant project can contribute to the work of the Fedora Workflow Working Group and the Fedora Preservation Working Group.

Completion of the Grant Project

Robert Dockins left his project position of Fedora Project Analyst on December 1, 2005 to take another position at the Digital Collections and Archives at Tufts. This change in staff and other factors prompted the principal investigators to apply for a three-month no-cost grant extension, pushing the completion of the Grant from December 31, 2005 to March 31, 2006. The NHPRC accepted the extension application and its accompanying plan of work on January 3, 2006.

During the first three months of 2006 we plan to:

- Have Neil Guo and Roy Lechich continue their grant work at Yale.
- Hire someone with archival and editorial experience edit the project reports.
- Hire someone to create a sophisticated web version of the Ingest Guide.
- Hire an expert in the digital preservation and electronic records field to review and suggest necessary changes to our combined Step 1 and Step 2 documents on requirements for trustworthy recordkeeping and preservation systems.
- Hire someone to conduct further investigation on turning some of the resources articulated in the Step 3-1 Ingest Guide into actual tools.

Project Dissemination

Kevin Glick and Eliot Wilczek have become members of the Fedora Preservation Working Group. This grant project, particularly its production of the Maintain Guide, has strongly influenced the work of the Preservation Working Group and potentially the future development of Fedora.

The project staff has given the following presentations on the project between July 1 and December 31, 2005:

- Eliot Wilczek gave a brief update on the grant project at the Electronic Records Section meeting at the Society of American Archivists conference, New Orleans, August 19, 2005.
- Eliot Wilczek, "Preserving Electronic Records at Tufts University," International Council on Archives, Section on University and Research Institution Archives Seminar (ICA/SUV), Michigan State University, September 7, 2005. The presentation focused on the grant project's Requirements of a Trustworthy Electronic Recordkeeping System and Preservation System and the Ingest Guide.

The project staff anticipates giving the following presentations concerning the grant project in 2006:

- Eliot Wilczek, “Institutional Repositories: It’s Not Just the Technology,” New England Archivists Conference, Boston College, March 11, 2006. The presentation will include general findings of the grant project.
- Kevin Glick, Eliot Wilczek, and Robert Dockins, “The Ingest and Maintenance of Electronic Records: Moving from Theory to Practice,” Poster Presentation, Joint Conference on Digital Libraries, Chapel Hill, North Carolina, June 11-15, 2006. The presentation will focus on the grant project’s work on ingest and maintain.
- Robert Dockins, Fedora Users Conference, University of Virginia, June 19-20, 2006. The presentation will focus on Fedora’s ability to meet the requirements for preserving electronic records articulated by the grant project.
- Kevin Glick, ICA/SUV, Reykjavik, Iceland, September 13-20 2006. The presentation will review the general findings of the grant project.

In addition to the project findings and reports that will be posted on the project website at <http://dca.tufts.edu/features/nhprc>, the project staff anticipates publishing articles that give an overview of the project, discuss the Ingest Guide, and examine the concepts of Designated Community and Knowledge Base (based on work on the Maintain Guide) in such venues as *RLG DigiNews* and *D-Lib Magazine*.