Fedora and the Preservation of University Records Project

1.1 Project Overview

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:UA069.004.001.00001
(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

Project Overview **1**
Summary of Products **4**
Acknowledgements **7**

PROJECT OVERVIEW

This grant project, "Fedora and the Preservation of University Electronic Records," combines electronic records preservation research and theory with digital library practice to investigate three areas of research: requirements for trustworthy recordkeeping systems and preservation activities, ingesting records into a Preservation System, and maintaining records in a preservation system. The Digital Collections and Archives of Tufts University and Manuscripts and Archives of Yale University undertook this project with support of a National Historical Publications and Records Commission (NHPRC) electronic records research grant (grant number 2004-083).

The essential nature of the modern office at colleges and universities—complete with hybrid paper/electronic systems, digital environments established to support manipulation and repurposing of data at the expense of recordkeeping, obsolescence of hardware and software, media decay, the proprietary and idiosyncratic nature of applications, and other problems—makes it difficult for archivists to provide for the long-term preservation of authentic electronic records and maintain the accountability of the organizations and operations which those records are supposed to document. This nature of the modern office leads institutions to create and maintain electronic records that they cannot automatically trust and depend on in the same way that institutions trust and depend on traditional paper records. In general, archivists have difficulty preserving electronic records that fail to be (1) accessible, readable, or intelligible due to compatibility and obsolescence issues; (2) identifiable and retrievable due to an incongruence of classifications and/or taxonomies; and (3) reliable in the accuracy of their content due to the ease of updating and altering records, either inadvertently or purposefully.

In order to address these issues, an organization must recognize that the goal of electronic records preservation is to physically and intellectually protect and technically stabilize 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. This is a continuous process that begins even before the moment of records creation.

The Tufts-Yale Project contributes to a growing body of research on electronic records preservation by producing a dual set of functional requirements for recordkeeping systems and preservation activities, a guide defining the necessary steps for a trustworthy Ingest process, and a guide defining the necessary steps for a trustworthy maintain process.[1] Our research depends greatly on existing research and standards, particularly the Reference Model for an Open Archival System Information System (OAIS).[2] We have used the Reference Model as a conceptual framework guiding our research. Whenever possible, we have attempted to adopt OAIS terminology. Recorkeepers are Producers; the Archive is the entire preservation environment, including the university archives and all of its affiliated and supporting units; and the users or patrons of the archives are the Consumers. The Ingest Guide describes the Ingest

[1] The project produced a total of twelve reports that center on the requirements report and the two guides. For a list and brief description of the reports, see the "Summary of Products" section below.

[2] ISO 14721:2003*:* Space data and information transfer systems--Open archival information system--Reference model (Geneva: International Organization for Standardization, 2003). Available at <http://public.ccsds.org/publications/archive/650x0b1.pdf>.

function as well as much of Establish Standards and Policies, Audit Submission, and Negotiate Submission Agreement activities within the Administration function. The Maintain Guide covers the Data Management and Archival Storage function. The requirements for recordkeeping attempt to guide the activities of a Producer, while the requirements for preservation activities attempt to guide the activities of an Archive and thus cover every single functional area of the OAIS Reference Model.

One will be able to make the best use of these documents, resources, and services by understanding that the OAIS Reference Model, the two sets of functional requirements, the ingest and maintain guides, the resources and services that support the guides, and the implementation of the guides, are all parts of a set. All of the documents relate to and build upon each other. The OAIS Reference Model is the overarching conceptual structure for preservation activities and systems. Beneath OAIS sits a layer of requirements for preservation activities or systems such as our preservation requirements. These requirements add further articulation to OAIS by describing the attributes of preservers that fit within the context of the Reference Model. Beneath these requirements are conceptual guides or well-defined models, like the Ingest Guide and Maintain Guide, which translate requirements into actions for those functional areas of preservation. In turn resources and services—ideally, standardized and openly available—support the execution of the activities defined in the guides. Individual institutions will still have implementation decisions to make within the context of the guides, resources, and services. Institutions cannot simply take the guides and call them their procedures. This interconnectedness reinforces each level, giving context to the frameworks, requirements, guides, resources and services, and implementation decisions, helping to enable their intelligent utilization.

The Tufts-Yale Project was originally conceived of as a systems development/analysis project focused primarily on the Fedora repository system. Fedora was already being implemented by a number of universities for very interesting digital library repository projects.[3] The flexibility and extensibility of the Fedora architecture and object model and the modularity of the Fedora repository system led the project team to hypothesize that it could be extended to serve as an electronic records preservation system. Thus, the plan was to test the hypothesis by simply accessioning records into a Fedora repository and analyzing the results. In practice, this hypothesis was very difficult to test. There were very few evaluation criteria, and there was still much for the field of digital preservation to learn. In light of this, the Tufts-Yale Project was reconceived to combine electronic records preservation research and theory with that existing digital library research and practice. The project shifted much of its attention away from Fedora because a Fedora instance (or instances), serving as the repository core of a preservation system, would only be one part of an overall preservation environment. Significant portions of Ingest, Access, and Preservation Planning activities occur outside of any Fedora instance. Rather than an out-of-box, limited repository solution, Fedora is a repository architecture upon which an institution can shape a repository in many different ways. Thus, the suitability of Fedora as the basis of a preservation system depends significantly on its implementation.[4] The primary focus

[3] See Fedora "Community," <http://www.fedora.info/community/> for more information.
[4] Despite this shift in focus the Tufts-Yale Project does make assessments of Fedora's ability to support maintain and preservation activities respectively in 3.2 Checklist of Fedora's Ability to Support Maintain Activities" and 4.1 Analysis of Fedora's Ability to Support Preservation Activities, respectively.

of the research and its most significant contributions are a synthesis of digital library, preservation, electronic records, and archival literature and standards in order to present in depth guides to portions of the preservation process and explicit criteria with which to evaluate recordkeeping and preservation environments. These environments include not only the repository application, but also the wider context of people, policies, procedures, infrastructure, and the institution as a whole.

The Tufts-Yale Project is aimed at university archivists and focuses primarily on university records because the project team feels that the existing electronic records research has been aimed at government recordkeepers and because both research partners, whose primary responsibility is to preserve university records, are based at universities. However, the findings of this project are not particularly university-specific and are applicable to the management and preservation of electronic records in most industries. In addition, most of the research could benefit libraries or archives dealing with the preservation of any type of digital object.[5]

[5] The strict adherence authenticity and the characteristics of trustworthy systems may be much less important for preservers dealing with non-records, leaving some of the project's instructions to seem cumbersome, time-consuming, or expensive to implement in proportion to the value of the digital object.

SUMMARY OF PRODUCTS

The output of Fedora and the Preservation of University Records Project consists of twelve reports and a ingest prototype tool. The reports fall into four groups: Introduction, Ingest, Maintain, and Findings. All reports and the ingest prototype tool are available through the project website at http://dca.tufts.edu/features/nhprc/index.html. The individual reports are available as PDF documents at the locations listed below. Tufts University and Yale University jointly published Version 1.0 of all reports in September 2006.

The deliverables of the project are:

Part One: Introduction

1.1 Project Overview
http://dl.tufts.edu/view_pdf.jsp?urn=tufts:central:dca:UA069:UA069.004.001.00001
An introduction to the project and outline of reports

1.2 System Model
http://dl.tufts.edu/view_pdf.jsp?urn=tufts:central:dca:UA069:UA069.004.001.00002
A description of the components that comprise a recordkeeping system, preservation activities, and the relationship between recordkeeping systems and preservation activities

1.3 Concerns
http://dl.tufts.edu/view_pdf.jsp?urn=tufts:central:dca:UA069:UA069.004.001.00003
Six attributes implicit in all requirements of the Requirements for Trustworthy Recordkeeping Systems and the Preservation of Electronic Records in a University Setting, and all the steps in the Ingest Guide and the Maintain Guide

1.4 Glossary
http://dl.tufts.edu/view_pdf.jsp?urn=tufts:central:dca:UA069:UA069.004.001.00004
Definition of terms used throughout all project reports. Most capitalized words in the project reports are terms defined in the Glossary

1.5 Requirements for Trustworthy Recordkeeping Systems and the Preservation of Electronic Records in a University Setting
http://dl.tufts.edu/view_pdf.jsp?urn=tufts:central:dca:UA069:UA069.004.001.00005
Requirements for trustworthy recordkeeping systems and requirements for preservation activities

Part Two: Ingest

2.1 Ingest Guide
http://dl.tufts.edu/view_pdf.jsp?urn=tufts:central:dca:UA069:UA069.004.001.00006
A guide defining the necessary steps for a trustworthy Ingest process and a description of the resources needed to operate this process in a semi-automated manner. In addition to the

Guide's PDF version, a web version exists at http://dca.tufts.edu/features/nhprc/reports/ingest/index.html.

2.2 Ingest Projects
http://dl.tufts.edu/view_pdf.jsp?urn=tufts:central:dca:UA069:UA069.004.001.00007
An examination of three ingest projects undertaken according to the Ingest Guide

2.3 Ingest Tools
http://dl.tufts.edu/view_pdf.jsp?urn=tufts:central:dca:UA069:UA069.004.001.00008
A description of the Ingest tools examined and developed by project staff, including the Tufts Ingest Prototype System (TIPS), which is available at http://dca.tufts.edu/features/nhprc/reports/tips/index.html. TIPS is available as two files: TIPS-alpha1-nolibs.tar.gz which includes source code and documentation without the library dependencies and TIPS-alpha1.tar.gz which includes source code and documentation with most of the library dependencies

Part Three: Maintain

3.1 Maintain Guide
http://dl.tufts.edu/view_pdf.jsp?urn=tufts:central:dca:UA069:UA069.004.001.00009
A guide defining the necessary steps for a trustworthy maintain process

3.2 Checklist of Fedora's Ability to Support Maintain Activities
http://dl.tufts.edu/view_pdf.jsp?urn=tufts:central:dca:UA069:UA069.004.001.00010
A description of abstract services needed to support steps described in the Maintain Guide and an analysis of Fedora's ability to support these services

Part Four: Findings

4.1 Analysis of Fedora's Ability to Support Preservation Activities
http://dl.tufts.edu/view_pdf.jsp?urn=tufts:central:dca:UA069:UA069.004.001.00011
An overview of Fedora's ability to support services needed for preservation activities and Fedora's current and potential role in records preservation

4.2 Conclusions and Future Directions
http://dl.tufts.edu/view_pdf.jsp?urn=tufts:central:dca:UA069:UA069.004.001.00012
A general discussion of project findings and opportunities for building on the work of this project

Supporting Documents
The following documents concern the Tufts-Yale Project but are not products of the Project:

Project Narrative
http://dl.tufts.edu/view_pdf.jsp?urn=tufts:central:dca:UA069:UA069.004.001.00013

Revised Plan of Work
http://dl.tufts.edu/view_pdf.jsp?urn=tufts:central:dca:UA069:UA069.004.001.00014

Interim Narrative Report
January 31, 2005
http://dl.tufts.edu/view_pdf.jsp?urn=tufts:central:dca:UA069:UA069.004.001.00015

Interim Narrative Report
August 24, 2005
http://dl.tufts.edu/view_pdf.jsp?urn=tufts:central:dca:UA069:UA069.004.001.00016

Interim Narrative Report
February 27, 2006
http://dl.tufts.edu/view_pdf.jsp?urn=tufts:central:dca:UA069:UA069.004.001.00017

Final Narrative Report
September 27, 2006
http://dl.tufts.edu/view_pdf.jsp?urn=tufts:central:dca:UA069:UA069.004.001.00018

ACKNOWLEDGEMENTS

The staff of the Tufts-Yale Project included Kevin L. Glick, Co-Principal Investigator, Manuscripts and Archives at Yale University; and Eliot Wilczek, Co-Principal Investigator, and Robert Dockins, Project Analyst, Digital Collections and Archives at Tufts University.

Additional members of the project team at Yale included Stephen Yearl and Raman Prasad, Manuscripts and Archives; Roy Lechich, Integrated Library Technology Services; and David Gewirtz and Neil (Xinjian) Guo, Academic Media and Technology. Addition members of the Project team from Tufts included Anne Sauer, Robert Chavez, and Greg Colati, Digital Collections and Archives.

The Project team consulted with and benefited greatly from the contribution of several additional people who all deserve thanks for their efforts.

The National Historical Publications and Records Commission (NHPRC) provided financial support for this project through an electronic records research grant. In addition to its financial support, the NHPRC staff provided us with crucial grant development administration assistance. In particular, Mark Conrad provided critical guidance for turning the initial grant proposal into a funded grant project. The Yale University Library Business Office and the Office of the Vice Provost and Sponsored Programs Accounting at Tufts also provided essential grant administration support.

Nancy McGovern, formerly of Cornell University and currently at the Inter-University Consortium for Political and Social Research, provided invaluable guidance to the project team's formulation of the Requirements for Trustworthy Recordkeeping Systems and the Preservation of Electronic Records in a University Setting. She helped the project team think through some of the tricky issues and problems concerning the requirements and helped the team shape them into a logical order.

Thornton Staples, University of Virginia, provided the project team with invaluable insights about Fedora and ideas for using Fedora to help manage records throughout the duration of the grant project and particularly at a two-day meeting with the project team in September 2005.

During the course of the grant project Kevin L. Glick and Eliot Wilczek became members of the Fedora Preservation Services Working Group.[6] Members of the Working Group have given the principle investigators a deeper understanding of Fedora—particularly its service framework guiding the development of future Fedora services. This group has also provided valuable comments on the Maintain Guide. Members of the Working Group in addition to Glick and Wilczek include Ron Jantz (Chair) and Grace Agnew, Rutgers University; Dan Davis, Harris Corporation; and Sandy Payette, Cornell University. Sandy Payette also provided explicit technical insights on Fedora for Checklist of Fedora's Ability to Support Maintain Activities.

[6] See <http://www.fedora.info/wiki/index.php/Working_Group:_Preservation> for more information.

Patsy Baudoin, an independent digital archivist, edited several of the project reports. She caught many errors and greatly improved the writing of the reports she edited.

Liz Chrastil, a former graduate assistant at the Digital Collections and Archives at Tufts University, conducted research that supported the Project's preliminary explorations of creating and managing Producer Records, a resource described in the Ingest Guide.

Raman Prasad and Niloufer Moochhala of Nymdesign <www.nymdesign.com> designed and encoded the web version of the Ingest Guide.

The following people provided comments on various project reports: Tom Hyry and Richard Szary, Yale University; Cal Lee, University of North Carolina at Chapel Hill; Rebecca Hatcher, Northeast Document Conservation Center; Luke Meagher, University of British Colombia; and Jim Suderman, Library and Archives Canada.

Although many contributed to the research of the project, it should be noted that any project shortcomings or errors are solely the responsibility of the Co-Principal Investigators.