

Document Title Project Narrative

Document Number 0.2

Version Final

Date 05/27/04

NHPRC Grant Number 2004-083

Introduction

The Digital Collections and Archives (DCA), Tufts University, in conjunction with Manuscripts and Archives (MSSA) of Yale University Library, is seeking a grant from the National Historical Publications and Records Commission (NHPRC) 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. The DCA and MSSA propose a research project to test the potential of Fedora (the Flexible Extensible Digital Object and Repository Architecture) to serve as the architecture for such an electronic records preservation system.

While archivists have been researching electronic records preservation with minimal program development, the digital library community has focused on building prototype systems that store, manage, and deliver content without focusing, at least initially, on addressing long-term preservation issues. Many successful projects have focused on the e-print community and on increasing the availability of scholarly communications. Although these groups often develop so-called “e-print archives,” they have generally lacked a real concern about the long-term preservation of, and access to,

the resources within the disparate repositories.¹ Without adherence to archival principles these digital tools cannot adequately preserve electronic records over the long-term. With the body of knowledge it has developed in the 1990s, the archival community has much to teach the digital library community about records persistence, trustworthiness, and authenticity. Archivists have always been concerned with preserving authentic and trustworthy records, and the important theoretical research conducted by the Minnesota Historical Society and the InterPARES Project shows that archivists are knowledgeable about and committed to authenticity and digital preservation.

The Tufts/Yale project will synthesize archival theory and principles with digital library practice by injecting these theories and principles into Fedora, one of the most significant systems to emerge from the digital library community. Although Fedora is not currently designed to function as an electronic records preservation system, we feel it has the potential to do so through the application of archival theories and principles. The project will test the capabilities of a major tool from the digital library community against archival preservation and access standards for university electronic records.

Fedora

The Fedora digital object repository management system is the result of a digital library research project funded by the Andrew W. Mellon Foundation.² The system is freely available and now supported by the open-source community. The Fedora architecture is based on the concepts of object-oriented programming, a software design

¹ The report of the October 2002 meeting of the Open Archives Initiative describes the group's lack of concern for long-term preservation. <<http://library.cern.ch/mcmahon.htm>>

² "The Fedora Project: Developing An Open-Source Digital Repository Management System," <http://www.fedora.info>.

method that models digital objects using classes that define properties of the objects and operations that may be performed by those objects. Fedora employs XML and METS (Metadata Encoding Transmission Standard) to create digital objects by encapsulating content, along with metadata about the content, and actions that can be performed on the content. This linkage of the content to the applications that are used to search, render, and display it distinguishes Fedora from other digital repository systems. Clients (administrative or patrons) interact with the Fedora repository through an application-programming interface (API) that provides management and access services. Fedora developers assert that the object-oriented architecture of Fedora makes migrations less risky and less complex for repositories storing heterogeneous collections. Yale and Tufts have chosen to test Fedora because the architecture: (1) is format neutral, (2) has the flexibility to deal with the diversity of electronic record types, (3) supports versioning to preserve access to former instantiations of both content and services, (4) can manage digital objects as items within a collection and manage collections with complex structures, and (5) includes a technology suite that incorporates XML and Web Services, which are ideally suited to nourish and develop preservation networks.

Yale's and Tufts' Commitment to Electronic Records Preservation

Tufts and Yale come to this project from very different backgrounds and situations, but both have already made firm commitments to developing preservation functions for university electronic records. These different backgrounds and situations will benefit the project because each institution will add its particular expertise and be able to evaluate Fedora with different content and resources, allowing the findings to be more generalizable.

Yale has been considering digital preservation questions for a long time and brings institutional knowledge of digital preservation to the project. The Yale University Library, of which MSSA is a unit, has one of the oldest data archives in American universities, first acquiring numeric data sets in 1972. Since that time, Yale has continued to preserve data sets and was even contracted to produce research on preserving social science data for the Digital Library Federation.³ In 2002, Yale completed a major research collaboration with Elsevier Science, funded by the Andrew W. Mellon Foundation, to study the challenges and opportunities for long-term preservation of commercially published scientific journals.⁴ Recently, Yale University Library completed a yearlong strategic planning initiative to create a vision of a mature integrated library that is a steward and a custodian of both analog and digital content. A core value of the integrated library is to build an infrastructure for the long-term preservation of scholarly outputs of the Yale community. As part of the University Archives program, work is underway to establish an electronic records program. A permanent Electronic Records Archivist position has been established and project funding has been appropriated to research, test, and develop prototype systems for the preservation of Yale's electronic records. The project group has assessed the current landscape of electronic records at Yale, examined the records landscape at peer institutions, and has begun to map out a vision for potential electronic records preservation systems. This NHPRC grant project would enable Yale to develop and test one particular prototype preservation system.

³ Ann Green, JoAnn Dionne, and Martin Dennis, *Preserving the Whole: A Two-Track Approach to Rescuing Social Science Data and Metadata*, Washington, DC: The Digital Library Federation, 1999.

⁴ *YEA: The Yale Electronic Archive, One Year of Progress*. Report on the Digital Preservation Planning Project, A Collaboration between Yale University Library and Elsevier Science, New Haven, 2002.

The Digital Collections and Archives (DCA), Tufts University was created in July 2001 to be the central repository for permanently valuable intellectual content produced at Tufts. It is a central university office with a permanent budget, staffing, and infrastructure. The DCA has implemented Fedora to manage its digital repository of permanent digital objects. The DCA will have all of its current digital collections, which include texts, images, audio files, and finding aids under Fedora management by Fall 2003. The DCA also plans to add archival electronic records generated by the University to the digital repository. This project is part of the Digital Collections and Archives' five-year strategic plan for July 1, 2003 to June 30, 2008. The mission of the DCA states in part, "The Digital Collections and Archives (DCA) is the steward of the University's permanently valuable records and collections created in any format, ensuring their permanent preservation and accessibility." The vision of the DCA includes being an organization that (1) provides long-term access to managed, permanently valuable resources created in all formats in a user-centered way; (2) that accepts responsibility for the long-term maintenance of all resources on behalf of its depositors and for the benefit of current and future users; and (3) that designs its systems in accordance with commonly accepted conventions and standards to ensure the ongoing management, access, and security of materials deposited within it. In order to meet its mission and vision the DCA has a goal of developing and implementing a system to manage electronic records. This project is one of the actions we plan to undertake to achieve this goal. The grant will help fund the extra resources needed to develop and implement a strategy for accessioning and managing archival electronic records in Fedora. Once this strategy is implemented, the

DCA is committed to managing archival electronic records in the digital repository as part of its mandate and as a routine part of its operations.