



TAPER
TUFTS ACCESSIONING PROGRAM FOR ELECTRONIC RECORDS

National Historical Publications and Records Administration

Electronic Records Project: Program Expansion Project

Application: Project Narrative May 31, 2007

PROJECT PURPOSE AND GOALS

Introduction

The Digital Collections and Archives (DCA) of Tufts University is applying for a three-year NHPRC electronic records Program Expansion Project grant to support the Tufts Accessioning Program for Electronic Records (TAPER). The TAPER Project will make key contributions to systematizing and semi-automating the DCA's process for accessioning electronic records and will play a crucial role in the DCA's effort to further develop its program of preserving and making accessible electronic archival records. Although the DCA has implemented a Fedora-based repository for managing and preserving electronic records along with teaching and research digital objects, its current process for accessioning electronic records needs improvement and systemization to meet the growing demand for preserving electronic records that document the activities of Tufts and other institutions served by the DCA. The current process for transferring electronic records from creators to the DCA is not scalable and does not properly document the records and their terms of transfer. In order to continue its development of a sustainable electronic records preservation program, the DCA needs to have a semi-automated and scaleable transfer process that richly documents the records and their terms of transfer.

The need to systematize our ingest process has become increasingly pressing for the DCA because Tufts' senior leadership recently approved a University Records Policy and Guidelines for Managing University Records. The Policy and Guidelines should increase



faculty and staff's awareness about properly managing electronic records and transferring records with enduring value to the DCA. The DCA is planning outreach programs to help ensure compliance with the Policy and Guidelines. The DCA anticipates a significant increase in the number of electronic records people will want to transfer to the DCA. An accessioning process that can meet this increased demand is critical to the success of the DCA's electronic records program. If the DCA's current accessioning process is not improved, it will become a drag on electronic records program through its slowness and inability to capture the documentation needed to rapidly and systematically process and describe electronic records.

The TAPER Project will focus on developing and implementing Submission Agreements that would richly and systematically document accessioned records and the terms of their transfer. The TAPER Project would enable the DCA to achieve two main goals:

1. Facilitate the regular and systematic transfer of archival electronic records from records creators to the DCA by making it easier for creators to transfer records and easier for the DCA to manage incoming records.
2. Provide a structure for capturing the necessary documentation for archival electronic records. This documentation will help the DCA regularize and semi-automate its preservation decisions and processing of electronic records.

Current State of Electronic Records Program

The DCA is the steward of Tufts University's permanently valuable records and collections created in any format, ensuring their permanent preservation and accessibility. The DCA supports the Tufts Digital Repository, the university-wide Records Management Program, and the University Archives and Manuscript Collections.

The DCA is committed to sustaining an active electronic records program. The DCA, in partnership with University Information Technology (UIT) Academic Technology, has



already established the Tufts Digital Repository (TDR). This gives the DCA a robust preservation environment for its archival electronic records. The TDR:

- Is Fedora-based (Flexible Extensible Digital Object Repository Architecture)¹
- Is supported and managed by the DCA and UIT Academic Technology with server support from University Information Technology Systems Group and the Network Operations Center.
- Contains approximately 70,000 digital objects that support teaching and research and document some aspects of Tufts history.
- Supports the Tufts Digital Library, the general search and delivery tool for the TDR, and other specialized access tools that provide access to specific digital collections from the TDR.
- Has the flexible architecture needed to meet the future digital preservation and access needs of the Tufts community, enabling it serve as Tufts' long-term institutional repository for digital objects and electronic records.

Last year the DCA completed an NHPRC electronic records research grant (2004-083) with Yale University, "Fedora and the Preservation of University Records."² One of the project's deliverables was findings on Fedora's ability to serve as the basis for a trustworthy preservation system.³ As a flexible repository architecture, Fedora allows institutions to configure a Fedora-based repository in a wide variety of ways that can range from a robust preservation system to something that is wholly inappropriate for preserving electronic records. A trustworthy preservation system includes many components beyond the repository application. Tufts and Yale concluded that a Fedora repository instance, surrounded by a proper set of policies, tools, services, and institutional commitment, can serve as the basis of a trustworthy preservation system. Therefore, Fedora gives the DCA a firm foundation for

¹ Fedora <<http://www.fedora.info>>.

² Fedora and the Preservation of University Records, <<http://dca.tufts.edu/features/nhprc/index.html>>.

³ "4.1 Analysis of Fedora's Ability to Support Preservation Activities," <<http://dca.tufts.edu/features/nhprc/reports/index.html>>.



managing and preserving archival electronic records in a trustworthy preservation system.

Trustworthy management and preservation is predicated on a regular, systematized accessioning process in which vital information is captured. The TAPER project, by enabling this process, will allow the DCA to dramatically improve and expand its electronic records program.

The DCA currently manages and preserves 190 university records collections and 137 manuscript collections. 20 of those collections contain electronic records. Over the past five years the DCA has averaged 41 accessions a year. During this same time period, the DCA has averaged a few accessions each year of electronic records—although this has increased to five accessions with electronic records in the current fiscal year.⁴ Most of the accessioned electronic records have come to the DCA on storage media such as CDs and diskettes. The DCA has recently implemented procedures to immediately move electronic records from accessioned media to a secure and regularly backed-up network storage environment. This process includes conducting virus checks, creating checksums, and generating documentation to establish records' chain of custody and provide evidence of their authenticity.⁵ The DCA has also begun to accession electronic records from some Tufts offices via the Tufts network, eliminating the need to manage transfer media.⁶

Tufts University is committed to supporting the DCA and properly managing its electronic university records, including the preservation of electronic records with enduring value. To bolster that commitment Tufts recently approved the addition of two permanent staff positions to the DCA, increasing its permanent staff from five to seven starting July 1,

⁴ These figures do not include the thousands of digital objects the DCA ingests into the Tufts Digital Repository each year.

⁵ See the Digital – Media Stabilization Accession Procedures in the Supplementary Materials.

⁶ See the Digital – Network Stabilization Accession Procedures in the Supplementary Materials.



2007. In addition, the Provost and the Executive Vice President approved the aforementioned new University Records Policy that defines university records at Tufts (including electronic records); establishes the DCA's authority to appraise and preserve university records; and delineates the responsibilities of all Tufts employees and departments to manage university records properly. In collaboration with University Counsel and Audit & Management Advisory Services, the DCA issued Guidelines for Managing University Records to help members of the Tufts community properly follow the Records Policy. The Guidelines give advice on creating and storing records; determining the appropriate disposition for records, including transferring records with enduring value to the DCA; confidentially destroying records; and complying with appropriate laws, regulations, and ethics concerning the management of records.

The DCA has also engaged itself in extensive research on electronic records with its recently completed NHPRC electronic records research grant with Yale. That project produced four main deliverables:

1. Requirements for recordkeeping systems and preservation activities
2. Ingest Guide for ingesting records in a preservation system in a trustworthy manner
3. Maintain Guide for maintaining records in a preservation system in a trustworthy manner
4. Findings on Fedora's ability to serve as the basis for a trustworthy preservation system.

Producing work for the NHPRC grant on requirements for recordkeeping systems and preservation activities and trustworthy ingest and maintain programs helped the DCA develop a roadmap for expanding its electronic records program. The TAPER project represents a critical step in that ongoing development and growth of the DCA's program.



PROJECT'S SIGNIFICANCE TO THE NHPRC'S PROGRAMS AND GOALS

The TAPER Project will support the NHPRC's goal of supporting institutions in meeting the challenges of preserving and managing electronic documentation. The TAPER Project will support the DCA's efforts to systematically and comprehensively accession, preserve, and make appropriately available the electronic records of enduring value that document the activities of Tufts University. In addition, the DCA serves as the archival repository of several independent or Tufts-affiliated organizations in the fields of environmental stewardship, international development, and active citizenship. These organizations include the World Peace Foundation, the Mystic River Watershed Association, and the Center for Health, Environment and Justice. This project will help preserve the electronic documentary heritage of these fields.

The work of the TAPER Project can also serve as a model for other institutions charged with the responsibility of accessioning and preserving electronic records of enduring value. As mentioned in the Project Deliverables section below, the DCA will make its Submission Agreement deliverables freely available for use by others. In addition, the DCA will openly share its methods and experiences in expanding its electronic records program.



PLAN OF WORK

Introduction

In order to sustain and expand its electronic records program, the DCA needs to have a semi-automated and scalable accessioning process that can systematically capture and document records in a manner that facilitates their semi-automated and regularized processing. A machine-readable Submission Agreement serves as the lynchpin of this desired process. The development and implementation of this Submission Agreement is the main focus of this project's work.

Submission Agreement

The Ingest Guide, which the DCA developed with Yale University in its NHPRC electronic records research grant, describes the steps and tools needed to have a scalable and trustworthy process for ingesting records into a preservation system. In the first section of the Guide, Negotiate Submission Agreement, an archives and records creator define and agree to the scope and terms of a records transfer, which they document in a Submission Agreement. In the second section, Transfer and Validation, the records creator transfers the records and the archives validates and prepares the records and submits them to its preservation repository.⁷

The Submission Agreement is a core component of the trustworthy ingest process described by the Guide. The Submission Agreement documents the scope and terms of transfer and provides the framework for semi-automating an archives' electronic records processing work.

The Guide envisions that Submission Agreements could exist as machine-readable files that

⁷ 2.1 Ingest Guide, <<http://dca.tufts.edu/features/nhprc/reports/index.html>>. The Ingest Guide is composed of 17 parts and 88 steps. It also describes 30 resources needed for a semi-automated and scalable trustworthy ingest process. The Guide is based on: Consultative Committee for Space Data Systems, *Producer-Archive Interface Methodology Abstract Standard*, CCSDS 651.0-B-1, Blue Book, May 2004, <<http://www.ccsds.org/CCSDS/documents/651x0b1.pdf>>. More broadly, the Ingest Guide is based on: ISO 14721:2003, Space data and information transfer systems – Open Archival Information System – Reference model.



an archives could use to validate the records it accessions and the Agreements could also prescribe many of the preservation activities an archives undertakes on its records.

A Submission Agreement would be composed of several elements that would document essential aspects of accessioned records. Based on the Ingest Guide, these elements include:

- Reference to associated records survey
- Archives/Repository identification
- Producer identification
- Description of records
- List of record types
- Rules for preparing records for transfer (Submission Information Package creation rules)
- Transfer procedures and schedule
- Copyright status
- Access restrictions/security profile
- Archival collection(s) transferred records will belong to
- Archival naming scheme for transferred records
- Validation procedures
- Archival metadata formats.

During the course of the TAPER Project, Tufts may adjust this list of elements as it builds and implements a machine-readable version of the Submission Agreement.

Tufts will also implement methods to associate a Submission Agreement with the records it describes. For its Fedora repository environment, Tufts will create a Submission Agreement content model and use RDF metadata to configure the correct relationships between the Submission Agreements and their records. This will enable the DCA ensure that archival records' terms of use, ownership status, essential elements, and archival value are permanently and clearly identified and documented. This will help ensure that future DCA staff know the essential nature of the records held by the DCA and understand why the DCA has them.



In addition to this long-term documentation, a machine-readable Submission Agreement will help the DCA upgrade its archival processing operation by enabling it to systematically plan and execute its preservation activities. The Submission Agreement provides a structure for documenting the information the DCA needs to preserve electronic records. The Submission Agreement elements described in the Ingest Guide also act as a checklist to help the DCA ensure that it has captured all of the supporting documentation that should accompany accessioned records. To help ensure that the Submission Agreement is designed to capture the appropriate information, the DCA will evaluate metadata and documentation requirements for transferring electronic records or digital objects to archives or digital repositories at several leading institutions such as the National Archives and Records Administration, the University of Virginia, Harvard University, the Massachusetts Institute of Technology, and Rutgers University.⁸

The elements composing the Ingest Guide also establish a framework for making the DCA's archival processing of electronic records a largely automated and scalable process. Elements in the Guide that require information on record creators, record types, file formats, transformation procedures, SIP creation procedures, among other elements, provides with DCA with a structure and space to define these activities, rules, and standards. Within the scope of the Program Expansion grant, this information will be embedded within Submission Agreements. Beyond the grant, the DCA plans to have its Submission Agreements reference externally defined activities, rules, and standards. For example, a Submission Agreement may reference a records creator record (written to the ISAAR-CFP content standard and structured

⁸ Tufts will also check the Submission Agreement elements against the *Trustworthy Repositories Audit & Certification: Criteria and Checklist*, Version 1.0 February 2007, <<http://www.crl.edu/PDF/trac.pdf>>. Sections B1 and B2 of *Trustworthy Repositories* mandates the systematic capture and documentation of information about ingested digital objects to help ensure their understandability.



in EAC⁹), a record type record (an XML file that describes a type of record—such as meeting minutes—and defines its disposition and essential elements), its format type (referencing PRONOM or other format registry¹⁰), and set of procedures for preparing electronic records for transfer to the DCA. This would enable the DCA to shift its attention away from handling and processing individual or small sets of electronic records and focus on maintaining and developing rules, procedures, and references to standards that facilitate the DCA’s systematic and scalable processing and preservation of electronic records.¹¹ Work on implementing the Submission Agreement within the scope of the Program Expansion grant project is a crucial step towards this goal.

Tasks

See the Work Plan Chart in the Supplementary Materials for a timeline of the tasks.

Task One

Submission Agreement Elements and Use Cases

Months 1-6

University Records Manager (URM) and Digital Repository Program Manager (DRPM), and Digital Resources Archivist (DRA)

(See the Project Personnel section below for full details on project staff).

- The URM writes:

⁹ ISAAR(CFP) International Standard Archival Authority Record for Corporate Bodies, Persons, and Families, Second edition, <<http://www.ica.org/en/node/30230>>. Encoded Archival Context, <<http://www.iath.virginia.edu/eac/>>.

¹⁰ PRONOM <<http://www.nationalarchives.gov.uk/pronom/>>.

¹¹ Handling and “processing” electronic records in the same manner as archives have traditionally processed paper records is simply not feasible. It is a process that is ill-equipped to handle the volume and complexity of electronic records that the DCA and other archives are or will be responsible for preserving. In addition, traditional processing that focused on handling individual or small batches of records could not even address the challenges presented by modern paper collections. Mark A. Greene and Dennis Meissner, “More Product, Less Process: Revamping Traditional Archival Processing,” *American Archivist*, Volume 68, Number 2, (Fall/Winter 2005). Christine Weideman describes working with records creators to arrange and describe records before Manuscripts and Archives at Yale University accessions them. Although Weideman discusses this strategy in the context of paper records, the principle of ensuring that records are well-ordered and managed before records creators transfer them to an archives holds true for electronic records as well. See Christine Weideman, “Accessioning as Processing,” *American Archivist*, Volume 69 Number 2 (Fall/Winter 2006) ps 274-283. For more on shifting the focus of archival processing to the development and management of procedures and rules see Kevin Glick, Eliot Wilczek, and Robert Dockins, “Fedora and the Preservation of University Records Project,” *RLG DigiNews*, Volume 10, Number 5, (October 2006) <http://www.rlg.org/en/page.php?Page_ID=20987#article0>.



- Description of the Submission Agreements' role for accessioning electronic records;
- Elements set for Submission Agreements based on the Ingest Guide and to a lesser extent on the *Trustworthy Repositories Audit & Certification* and the metadata and documentation requirements of leading archival and digital repository institutions;
- Use cases for Submission Agreements illustrating how records creators and the DCA would produce Submission Agreements and how Submission Agreements would document electronic records accessions.
- The DRPM reviews the descriptions and use cases to ensure they contain the appropriate detail for the Project Developer (PD) to create an XML Submission Agreement schema. The URM makes adjustments to the descriptions and use cases as needed.
- The URM tests human-readable versions of the Submission Agreement on accessions of electronic records from some of the following offices
 - Meeting minutes and reports from various committees at the Fletcher School of Law and Diplomacy
 - Working papers from the Office of the Trustees
 - Project files (a mix of reports, interviews, presentations, papers, and photographs) from the Feinstein International Center, a research center at Tufts
 - Emails from select Tufts administrators
 - Health Sciences course syllabi from the Tufts University Science Knowledgebase
 - Datasets from the Office of Institutional Research and Evaluation
 - Reports, minutes, policies, and procedures from the Tisch Library (Arts, Sciences & Engineering library)
 - Publications from the Mystic River Watershed Association.The URM makes adjustments to the structure of the Submission Agreement as needed.
- The writing, reviewing, and testing work described above would occur iteratively during this six-month period.
- As the project moves to the Task Two, the URM will implement the use of human-readable Submission Agreements into the DCA's normal accessioning procedures for electronic and paper records. This will help project staff identify necessary modifications that are needed to incorporate the Submission Agreements in the DCA's workflow.
- The URM documents the paper Submission Agreement and its use in the DCA's procedures.

Task Two

Develop Submission Agreement Schema

Months 7-18

Project Developer (PD), URM, and DRPM



- Based on the elements set and use cases written in Task One, the PD creates an XML schema for Submission Agreements.
- The URM and DRPM review and comment on XML schema. The PD makes adjustments as needed.
- The PD and URM create Submission Agreement records (structured according to the Submission Agreement XML schema) with data from the human-readable Submission Agreements tested in Task One and from accessions of electronic records that come to the DCA during Task Two.
- The URM and DRPM review and comment on the Submission Agreement records. The PD makes adjustments as needed.
- The PD and URM document the Submission Agreement schema and use of electronic Submission Agreements in the DCA's procedures.

Task Three

Develop Submission Agreement Builder Tool

Months 16-24

PD, URM, and DRPM

- The URM writes use cases for the Submission Agreement Builder, a web-based tool that records creators and the DCA use to create Submission Agreements. The tool allows records creators and the DCA to add data to a Submission Agreement in stages starting shortly after initial contact, throughout the appraisal process, and finishing ahead of the actual transfer.
- Based on the uses cases, the PD creates the Submission Agreement Builder tool.
- The URM tests the Submission Agreement Builder by using it with records creators to generate Submission Agreements. These records creators include some listed in Task One and others transferring records to the DCA during Task Three.
- Based on the results of the testing, the PD makes adjustments to the tool as needed.
- The PD and URM document the Submission Agreement Builder tool and its use in the DCA's procedures.

Task Four

Implement Management of Submission Agreement in Preservation Repository

Months 19-30

PD, URM, and DRPM

- The PD and URM develop a plan for preserving Submission Agreement XML files in the preservation repository and associating Submission Agreements with the records they describe. The PD and URM write use cases for a Submission Agreement content model. The content model would allow Tufts' Fedora-based repository to recognize Submission Agreements objects in the repository as Submission Agreements (as defined by the content model) and in turn manage them and their relationships according to a set of rules defined by the content model.¹²

¹² "Content Model Forum," <http://www.fedora.info/wiki/index.php/Content_Model_Forum>.



- The DRPM reviews the uses cases for the content models. The PD and URM make adjustments as needed.
- The PD undertakes work on developing the content model. Around this time (month 24; end of December 2009), grant funding for the PD will end. When this occurs the DRPM or URM will take over the remaining content model development work as a Tufts Digital Repository project task.
- Project staff iteratively tests the content model by submitting Submission Agreement objects (describing records listed in Task One and other records accessioned by the DCA during Task Four) to the preservation repository during the content model development process.
- The PD, URM, and DRPM document the Submission Agreement content model and the preservation of Submission Agreement objects in the DCA's procedures.

Task Five

Integration of the Submission Agreement

Months 4-36

URM

- Throughout the project, DCA staff will implement the deliverables created in Tasks One through Four.
 - When project staff have established a firm set of Submission Agreement elements, the DCA will integrate a paper-based Submission Agreement into its accessioning procedures. This will already improve the DCA's documentation regimen for accessioning electronic and paper records.
 - When project staff have established a functional XML schema for Submission Agreements, the DCA will create electronic Submission Agreements for its accessions of electronic and paper records.
 - When project staff has developed the Submission Agreement Builder tool, the DCA will use the tool to generate Submission Agreements. This will increase the efficiency and ease of creating Agreements for the DCA and records creators. By making it easier to transfer electronic records, the tool will help the DCA document the activities of Tufts and the other institutions more comprehensively.
 - When project staff has developed the Submission Agreement content model, the DCA will submit Submission Agreement objects into its preservation repository. The Submission Agreement objects, defined by the content model, will help enable the DCA approach its preservation planning systematically and execute its preservation activities appropriately.
- The URM will integrate instructions for using Submission Agreements into the training program the DCA is developing in 2007 to support the recently approved University Records Policy and Guidelines for Managing University Records.



PROJECT DELIVERABLES

The DCA will produce the following project deliverables:

1. Submission Agreement XML schema
2. Submission Agreement Builder tool
3. Submission Agreement content model
4. Accessioning procedures that incorporate the Submission Agreement
5. Training materials on managing records (including electronic) that include information on transferring records to the DCA using Submission Agreements.

The DCA will disseminate the deliverables and information about the project in the following manner:

1. The DCA will make all project deliverables freely available through a project website.
2. At Tufts, the DCA will train staff on transferring records to the DCA using a variety of training methods (in-person classes, video, and/or web).
3. Beyond Tufts, the DCA will give presentations on the project at appropriate venues such as the Joint Conference on Digital Libraries and the Society of American Archivists' annual conference.
4. In addition to its progress and final narrative reports to the NHPRC, the DCA will produce an article(s) describing the project, its deliverables, and lessons learned in an appropriate venue(s) such as *D-Lib* or *American Archivist*.



PROJECT PERSONNEL

Eliot Wilczek
University Records Manager
Program Expansion Project Director

17-34% of time contributed to project, see Budget for details

Eliot Wilczek manages the university-wide Records Management Program, which is part of the Digital Collections and Archives. His responsibilities include helping offices and departments (including outside institutions that the DCA serves) transfer appropriate records to the DCA and assist with accessioning those records into the DCA's holdings. Mr. Wilczek has experience as an archivist and records manager and knowledge of records and archival theory and electronic records and digital preservation issues. Mr. Wilczek served as Co-Principal Investigator of an NHPRC electronic records research grant with Yale University, "Fedora and the Preservation of University Records."

Robert Chavez
Digital Repository Program Manager

2.5% of time contributed to project

Robert Chavez manages the Tufts Digital Repository, which contains over 70,000 digital objects primarily pertaining to teaching, research, and Tufts history. Mr. Chavez has extensive knowledge of the Fedora repository system, XML, metadata and digital object creation and management, system administration, several programming languages, and digital preservation issues.

Position to be Filled
Project Developer

100% of time contributed to project

This NHPRC-funded position will be devoted to the following project tasks: 1) create XML schema for Submission Agreements, 2) create web-based tool for creating Submission Agreements, 3) design and start development of a Submission Agreement content model. The position requires considerable knowledge of XML, JSP, Java, PHP, and Perl. Knowledge of digital repository systems, particularly Fedora, and digital preservation issues preferred. Depending on the experience of Project Developer, he or she will devote part of his or her time to learning about Fedora. This will help prepare the Project Developer for work on the Submission Agreement content model.



PERFORMANCE OBJECTIVES

The DCA plans to meet the following objectives by the conclusion of the three-year Program

Expansion Project:

- The DCA will document all of its accessions of electronic and paper records with Submission Agreements.
- The DCA will preserve its Submission Agreements in its preservation repository and manage a permanent connection between the Submission Agreements and the records they describe.
- The DCA will double its current average annual accession of records to 80 accessions per year). Half of those accessions (40 accessions per year) will include electronic records.
- The DCA will produce a Submission Agreement XML schema, a tool for generating Submission Agreements, and a Submission Agreement content model.