Fedora and the Preservation of University Records Project

4.2 Conclusions and Future Directions

Version
1.0

Date
September 2006

Co-Principle Investigators
Kevin Glick, Yale University
Eliot Wilczek, Tufts University

Project Analyst
Robert Dockins, Tufts University

Fedora and the Preservation of University Records Project

TABLE OF CONTENTS

FUTURE DIRECTIONS

Implementing the work of "Fedora and the Preservation of University Records"
The work of the Tufts-Yale Project suggests several different avenues of future work that can build on the output of this grant project. This future work centers on ways to implement— sometimes in a semi-automated fashion—the Project's Requirements for Trustworthy Recordkeeping and Preservation, the Ingest Guide, and the Maintain Guide.

Understanding the Tufts-Yale Project within the context of the Reference Model for an Open Archival System Information System (OAIS)[1] will help people understand how the products of this research project can help their work. All three main products of the project map to OAIS functions. The Ingest Guide describes the Ingest function as well as much of Establish Standards and Policies, Audit Submission, and Negotiate Submission Agreement within the Administration function. The Maintain Guide covers the Data Management and Archival Storage functions. The recordkeeping requirements from the Requirements for Trustworthy Recordkeeping and Preservation, although organized according to a section of ISO 15489-1, still map to the activities of a Producer, while the requirements for preservation activities from the same report guide the activities of an Archive and thus cover all the functional areas of the OAIS Reference Model.

Viewing the OAIS Reference Model, the Requirements for Trustworthy Recordkeeping and Preservation, the Ingest Guide, and Maintain Guide, the resources and services that support the two guides, and the implementation of the guides, as a tightly related set of steps which build on each other will help archives and institutions make the best use of these documents, resources, and services. The OAIS Reference Model is the overarching conceptual structure for preservation activities and systems. Beneath OAIS sits the preservation requirements adding further articulation to OAIS by describing the attributes of preservers that fit within the context of the Reference Model. Beneath these requirements are the Ingest Guide and Maintain Guide, which translate requirements into actions for those two functional areas of preservation.[2] Then resources and services—ideally, standardized and openly available—support the execution of the activities defined in the guides. Individual institutions and archives will still have implementation decisions to make within the context of the guides, resources, and services. Archives or 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.

---

[1] 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>
[2] The Tufts-Yale Project team did not develop guides for all functional areas of preservation, such as Access and Preservation Planning.

Implementing the Requirements for Trustworthy Recordkeeping and Preservation
The Requirements for Trustworthy Recordkeeping and Preservation can assist institutions or university archives informally evaluating existing recordkeeping systems or preservation programs, particularly in giving them an outline of issues to address. However, the requirements are currently just a set of requirements; they are not a true evaluation tool. To be effectively used in an assessment, the Requirements must be turned into a true evaluation tool. Archives may be able to leverage the work of other projects like the Center for Research Libraries' "Auditing and Certification of Digital Archives" project to turn the Requirements for Trustworthy Recordkeeping and Preservation into a true evaluation tool or the PLEDGE Project (PoLicy Enforcement in Data Grid Environments), which is developing tools and mechanisms to enable scalable policy expression in digital repositories.[3]

As mentioned earlier, the project staff had great difficulty arriving at an appropriate framework for organizing the requirements, particularly the set for recordkeeping. It may be that rather than fixing them in textual linear document where they are in a set, numerical order, the requirement are instead best served by residing in database or other environment that allows users to flexibly arrange the individual requirements to best suit their needs. For example, a user may only want to see the mandatory requirements or only the requirements for a recordkeeping or preservation application.

Implementing the Ingest Guide and the Maintain Guide
The Ingest Guide and the Maintain Guide essentially take the next step after Requirements for Trustworthy Recordkeeping and Preservation for the ingest and maintain functions. Both Guides prescribe the actions an Archive needs to take to meet the expectations of the requirements for trustworthy ingest and maintain activities. Although prescriptive, both guides are not procedure manuals. The guides describe what actions to take but not precisely how to undertake those actions. For example, the Ingest Guide says an Archive needs to have a Formats Standards Policy declaring their preservation formats, but the Guides does not prescribe what formats the Archive should use as preservation formats. Each archive has to make that decision as part of their policies. Thus, the Guides give archives a detailed framework for creating their own polices and procedures.

Ideally, university archives would implement the Guides in a semi-automated fashion, allowing them to manage and preserve a large volume of electronic records in a scaleable manner. A semi-automated ingest or maintain process will require university archives to implement machine-readable versions of the resources described in the Ingest Guide and the abstract services for maintain activities described in the Checklist of Fedora's Ability to Support Maintain Activities. It will also require many university archives to extensively re-engineer their accessioning, storage, and handling workflows. One promising avenue for future development might be utilizing Business Process Execution Language (BPEL) for turning the Maintain Guide and especially the Ingest Guide into semi-automated, scaleable, trustworthy, processes at individual archives.[4]

---

[3] For more information on the Digital Repository Certification project, see
<http://www.rlg.org/en/page.php?Page_ID=580>, for more information on the PLEDGE project, see
<http://pledge.mit.edu/>.
[4] Internal Report from Thornton Staples to project staff, October 10, 2005.

The Ingest Guide identifies thirty resources needed to support a trustworthy and viable ingest process and the Checklist of Fedora's Ability to Support Maintain Activities identifies nineteen abstract services to support a trustworthy and viable maintain process. Successfully implementing trustworthy, semi-automated, and scalable ingest and maintain process will require university archives to implement a significant portion, if not all, of these resources and abstract services. Some of the services and resources exist; many do not. The preservation, records, digital library, and information science communities would have to either create these abstract services and resources from scratch or adapt existing tools. For example, the Fedora community could use formal content models, rules it is currently developing for defining digital object types, as a syntax for articulating machine-readable Record Types Records.[5] In another adaptive example, archives may be able to use Encoded Archival Description (EAC) as the data structure standard and Internal Standard Archival Authority Record for Corporate Bodies, Persons and Families (ISAAR (CFP)) as the data content standard for Producer Records.[6]

However, implementing these services and resources will require more than simply adapting a metadata standard and syntax. For example, the project team undertook a brief examination of how to implement Producer Records. The team found that Producer Records depend on an authoritative naming and definition of producers. In a university setting, creating such an authoritative list would be difficult to establish and maintain as departments change names and responsibilities so frequently. In order to effective use Producer Records many archives would have to rely, at least in part, on a university-wide identity management office to accurately identify and define all offices and departments at the institution.

---

[5] 2.1 Ingest Guide, p. 89.
[6] 2.1 Ingest Guide, p. 88.

List of Resources and Services Identified in the Ingest Guide and Maintain Guide

Resources described in 2.1 Ingest Guide	Abstract Services described in 3.2 Checklist of Fedora's Ability to Support Maintain Activities
Access Controls Policy	AIP Module
Accession Log	Data Backup Protocol
Activity Log	Data Management Database
Archival Information Package Configuration Rules	Alerting Service
Archive Naming/Identification Scheme	Format Transformation Service
Archives Directory	Format Validation Service
Collection Policy	Integrity Checking Protocol
Copyright Policy	Knowledge Base Module
Copyright Transfer/License	PDI Module
Designated Community Description	Persistent Identifier Manager
Format Representation Information System	Repository History
Format Standards Policy	Repository Stasis
Institutional Identity Management System	Request Service Manager
Metadata Encoding Rules	Retention and Disposition Module
Preservation System Capabilities Report	Search Service
Producer Record	Security Audit
Record Security Profile	Security Protocol
Record Type Record	Storage Management Module
Recordkeeping System Evaluation Tool	System Administration Protocol
Recordkeeping System Internal Rules	
Recordkeeping System Report	
Records Authority Statement	
Records Retention Policy	
Representation Information	
Submission Information Package Creation Procedures	
Survey Instrument	
Survey Procedures	
Transfer Procedures	
Transformation Policy	
Validation Procedures	

CONCLUSIONS

This grant project, "Fedora and the Preservation of University Electronic Records," has combined 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. Work on these three issues has allowed the project team to draw conclusions about the capability of archives and institutions to preserve electronic records, reengineering archival work to preserve electronic records, and the state of electronic records and recordkeeping research.

Preservation Capabilities
One of the key findings of the Tufts-Yale Project is that long-term preservation of archival university records is a difficult and costly endeavor. The Maintain Guide in particular gives a sense of the significant hardware, software, network, and personnel resources needed for simply maintaining electronic records. The Ingest Guide indicates the extensive policy and procedure development and commitment needed to develop and sustain a trustworthy ingest process. The Ingest Guide also describes the extensive range of resources needed to make that process scalable. Ingest and Maintain are just two of the activities needed for a successful preservation program—considerable additional work will be necessary to properly undertake the activities of preservation planning, access, and common services.[7]

Many—if not most—university archives and academic institutions (along with archives and institutions in other industries) that are responsible for preserving electronic records and other digital objects simply do not have the resources to establish and sustain their own trustworthy and scalable digital preservation program. Most archives will need to develop partnerships with other departments within its parent institution, peer archives and institutions, consortiums, or vendors in order to successfully preserve electronic records and digital objects. In addition, because the development of application tools, descriptive standards, and metadata schemas can represent a significant expenditure of effort, archives should look to employing existing tools and schemas—ideally ones that are standard, open, and widely supported by the appropriate communities.

For example, an Archive may contract with a commercial vendor to handle its maintain activities, particularly the sub-activities of data storage and back-up management, while it uses the services of a consortium repository to handle its access, data management, preservation planning, and part of its administration needs. In addition the Archive may employ metadata standards such as Dublin Core and METS in accordance with the rules of the consortium repository. However, this still leaves the Archive with a variety of responsibilities, such as creating and agreeing to a submission agreement with the producer, receiving the Submission Information Packages (SIPs) from the Producer, and ensuring the SIPs, content data-streams, and metadata data-streams are properly configured for submission to the consortium repository.

---

[7] ISO 14721:2003*: Space data and information transfer systems--Open archival information system--Reference model* (Geneva: International Organization for Standardization, 2003)

The fact that most archives cannot develop and sustain a trustworthy and scaleable preservation program by themselves should not be taken as a cue for archivists to do nothing. Archivists charged with preserving electronic records or digital object have a responsibility to do all that they can do even if that is not all that they need to do. For example, an Archive with few technical resources can still undertake a significant amount of essential policy work before finding a partner with the necessary technology. Archives must not be paralyzed into inaction.

Reengineering Archival Work
If archives are going to have any chance of preserving the increasingly complex and voluminous electronic records they are charged with preserving, archives are going to have to refocus their work away from processing and handling individual records and collections to managing the resources, abstracts services, tools, and policies that manage archival records in bulk. In short, archivists need to become a step removed from the records they manage if they are going to have any chance of preserving them. They are going to have to increasingly rely on semi-automated, regularized processes in their work. The Ingest Guide, for example, is geared towards enabling archives to take in records in a semi-automated and scaleable manner by helping them regularize and streamline many decisions-making steps. In addition, university archives could manage many of the resources described in the Guide as machine-readable objects. The more machine-readable resources a university archives has, the more it can automate its Ingest process.

These semi-automated, regularized ingest processes would remove university archives from directly handling records, manually arranging and describing them; work characterized as traditional "processing" work. This workflow is not scalable and cannot meet the challenges of electronic records.[8] Instead, archivists would spend the majority of time tending to their ingest policies and machine-readable resources, ensuring their continuing performance, making adjustments when necessary, and expanding their suite of resources to handle a broader range of records. Thus university archives would work at a policy and resource level that sits above the level of the records that they manage. Archivists would only dip down to manually handle individual or small groups of records that present exceptional issues or problems, and only when time and resources permit.

The Tufts-Yale Project also echoes the call many have made before: working with records creators and producers as they create their records and recordkeeping systems is essential to electronic records preservation. Many archives have traditionally accepted unorganized paper records with no descriptive information, which then forces them to spend considerable effort manually arranging and describing the records after—sometimes long after—the accession. This emphasis on rescuing disheveled records that come to the archives will not allow archivists to successfully preserve all of the records they need to preserve. First, electronic records are simply too voluminous and complex to reassemble their "order" and context after the fact. Second, nearly all electronic records sent in a disheveled state would need significant and immediate preservation work—a task that may be too burdensome for most archivists. Third, electronic records delivered to a university archives in a haphazard manner or after years of neglect have been, by definition, managed by the Producer in an untrustworthy manner. This severely

---

[8] This model did not meet the challenges of twentieth century paper records very well either. Mark A. Greene and Dennis Meissner, "More Product, Less Process: Revamping Traditional Archival Processing," *American Archivist*, Volume 68, Number 2, (Fall/Winter 2005).

jeopardizes Consumers' ability to presume the authenticity of those records—something that cannot be recovered by the university archivist, who can only maintain, not improve, the authenticity of the records it receives from a Producer.

The submission agreement described by the Ingest Guide provides a framework for university archives to help ensure that Producers properly prepare records for transfer to an Archive. This is designed to ensure that the Producer transfers the records to the Archive in an orderly fashion in the format and with the descriptive and contextual information that both the Archive and Producer deem necessary. In forcing a university archives to carefully articulate the terms of transfer, the submission agreement encourages the Archive to work closely with Producers. This working relationship would, ideally, enable archivists to communicate to Producers the requirements for trustworthy records systems and influence Producers' recordkeeping practices. This would involve a shift in the focus of archival work away from arrangement and description or processing of records and towards systems analysis and business process analysis.[9] This shift would entail the archival community changing its traditional skill set.[10]

Electronic Records Research
The project team had considerable difficulty in attempting to find an appropriate framework for its recordkeeping requirements in Requirements for Trustworthy Recordkeeping and Preservation. After giving careful consideration to fitting the requirements within the structure of *Trusted Digital Repositories: Attributes and Responsibilities*, the project team settled on using the "Record management processes and controls" section of ISO 15489-1: 2001, *Information and documentation – Records management – Part 1: General*, as the requirements' framework. While ISO 15489 gave us a satisfactory conceptual framework upon which to shape the project's recordkeeping requirements, there appears to be no consensus within the university records community for a framework for recordkeeping system requirements as the OAIS Reference Model has become the consensus framework for digital preservation requirements. Developing such a consensus framework would help enable institutions to make better sense and use of the recordkeeping requirements literature that has emerged in the 1990s and 2000s.

---

[9] Kenneth Thibodeau, "Archival Science and Archival Engineering: Building a New Future for the Past," *Archival Outlook*, (May/June 2006).

[10] Richard Pearce-Moses, "President's Message," *Archival Outlook*, (September/October 2005, January/February 2006, May/June 2006, July/August 2006).