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/

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

TABLE OF CONTENTS

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.

Fedora and the Preservation of University Records Project

1.2 System Model

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/

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:UA069.004.001.00002
(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

TABLE OF CONTENTS

Issues with the Preservation of Electronic University Records

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 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 creation, and pervades every single recordkeeping activity.

Authenticity is the trustworthiness of the record as a record—that the record is what it purports to be and has not been tampered with or corrupted in essential respects. A person cannot automatically presume the authenticity of an electronic record; he or she must weigh the evidence that the record either is or is not what it purports to be and either has or has not been modified or corrupted in essential respects—and then judge whether the record is authentic or not. Authenticity is not a component of a record but the judgment a person makes about a record. When reports from this project refer to “authentic records” or “authentic electronic records” it is shorthand for records that a reasonable person would judge as authentic. One cannot judge the authenticity of a recordkeeping or preservation system, only its trustworthiness.

In order to be able to reasonably judge a record as authentic, one must be able to establish its identity and demonstrate its integrity. One must ensure that electronic records are clearly identifiable, of demonstrable integrity, and that accidental corruption or purposeful tampering has not occurred since they were created and set aside. One can accomplish this by maintaining the records in a trustworthy records system. A trustworthy records system ensures the preservation of a record’s identity and integrity, protecting it from corruption and tampering. Therefore, a reasonable person can presume that a record created/captured and managed in a trustworthy records system is authentic.

The archival, records, and information management communities have used the terms electronic recordkeeping system in a number of ways. Some conceive of a recordkeeping system broadly as the entire framework of recordkeeping from creation to preservation and access, while others

conceive of it more narrowly, describing the specific computer application tasked to store and manage records. In addition, the elements that make up a recordkeeping system and the factors (controls) that influence recordkeeping have been described interchangeably. The imprecise use of the term electronic recordkeeping system in documents articulating requirements for such systems may have hindered records professionals' efforts to turn these documents into evaluation tools and detailed implementation and application development guidelines. In order to alleviate this problem, this project has explicitly defined a trustworthy electronic recordkeeping system and the composition of its elements. In addition it has also differentiated between the components of a records system and the records controls that influence and shape an institution's recordkeeping activities.

TRUSTWORTHY ELECTRONIC RECORDKEEPING SYSTEM

A trustworthy electronic recordkeeping system is the combination of all of the records components—people (natural and juridical), institutions, applications, infrastructure, and procedures—necessary for records to be created, collected, organized, and categorized to facilitate their preservation, retrieval, use, and disposition in a manner that provides a circumstantial probability of the authenticity of those records and a likelihood that a reasonable person would judge those records as authentic. An institution uses a combination of records components to help it meet its records needs and expectations that are articulated or manifested in requirements, policies, responsibilities, and practices.

INFLUENCE OF THE LIFECYCLE MODEL

The research work of this project is based largely on the conceptual underpinnings of the records lifecycle model, presuming that a Producer will create, acquire, use, and manage records in a recordkeeping system to suit its current business needs, and later the Archive will ingest some of those records into a separate preservation system that the Archive administers. In this model, the Archive acts as a neutral third party in the recordkeeping process, acting first and foremost on behalf of broader societal needs rather than on behalf of the Producer. As a neutral third party, the Archive has no stake in the content of the records and no reasons to alter records in its custody, and it should not allow anybody to alter the records either accidentally or on purpose. Many archivists have rejected the lifecycle model in favor of the records continuum concept, where recordkeeping is seen as a continuous process that is not time-based, separated into a series of clearly defined steps, or administered by completely separate juridical entities. Many Producers and Archives operate in a mixed world between these two models. For example, many Archives operate separately from a Producer, are part of the same organization as the Producer, and do not act as a neutral third party.

These two types of records systems (recordkeeping and preservation) are distinguished from one another in this project because such a distinction helps best describe the records environment of many colleges and universities in the United States. However, separate recordkeeping and records preservation systems are not necessary to make up a trustworthy records system at a college or university. Recordkeeping systems, because they must preserve records in order to "keep" them, naturally fulfill many of the same requirements placed upon records preservation systems. The opposite is not true. A trustworthy records preservation system cannot exist without a trustworthy recordkeeping system. It is not possible for a records preserver to preserve authentic records if such records were not created and managed in trustworthy recordkeeping systems fulfilling the recordkeeping requirements.[1]

[1] It would be possible for a preserver to preserve authentic copies of records whose authenticity might be in question. For example, the preserver can not alter a forgery so that it can be presumed authentic, but the preserver can faithfully preserve authentic copies of that forgery.

Fedora and the Preservation of University Records Project

1.3 Concerns

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/

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:UA069.004.001.00003
(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

Table of Contents

Overview

This project makes a set of intellectual assumptions, or overall concerns, that inform all of its findings. Every requirement and step described in the Requirements for Trustworthy Recordkeeping and Preservation the Ingest Guide, and the Maintain Guide implicitly carry with them the following nine concerns: Audit, Authorization, Automation, Compliance, Documentation, Financial Sustainability, Metadata, Reporting, and Training. In order to meet the requirements or undertake the steps of the guides in a trustworthy manner, an Archive or institution must address all nine concerns.

AUTHENTICITY

The goal of 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.

Authenticity is the trustworthiness of the record as a record—that the record is what it purports to be and has not been tampered with or corrupted in essential respects. A person cannot automatically presume the authenticity of an electronic record; he or she must weigh the evidence that the record either is or is not what it purports to be and either has or has not been modified or corrupted in essential respects—and then judge whether the record is authentic or not. Authenticity is not a component of a record but the judgment a person makes about a record. When reports from this project report refer to "authentic records" or "authentic electronic records" they are using shorthand for records that a reasonable person would judge as authentic.

In order to be able to reasonably judge a record as authentic, one must be able to establish its identity and demonstrate its integrity. One must ensure that electronic records are clearly identifiable, of demonstrable integrity, and that accidental corruption or purposeful tampering has not occurred since they were created and set aside. One can accomplish this by maintaining the records in a trustworthy electronic records system. A trustworthy records system ensures the preservation of a record's identity and integrity, protecting it from corruption and tampering. Therefore, a record created/captured and managed in a trustworthy records system can be presumed to be authentic. A person judges the trustworthiness of records systems and the authenticity of records.

Several recordkeeping and preservation requirement sets developed primarily by the archival and records management professions over the past two decades have specific requirements devoted to authenticity. However, authenticity—creating the likelihood that a reasonable person would judge records as authentic—is the goal of a records system; it is not a required attribute or activity of a system. Therefore, none of this project's reports explicitly discuss authenticity as a specific requirement or step, instead it is the implicit, core aim, and purpose of every requirement and step.

List of Concerns

Below are concerns that permeate every aspect of recordkeeping and preservation in an effort to reach the ultimate goal of reproducing authentic copies of records. Many recordkeeping and preservation requirements developed in the last twenty years express these concerns—especially audit, compliance, and metadata—as specific requirements related to but distinct from other requirements. However, unlike requirements or steps related to relatively discrete functions (such as records capture), requirements like audit, compliance, and metadata are not distinct functions, they are instead concerns that permeate every function. For example, records capture, disposition, and delivery are all functions that must be auditable.

Audit

Every action taken in a records system to create, collect, organize, categorize, maintain, preserve, retrieve, use, or execute the disposition of a record must be auditable. This means that every action must produce an account of itself that an external entity can audit, and the records system must support a process that can execute audits. A records system must ensure that actions performed on records, their metadata, and the system itself are auditable. Institutions must keep unalterable audit trails and preserve those audit trails for as long as the appropriate auditor requires them for review. Audits should reveal information on the nature of the action, the entity undertaking the action, and the time of the action.[1]

Authorization

Every action taken in a records system to create, collect, organize, categorize, maintain, preserve, retrieve, use, or execute the disposition of a record must be undertaken by a person or unit within an institution, or a designee of the institution, who has the authority to undertake that action. This includes a person's or entity's authorization to view records (read access) and a person's or entity's authorization to take actions upon records (write access). A person's or entity's authorization to view or take actions upon records is based on a person's or entity's rights, security clearance, position within an organization or society, or training. For example, a citizen has the right—the authorization—to view non-classified government records because of various government laws. A person's or entity's authorization to view records and in particular take actions upon records also depends on policies, procedures, and the state of records. For example, a person who is authorized to confidentially destroy certain records is only authorized to destroy those records when they have exceeded their retention period and they have had the proper final review. Institutions need to document, monitor, enforce, and update their documentation of authorizations, which can become quite complex.

Automation

Many actions taken in a records system to create, collect, organize, categorize, maintain, preserve, retrieve, use, or execute the disposition of a record must be able to be undertaken in a automated manner that would enable an institution to implement the action in a scalable

[1] Recordkeeping requirement sets that extensively discuss auditing include Indiana University, *Requirements for Electronic Records Management Systems*; *Model Requirements to Ensure the Creation, Maintenance, and Preservation of Electronic Records*, prepared for the IDA Programme of the European Commission by Cornwell Affiliates; and *Design Criteria Standard for Electronic Records Management Software Applications.*

production workflow. Archives and institutions cannot automate every action, but the more they can automate, the more they can feasibly manage the daunting volume of electronic records many institutions and Archives face. Strictly speaking, automation is not absolutely required for a trustworthy records system. However, most electronic records systems handle such a large volume of records that an entirely non-automated system could not keep up with the sea of records it is responsible for. The institution could only manage or preserve a small percentage of records in a trustworthy manner while leaving the remaining records unattended.

Compliance

All recordkeeping and preservation activities are undertaken within the context of a legal, regulatory, and administrative environment. Institutions must identify, track changes to, and comply with the laws, regulations, standards, best practices, and professional ethics that affect its recordkeeping activities.[2] People must understand the laws, regulations, standards, best practices, and professional ethics that affect their recordkeeping activities.[3] Recordkeeping applications must not include any features that do not comply with the laws, regulations, standards, best practices, and professional ethics that affect the recordkeeping activities of the institution that the application serves.[4] Institutions should be able to demonstrate their compliance with the laws, regulations, standards, best practices and professional ethics that affect its recordkeeping activities.[5]

Documentation

All records systems need documentation that describes how to execute every action taken in a records system to create, collect, organize, categorize, preserve, retrieve, use, or execute the disposition of a record. In essence, institutions need to create written procedures for all of their substantive recordkeeping and preservation activities. Institutions must determine the appropriate detail and retention of their recordkeeping and preservation documentation. This concern does not include the creation of audit trails of individual recordkeeping and preservation actions, which many research projects refer to as documentation.

Financial Sustainability

Every records system must be financially sustainable if it is to persist long enough to fulfill its recordkeeping or preservation goals. An institution needs to follow sound business practices and have long-term plans in place supported by short- and long-term financial planning. Ensuring preservation of digital resources requires substantial and ongoing financial commitments over time—potentially more so than for traditional records. Electronic records preservation is dynamic; responses to technological obsolescence or media decay must be taken quickly and the life expectancy of a preservation treatment is short because the technologies utilized evolve rapidly. Consequently, preservation strategies must be periodically monitored and reassessed as the technological environment that supports standards, protocols, and formats, etc. evolves.

---

[2] University of Pittsburgh, *Functional Requirements for Evidence in Recordkeeping* 1a, 1a1-3, 1c; Indiana 1.l, 1.1.1; MoReq 11.4, 11.5, 11.5.2-3, 11.5.5; Public Record Office, *Functional Requirements for Electronic Records Management Systems* A.10.1, A.10.2; *ISO 15489-1: Information and documentation – Records management* 5, 5a-e, 7.1.h, 8.2.4.
[3] ISO 8.2.4.
[4] MoReq 11.5.4.
[5] Pitt 1; ISO 5, 5a-e, 8.2.4.

Institutions must be able to bear the financial cost of any recordkeeping or preservation activity to ensure that it can devote to that activity the resources needed to ensure its trustworthiness.[6]

Metadata
Many of the actions taken in a records system to create, collect, organize, maintain, categorize, preserve, retrieve, use, or execute the disposition of a record along with the audit trails and documentation of these actions manifest themselves as metadata. Encoding actions as metadata is critical for enabling records systems to have regularized, machine-readable, automated, and scalable workflows. Metadata is critical for documenting actions for audit purposes and generating timely and accurate reports for records system administrators and managers.[7]

Reporting
Records systems must be able to produce reports on most of the actions taken to create, collect, organize, maintain, categorize, preserve, retrieve, use, or execute the disposition of a record for records systems administrators and managers. These reports come in many forms and may be machine-readable or human-readable, delivering information to a service or person. Reports may describe individual actions or aggregate many actions. Reporting capabilities enable managers and administrators to manage records systematically, monitor records and usage, detect problems such as data failure or unauthorized access, and plan resource allocation and preservation strategies.

Training
Every person undertaking an action in a recordkeeping or preservation system to create, collect, organize, categorize, maintain, preserve, retrieve, use, or execute the disposition of a record must have the appropriate training needed to execute the action successfully. A person must be trained to a level that allows that person to undertake an action in a records system in a trustworthy manner. A person's breadth and level of training may be closely tied to his or her authorization to perform recordkeeping or preservation activities.

[6] *Trusted Digital Repositories: Attributes and Responsibilities* (Mountain View, CA: RLG, 2002) and *An Audit Checklist for the Certification of Trusted Digital Repositories*, Draft for Public Comment (Mountain View, CA: RLG, 2005).
[7] Several recordkeeping requirement documents have significant requirements concerning metadata, including, Indiana; MoReq; PRO; and DoD 5051.2. The *Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group* came out in 2005 as a set of preservation metadata.

Fedora and the Preservation of University Records Project

1.4 Glossary

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/

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:UA069.004.001.00004
(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

TABLE OF CONTENTS

GLOSSARY GUIDE

This glossary has definitions of terms used throughout the project documents. To the degree possible, the project team has tried not create its own definitions of key terms when sufficient definitions could be found in an existing glossary or dictionary. Because there is a wide range of stakeholders involved in electronic records preservation, the project team has utilized the terminology from the OIAS Reference Model as much as possible. The Reference Model is aimed at a range of different stakeholder communities that are involved in the preservation of information and records in digital form. The OAIS terms do not exactly match all of those terms familiar to any particular discipline (e.g., traditional archives, digital libraries, science data centers). Instead, OAIS uses terms that are not already overloaded with meaning to reduce conveying unintended meanings across stakeholder communities.

Terms citing the Tufts-Yale project as its source are developed by the project team.

Each term in this glossary has four elements:

Term: the term used in the project documents.
Definition: the standard definition utilized by the project team.
Source: the source of the standard definition.
Project Reports: the primary reports that use the term.

The primary sources for the definitions are:

ISO 14721:2003, Space data and information transfer systems -- Open Archival Information System -- Reference model. <http://public.ccsds.org/publications/archive/650x0b1.pdf>.
Cited as: *ISO 14721:2003*

InterPARES2 Terminology Database <http://www.interpares.org/ip2/ip2_terminology_db.cfm>.
Cited as: *InterPARES2*

Richard Pearce-Moses, *A Glossary of Archival and Records Terminology* (Archival Fundamentals Series II) Chicago: Society of American Archivists, 2005.
<http://www.archivists.org/glossary/>
Cited as: *SAA Glossary*

Oxford Univeristy Press, *Oxford English Dictionary*, 3rd edition, 2006. <http://www.oed.com>

Wikipedia: The Free Encyclopedia <http://en.wikipedia.org/wiki/Main_Page>
Cited as: *Wikipedia*

Computer Desktop Encyclopedia. Computer Language Company Inc., 2006. Answers.com.
<http://www.answers.com/topic/infrastructure>
Cited as: *Computer Encyclopedia*

TERMS

Access
The OAIS entity that provides the services and functions that support Consumers in determining the existence, description, location and availability of information stored in the OAIS, and allows Consumers to request and receive information products. Access functions include communicating with Consumers to receive requests, applying controls to limit access to specially protected information, coordinating the execution of requests to successful completion, generating responses (Dissemination Information Packages, result sets, reports) and delivering the responses to Consumers.
ISO 14721:2003
Requirements for Recordkeeping and Preservation

Administration
The OAIS entity that provides the services and functions for the overall operation of the archive system. Administration functions include soliciting and negotiating submission agreements with Producers, auditing submissions to ensure that they meet archive standards, and maintaining configuration management of system hardware and software. It also provides system engineering functions to monitor and improve Archive operations, and to inventory, report on, and migrate/update the contents of the Archive. It is also responsible for establishing and maintaining archive standards and policies, providing customer support, and activating stored requests.
ISO 14721:2003
Maintain Guide

Administration Metadata Store
A logical storage area where AIPs and PDI are stored. This is generally a hard disk in the Preservation Application Hardware Environment, and represents a fairly high per-unit storage cost.
ISO 14721:2003
Maintain Guide

Application
(see Recordkeeping Application and Preservation Application)

Archival Information Package (AIP)
An aggregation of records components and metadata, or information package, consisting of the Content Information and the associated Preservation Description Information (PDI), which is preserved within an Archive.
ISO 14721:2003
Maintain Guide

Archival Storage
The OAIS entity that contains the services and functions used for the storage and retrieval of Archival Information Packages (AIP). Archival Storage functions include receiving AIPs from Ingest and adding them to permanent storage, managing the storage hierarchy, refreshing the

media on which archive holdings are stored, performing routine and special error checking, providing disaster recovery capabilities, and providing AIPs to Access to fulfill orders.
ISO 14721:2003
Maintain Guide

Archive
An organization of people and systems that has accepted the responsibility to preserve information and make it available for a Designated Community. It meets a set of responsibilities defined in ISO 14721:2003. Such an Archive may be distinguished from other uses of the term archive or archives.
ISO 14721:2003
Maintain Guide

Authenticity
Authenticity is the trustworthiness of the record as a record—that the record is what it purports to be and has not been tampered with or corrupted in essential respects. A person cannot automatically presume the authenticity of an electronic record; he or she must weigh the evidence that the record either is or is not what it purports to be and either has or has not been modified or corrupted in essential respects—and then judge whether the record is authentic or not.
Based in part on InterPARES2
Concerns

Checksums
A checksum is a form of redundancy check, a very simple measure for protecting the integrity of data by detecting errors in data that is sent through space (telecommunications) or time (storage). It works by adding up the basic components of a message, typically the asserted bits, and storing the resulting value. Later, anyone can perform the same operation on the data, compare the result to the checksum, and (assuming that the sums match) conclude that the message was probably not corrupted. There are more sophisticated types of integrity or redundancy checks, including Fletcher's checksum, Adler-32, cyclic redundancy checks (CRCs), and cryptographic hash functions, such as SHA-256 are designed to address these weaknesses by considering not only the value of each byte but also its position. However, these integrity checks are commonly referred to as checksum in the reports of this project.
Wikipedia
Maintain Guide

Consumers
The role played by those people, or client systems, who interact with the services of the Archive to find preserved records of interest and to access those records in detail. This can include other Archives, as well as people or systems within the same institution.
ISO 14721:2003
Requirements for Recordkeeping and Preservation; Maintain Guide

Cryptographic Checksum or Cyptographic Hash Function
A hash function with certain additional security properties to make it suitable for use as a primitive in various information security applications, such as authentication and message integrity. A hash function takes a long string (or message) of any length as input and produces a fixed length string as output, sometimes termed a message digest or a digital fingerprint. It is less vulnerable to attack than Checksums and cyclic redundancy checks.
Wikipedia
Maintain Guide

Cyclic Redundancy Check
A cyclic redundancy check (CRC) is a type of hash function used to produce a CRC checksum against a block of data, such as a packet of network traffic or a block of a computer file.
Wikipedia
Maintain Guide

Data Management
The OAIS entity that contains the services and functions for populating, maintaining, and accessing a wide variety of information. Some examples of this information are catalogs and inventories on what may be retrieved from Archival Storage, processing algorithms that may be run on retrieved data, Consumer access statistics, security controls, and schedules, policies, and procedures.
ISO 14721:2003
Maintain Guide

Designated Community
An identified group of potential Consumers who should be able to understand a particular set of information. The Designated Community may be composed of multiple user communities.
ISO 14721:2003
Maintain Guide

Digital Object
A unit of digital information that includes properties of the object and may also include methods of performing operations on the object. This is the basic unit for information aggregation in Fedora. In Fedora, a digital object has consist of at least two elements: (1) an identifier or PID, that provides the key by which the digital object is accessed from the repository, and (2) Dublin Core metadata that provides a basic description of the digital object.
InterPARES2, Tufts-Yale
Maintain Guide

Dissemination Information Packages (DIP)
The aggregation of records components and metadata, or information package, derived from one or more AIPs, received by the Consumer in response to a request to the Archive.
ISO 14721:2003
Maintain Guide

Hot Spare
An extra disk drive in a RAID (redundant array of independent disks) configuration that is ready and waiting to be put into action automatically when another drive fails. Using the RAID algorithms, the missing data from the faulty drive is reconstructed and written to the hot spare. When the bad drive is replaced, it then becomes the hot spare. If a hot spare is not used, then the faulty drive must be manually removed and replaced with a new one.
Wikipedia
Maintain Guide

Infrastructure
The fundamental structure of a system or organization. The basic, fundamental architecture of any system (electronic, mechanical, social, political, etc.) determines how it functions and how flexible it is to meet future requirements; refers to system and development programs in contrast to applications. A computer system's infrastructure would include the operating system, database management system, communications protocols, compilers and other development tools.
Computer Encyclopedia
Requirements for Recordkeeping or Preservation

Ingest
The OAIS entity that provides the services and functions to accept Submission Information Packages (SIPs) from Producers (or from internal elements under Administration control) and prepare the contents for storage and management within the archive. Ingest functions include receiving SIPs, performing quality assurance on SIPs, generating an Archival Information Package (AIP) which complies with the Archive's data formatting and documentation standards, extracting Descriptive Information from the AIPs for inclusion in the archive database, and coordinating updates to Archival Storage and Data Management.
ISO 14721:2003
Maintain Guide

Institution
An establishment, organization, or association, instituted for the promotion of some object, e.g. a church, school, college, hospital, asylum, reformatory, mission, or the like; an established organization. An Institution creates and uses records to conduct its business. For this document it refers specifically to colleges and universities, although institutions in other industries may adopt this document.
Tufts-Yale
Requirements for Recordkeeping or Preservation

Juridical System
A collectivity organized on the basis of a system of rules; a social group that is organized on the basis of a system of rules and that includes three components: the social group, the organizational principle of the social group, and the system of binding rules recognized by the social group.
InterPARES2
Requirements for Recordkeeping and Preservation

Juridical Person
An entity having the capacity or the potential to act legally and constituted either by a position (a succession) or collection (an organization) of natural persons; role(s) taken by natural people; e.g. university archivist or university archives.
InterPARES2
Requirements for Recordkeeping and Preservation

Knowledge Base
A set of information, incorporated by a person or system, that allows that person or system to understand received information.
ISO 14721:2003
Maintain Guide

Natural Person
An individual human being, as distinguished from a corporate body, representative, or juridical person.
Tufts-Yale
Requirements for Recordkeeping and Preservation

Policy
An articulation of the goals and aims of the institution; a principle or course of action adopted or proposed as desirable, advantageous, or expedient; esp. one formally advocated by an institution or government. Policies exist as documents.
Oxford English Dictionary
Requirements for Recordkeeping and Preservation, Ingest Guide

Preservation
The act of physically and intellectually protecting and technically stabilizing 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.
Based in part on SAA Glossary
All documents

Preservation Application
The software package (or collection of software) which an Archive uses as part of its preservation system. This is intended to cover only the software actively involved in the management of the preservation system and not utility or operating system software.
Tufts-Yale
Maintain Guide

Preservation Application Hardware Environment
The hardware platform upon which the preservation application runs. This may be a single computer or a group of cooperating computers. Because of the high need for reliability, this hardware should be dedicated to the preservation application and should not participate in other functions. This hardware likely contains the Administrative Data Store primary media, probably in the form of hard disks. It may also contain the Records Components Store primary media.

Tufts-Yale
Maintain Guide

Preservation Description Information
The information which is necessary for adequate preservation of the Content Information and which can be categorized as Provenance, Reference, Fixity, and Context information.
ISO 14721:2003
Maintain Guide

Preservation Planning
The OAIS entity that provides the services and functions for monitoring the environment of the OAIS and providing recommendations to ensure that the information stored in the OAIS remains accessible to the Designated User Community over the long term, even if the original computing environment becomes obsolete. Preservation Planning functions include evaluating the contents of the archive and periodically recommending archival information updates to migrate current archive holdings, developing recommendations for archive standards and policies, and monitoring changes in the technology environment and in the Designated Community's service requirements and Knowledge Base. Preservation Planning also designs information package templates and provides design assistance and review to specialize these templates into SIPs and AIPs for specific submissions. Preservation Planning also develops detailed Migration plans, software prototypes and test plans to enable implementation of Administration migration goals.
ISO 14721:2003
Maintain Guide

Procedure
Procedures articulate the actions required to successfully complete a task. Procedures also describe how to execute those actions. Procedures articulate how a Juridical System will fulfill its policy requirements. Procedures exist as documents.
Tufts-Yale
Requirements for Recordkeeping and Preservation

Producer
The role played by the people, corporate bodies, administrative units, families, or client systems, who provide the records to be preserved; equivalent to donor, often, but not always the records creator (the physical or juridical person who makes, receives, and/or accumulates records by reason of its mandate/mission, functions or activities).
ISO 14721:2003
Maintain Guide

Recordkeeping Application
Any electronic program that creates and/or imports and maintains, stores, and distributes electronic records.
Tufts-Yale
Requirements for Recordkeeping and Preservation

Recordkeeping Component
The people, institutions, applications, infrastructure, and procedures necessary for records to be created, collected, organized, and categorized to facilitate the records' preservation, retrieval, use, and disposition.
InterPARES2
Ingest Guide, Maintain Guide

Recordkeeping Infrastructure
(See Infrastructure)

Recordkeeping Institution
(See Institution)

Recordkeeping Natural Person
(See Natural Person)

Recordkeeping Policy
(See Policy)

Recordkeeping System
A manual or automated system in which records are created, collected, organized, and categorized to facilitate their preservation, retrieval, use, and disposition. Recordkeeping Components compose a recordkeeping system.
Tufts-Yale
Requirements for Recordkeeping and Preservation

Records Component
A digital object that is part of one or more electronic records, including any metadata necessary to order, structure, or manifest the content, requiring a given preservation action.
InterPARES2
Maintain Guide

Records Component Store
A logical storage area where records components are stored. This would probably be a dedicated storage environment separate from the Preservation Application Hardware Environment, and may represent a lower per-unit storage cost than the Administration Metadata Store.
Tufts-Yale
Maintain Guide

Stasis
Setting the Preservation Application in a state that allows only system administrators to make changes application settings and data.
Tufts-Yale
Maintain Guide

Storage Hardware Environment
The hardware platform that is responsible for the storage functions of the Administrative Metadata Store or the Records Components Store. This can be a network appliance, a storage virtualization network, a tape robot, or some other storage subsystem. The Storage Hardware Environment also includes any networking or software components needed for the storage function to operate. The Records Components Store is usually implemented using a Storage Hardware Environment separate from the Preservation Application Hardware Environment, whereas the Administrative Metadata Store may or may not be.
ISO 14721:2003
Maintain Guide

Submission Information Packages
The aggregation of records components and metadata, or information package that is delivered by the Producer to the OAIS for use in the construction of one or more AIPs.
ISO 14721:2003
Ingest Guide

Trustworthiness
The quality of being dependable and reliable.
SAA Glossary
Requirements for Recordkeeping and Preservation

Trustworthy Electronic Recordkeeping System
The combination of all the recordkeeping components (people, institutions, applications, infrastructure, and processes) necessary for records to be created, collected, organized, and categorized to facilitate their preservation, retrieval, use, and disposition in a manner that provides a circumstantial probability of the authenticity of those records.
InterPARES2
Requirements for Recordkeeping and Preservation

Fedora and the Preservation of University Records Project

1.5 Requirements for Trustworthy Recordkeeping Systems and the Preservation of Electronic Records in a University Setting

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/

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:UA069.004.001.00005
(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

TABLE OF CONTENTS

I. Introduction

This report presents the results of an effort to illustrate trustworthy electronic recordkeeping and preservation at a college or university. It describes the necessary features, behaviors, and qualities of recordkeeping systems and preservation activities which may take place in recordkeeping systems or separate preservation systems. These requirements are articulated as two separate sets of functional requirements, one for recordkeeping and another for preservation.

Although the focus of this research project was aimed primarily at preservation, the project team felt it was necessary to establish a set of recordkeeping system functional requirements. Before designing new recordkeeping systems or evaluating existing systems for active records creation and management, it is necessary to define what types of functionality such a system should possess in order to facilitate preservation, whether that be undertaken inside the recordkeeping system or after transfer of the records to a separate preservation system.

This is not the first set of recordkeeping system functional requirements. In fact, these requirements are primarily a synthesis of recordkeeping requirements written by other groups in the 1990s and early 2000s into a single set that is appropriate for a university setting. There have been a number of significant efforts to define recordkeeping functional requirements, including national and international standards creating bodies. However, the project team believed it necessary to create a new set of recordkeeping requirements, because, "no one list [of functional requirements] has been endorsed by the [university archival] profession."[1] There are a number of reasons for this lack of profession-wide acceptance. The existing literature does not agree on terminology. The sets of requirements differ, in some cases significantly. The field of digital preservation had not reached full maturity during this period (some might argue that the field still has not yet reached full maturity). Many issues remained unexplored as each project began. The recordkeeping functional requirements literature from the period focused on specific aspects of digital perseveration, like records management software specifications, warrant, or policy development. This was all very important work, but each did not always consider the full scope of digital preservation or build upon the work of its predecessors. In addition, the majority of this work has been undertaken primarily by government recordkeeping professionals with the creation and maintenance of the government records in mind. This set is aimed at university records. This work addresses current needs of the college and university recordkeeping by leveraging the expertise from all the most significant projects over the last fifteen years.

A description of the necessary attributes of the activity of preservation in a college or university is necessary in order to preserve electronic records and ensure their continued accessibility and authenticity over time as the creating technologies become obsolete. In addition such a set of requirements is necessary in order to evaluate existing recordkeeping or preservation systems or to plan and/or build new systems capable of facilitating long-term preservation of authentic electronic records. Unlike the situation for recordkeeping, there has not been an extensive body of literature specifically on the functional requirements for preservation of electronic records. There is no single set of preservation requirements; no international standard. However, the project team found that many of the requirements necessary for long-term preservation are

[1] Philip Bantin, "Functional Requirements for Recordkeeping Systems – Evolution of the IU Functional Requirements," 2002 <http://www.indiana.edu/~libarch/ER/nhprcfinalfuncreq.doc>.

included in the existing recordkeeping requirements literature. This is because some of the same activities necessary to keep records accessible in active recordkeeping systems are also necessary to preserve records over the long term. As a result, many of the functional requirements of preservation activities have been gleaned from recordkeeping requirements literature.

The project team developed the separate sets of functional requirements for recordkeeping and preservation because recordkeeping and preservation environments are separately administered respectively by Producers and archivists (preservers) at both Tufts and Yale. Thus the project team is biased to presume that a Producer will create, acquire, use, and manage records in a recordkeeping system to suit its current business needs. While the central purpose of a recordkeeping system is to support the business needs of its Producer, the central purpose of preservation activities is to preserve records. In a pure records lifecycle model environment, an Archive will later ingest some records from a recordkeeping system into a separate preservation system (Archive) that the preserver administers, undertaking preservation activities in this system. In such a situation, the Producer would be responsible for meeting the recordkeeping requirements and the Archive would be responsible for meeting the preservation requirements. However, it may also be possible that this may be a continuous process where records do not move from a recordkeeping system to a separate preservation system administered by completely separate juridical entities. In this case, preservation activities will need to take place in the recordkeeping system if the electronic records are to persist over the long term. In this case, the recordkeeper is also acting as the preserver and should meet both the recordkeeping and the preservation requirements. This document should also apply in such situations.

The purpose of this report is not to serve as a records management or preservation application software specification. These functional requirements describe the entire recordkeeping system, not only the application. Thus it is best utilized by those who would benefit from a more holistic view of recordkeeping and preservation, like resource allocators and administrators responsible for those that buy, build, or manage recordkeeping or preservation systems, as well as archivists and records managers.

II. Form of the Requirements

The requirements for a recordkeeping system described in Section IV, or for preservation described in Section V, are for either the *Application* itself or for the *Natural or Juridical People, Institution, Procedure*, or *Infrastructure*. No requirements are expressed as requirements for *Recordkeeping System*. As *Records Controls* themselves impose requirements on records systems, the document does not include requirements for any *Controls*.

Each requirement includes only one of the six records system components. While some requirements may pertain in some way to multiple components, every requirement in this report only contains the most relevant component.

The requirements in this report are organized into two different chapters, one for recordkeeping system requirements and the other for records preservation requirements. The recordkeeping chapter is organized into six sections based loosely on the framework presented in the Records management and controls section of ISO 15489-1: *Information and documentation—Records management.*[2]:

1. Record Retention and Disposition
2. Records Capture and Registration
3. Classification
4. Storage and Handling
5. Access
6. Design and Performance

These six sections are further broken down into a total of fourty subsections that the Tufts-Yale project team created from the existing requirements. There was no attention paid to possible subsection topics that might exist if there were no recordkeeping requirements identified on that subject in the existing literature. Because of the strength of the existing literature, the project team did not make any effort to create requirements not found in the literature.

The records preservation requirements chapter is organized into seven sections, with the requirements grouped according to the six functional entities of the OAIS Reference Model, in addition to Common Services:

1. Common Services
2. Ingest
3. Archival Storage
4. Data Management
5. Administration

[2] ISO 15489-1: 2001, *Information and documentation – Records management – Part 1: General.* During this process the project team also gave careful consideration to mapping the recordkeeping requirements to *Trusted Digital Repositories: Attributes and Responsibilities*, but ultimately decided that the Records management processes and controls section of ISO 15489-1 was a better fit for the requirements. Organizing the requirements according to ISO 15489 gave the project team an existing conceptual framework upon which it could shape the requirements. It appears that that there is no consensus or preferred framework for recordkeeping system requirements in the way the OAIS Reference Model appears to play that role in describing a framework for preservation requirements.

6. Preservation Planning
7. Access

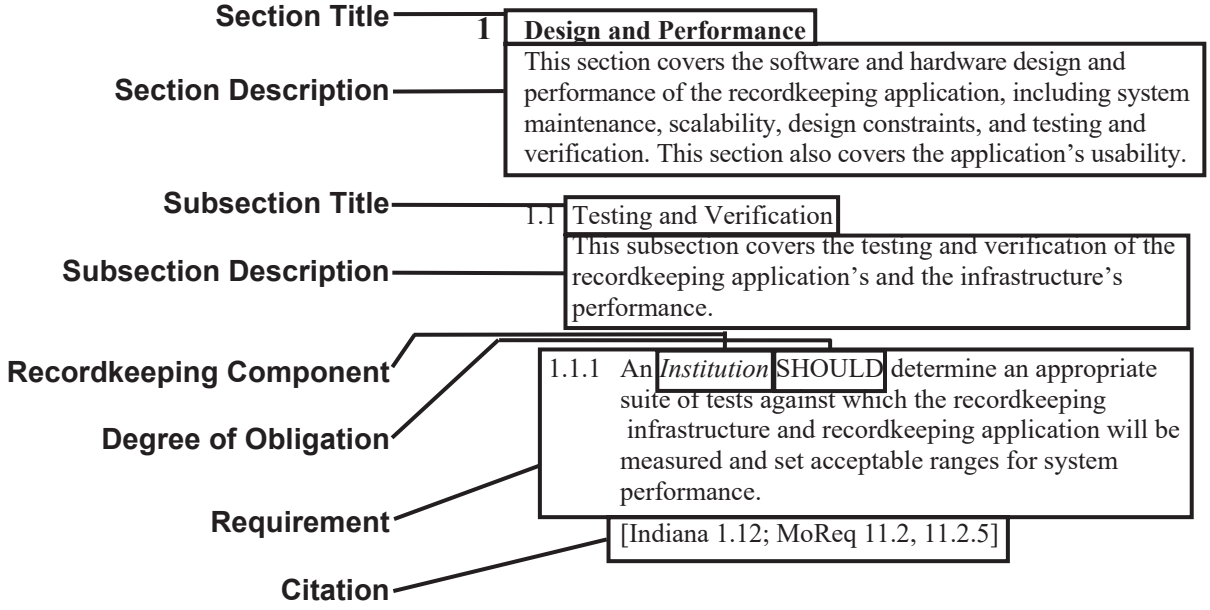
These seven sections are broken down into thirty-four subsections using the corresponding sub-functions of the OAIS model. In the case of the preservation requirements, the project team gave attention to subsection topics for which the requirements literature did not identify functional requirements in any obvious way. Because of the limited nature of the existing literature, the project team attempted to augment requirements found in the literature by creating its own requirements.[3] This was done only in cases where an analysis of the existing literature did not yield requirements the project team deemed necessary for preservation.

As stated above, every section has a number of different subsections. All of the requirements are nested within these subsections. The section and subsections have brief descriptions. These descriptions are not requirements; rather they are explanations defining the nature and scope of each section and subsection.

Each requirement will have a sequential number, the text of the requirement itself, and a citation to one or more of the research projects listed at the beginning of both requirement chapters. The text of the requirement itself will contain one of the five Records Components (Application, Infrastructure, Institution, Natural or Juridical People, or Procedures), a degree of obligation (MUST, MUST NOT, SHOULD, SHOULD NOT, or MAY), and then a description of the actual requirement. (See Figure 1)

[3] A primary reason for this need to identify additional requirements stems from the fact that the OAIS framework covers all the components of preservation, while much of the existing literature was focused on Application (software) specifications.

Figure 1
Example Requirement



III. Degrees of Obligation for each Requirement

To ensure clarity and accuracy, this report adheres to the RFC 2119 standard for defining requirement levels.[4] It is important to understand the precise meanings of each of these keywords, particularly because each does not necessarily represent the most commonly accepted meaning of the word. In this document each recordkeeping requirement is qualified by one of five modal auxiliary verbs used to express different degrees of obligation. These different verbs are: MUST, MUST NOT, SHOULD, SHOULD NOT, and MAY. The use of each keyword is described below:

MUST. This word means that the definition is mandatory, or an absolute requirement of this specification. If this requirement is not fulfilled, the system can not be considered to be a trustworthy recordkeeping system.

MUST NOT. This phrase means that the definition is an absolute prohibition of the specification.

SHOULD. This word means that there may be valid reasons in particular circumstances to ignore a particular item, but the full implications must be understood and carefully weighed before choosing a different course. Such a requirement is highly desirable. If the requirement is not fulfilled, the level of trust in the recordkeeping system will be diminished.

SHOULD NOT. This phrase mean that there may exist valid reasons in particular circumstances when the particular behavior is acceptable or even useful, but the full implications should be understood and the case carefully weighed before implementing any behavior described with this label.

MAY. This word means that an item may be desirable, but truly optional. One vendor may choose to include the item because a particular marketplace requires it or because the vendor feels that it enhances the product while another vendor may omit the same item. An implementation that does not include a particular option MUST be prepared to interoperate with another implementation that does include the option, though perhaps with reduced functionality. In the same vein an implementation that does include a particular option MUST be prepared to interoperate with another implementation that does not include the option (except, of course, for the feature the option provides.)

[4] Scott Bradner, *RFC 2119: Key words for use in RFCs to Indicate Requirement Levels*, Network Working Group <http://www.ietf.org/rfc/rfc2119.txt?number=2119>.

IV. Recordkeeping System Requirements

This chapter of the report describes the features, behaviors, and qualities of a trustworthy recordkeeping system at a college or university. It describes these features, behaviors, and qualities as requirements in ten sections. The large majority of these requirements are synthesized from existing research into the requirements for recordkeeping systems conducted by a number of organizations and research projects over the last two decades. The requirements documents examined include:

- Indiana University, *Requirements for Electronic Records Management Systems (ERMS)*, Bloomington, IL: 2002 <http://www.indiana.edu/~libarch/ER/requirementsforrk.doc>.
  In this document referred to as: Indiana

- University of Pittsburgh, *Functional Requirements for Evidence in Recordkeeping*, Pittsburgh, PA: 1996
  <http://web.archive.org/web/20001024112939/www.sis.pitt.edu/~nhprc/prog1.html>.
  In this document referred to as: Pitt

- Alan Kowlowitz and Kristine L. Kelly, *Functional Requirements to Ensure the Creation, Maintenance, and Preservation of Electronic Records*, Albany, NY: Center for Technology in Government, State Univeristy of New York, 1998
  <http://www.ctg.albany.edu/publications/reports/functional/functional.pdf>.
  In this document referred to as: CTG

- IDA Programme of the European Commission, *Model Requirements for the Management of Electronic Records*, 2001 < http://ec.europa.eu/idabc/servlets/Doc?id=16847>.
  In this document referred to as: MoReq

- Public Records Office, *Functional Requirements for Electronic Records Management Systems*, Surrey, UK: 2002
  <http://www.nationalarchives.gov.uk/electronicrecords/reqs2002/pdf/requirementsfinal.pdf>.
  In this document referred to as: PRO

- InterPARES I Project, "Requirements for Assessing and Maintaining the Authenticity of Electronic Records," *in The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project*, San Miniato, Italy: Archilab, 2005
  <http://www.interpares.org/book/interpares_book_k_app02.pdf>
  In this document referred to as: InterPARES

- U.S. Department of Defense, *Design Criteria Standard for Electronic Records Management Software Applications* (DoD 5015.2-STD), Arlington, VA: 2002
  <http://www.dtic.mil/whs/directives/corres/pdf/50152std_061902/p50152s.pdf>
  In this document referred to as: DoD

- International Organization for Standardization, *ISO 15489-I: Information and*

documentation—Records management
In this document referred to as: ISO

- San Diego Super Computer Center at the University of California, San Diego, *Preserving the Electronic Records Stored in a Records Management Application* (PERM Project), San Diego, CA: 2002 <http://www.sdsc.edu/PERM/Final-Report-December-20-2002.pdf>
In this document referred to as: PERM

Health Insurance Portability and Accountability Act, 45 C.F.R. § 160, 162, 164 (2005) <http://ecfr.gpoaccess.gov/cgi/t/text/text-idx?c=ecfr&tpl=/ecfrbrowse/Title45/45cfrv1_02.tpl>.
In this document referred to as: HIPAA

This report articulates a set of requirements based on a synthesis of these reports appropriate for a college and university setting and within the framework of the following assumptions below.

1. **Retention and Disposition**
This section covers the act of executing the disposition of records according to a records retention schedule. This usually means the act of removing records and their metadata from the recordkeeping application for either destruction or for transfer to a preservation application. The work also includes reviewing records before carrying out their disposition and the application of legal holds on records that are involved in a legal action, audit, or review. This section does not cover the creation and assigning of records retention schedules. See Subsection 3.2.

1.1. Execution
This subsection covers the execution of a record's disposition, which usually means either destruction or transfer to a semi-active, inactive, or preservation application.

1.1.1. An *Institution* SHOULD dispose of records that no longer have operational value, either by permitting (arranging for) their destruction, or by transferring (arranging for) their transfer to a preservation repository.
[ISO, 4.3.9, MoReq 5]

1.1.2. *Procedures* MUST articulate the management of records disposition, in particular the destruction or transfer of records to a preservation system.
[MoReq 5.2.10, 5.3.1; ISO 9.9]

1.1.3. *Procedures* MUST allow for the confidential destruction of all copies and instances of records scheduled for destruction.
[MoReq 5.3.9; PRO A.4.74, B.3.26; DoD c2.2.10.6; ISO 9.9]

1.1.4. An *Application* MUST confidentially destroy records scheduled for destruction in a manner that does not allow their recovery.
[Pitt 10; MoReq 5.2.13, 5.3.14, 9.3.2; PRO A.4.67-68; DoD c2.2.6.63; ISO 9.9.a; HIPAA 45CFR164.310]

1.1.5. An *Application* SHOULD be able to retain metadata about records that it destroys.
[Pitt 10C; MoReq 5.2.15-16; DoD c2.2.6.6.4]

1.1.6. An *Application* MUST be able to transfer records scheduled for long-term retention to a preservation system in a trustworthy manner.
[MoReq 5.3.3, 5.3.5, 5.3.7; ISO 9.9.c]

1.1.7. An *Application* SHOULD be able to retain metadata about records that it transfers to a preservation system.
[DoD c2.2.6.5.4]

1.1.8. An *Application* MAY track the actual time of disposition for a record based on the retention schedule assigned to that record.
[PRO A.4.29, A.4.35-36, A.4.49]

1.1.9. An *Application* SHOULD be able to export records to a preservation system.
[PRO A.4.50, A.4.58; PERM non dod 1]

1.1.10. An *Application* MUST, if it can export records to a preservation system, export records in a manner that preserves their recordness.
[PRO A.4.50-52; InterPARES A.8; DoD c2.2.6.5.3; PERM non dod 1]

1.2. Compliance with Schedules

This subsection covers the need for the disposition of records to be executed in compliance with appropriate retention schedules.

1.2.1. An *Institution* MUST base the disposition of its records and audit trails on appropriate, authorized, and approved records retention schedules.
[Indiana 1.4.2, 1.8, 1.8.1-2; MoReq 3.4.6; PRO A.1.46; ISO 7.1, 9.9]

1.2.2. An *Application* SHOULD be able to manage a variety of retention period configurations and disposition instructions.
[DoD c2.2.2.2, c2.2.2.4, c2.2.2.4.1-3, c2.2.2.5]

1.2.3. An *Application* MUST be able to adjust the scheduled disposition of a record if the content of the retention schedule that governs the record changes.
[DoD c2.2.2.6, c2.2.2.7]

1.3. Review

This subsection covers the review of records before executing their disposition as prescribed by their assigned retention schedule.

1.3.1. *Procedures* MUST articulate steps for reviewing records before their scheduled disposition is executed.
[MoReq 5.1.10, 5.2; ISO 9.9]

1.3.2. An *Application* SHOULD alert people of and present to them for review records, including vital records, that have a pending disposition.
[Indiana 1.8.4; MoReq 5.1.10, 5.2.1, 5.2.3-4, 5.2.7-8, 9.3.7; PRO A.4.32, A.4.45-46, A.4.64]

1.4. Legal Holds

This subsection covers managing the process of suspending the execution of a record's disposition that is a part of any ongoing or reasonably expected legal action or proceedings, litigation, audit, investigation, or review.

1.4.1. An *Institution* MUST be aware of ongoing and reasonably expected legal action or proceedings, litigation, audit, investigation, or review that involves or may involve its records and identify any records so affected.
[Indiana 1.8.5; ISO 9.9]

1.4.2. *Procedures* MUST allow for the interruption of the scheduled disposition of records with legal holds that are or are expected to be involved in legal action or proceedings, litigation, audit, investigation, or review.
[Indiana 1.8.5; PRO A.4.25-26, A.4.38; DoD c2.2.6.4.1; ISO 9.9]

1.4.3. *Procedures* MUST allow for the appropriate lifting of legal holds on records and the resumption of their scheduled disposition.
[PRO A.4.27; DoD c2.2.6.4.3]

2. Records Capture and Registration

This section covers the creation and capture of records through recordkeeping systems in a manner that preserves the records integrity and essential elements of form or recordness. It covers the requirements to create records that document activities. It discusses the creation and capture of a variety of standard document types, complex documents, metadata, and relationships between records, along with the process of assigning unique identifiers and normalization during the creation and capture process.

2.1. Generate Records

This subsection covers the need to create required records to successfully conduct business activities

2.1.1. An *Institution* MUST document its activities by creating or capturing records when those activities commit the institution to action, render the institution accountable, or document an action, decision, or decision-making process.
[ISO 9.1]

2.1.2. An *Institution* SHOULD generate records that document all of its defined functions and activities.
[Indiana 1.2.1; ISO 7.1.a, 7.2.1, 8.2.5]

2.1.3. An *Institution* SHOULD ensure its recordkeeping applications are able to capture all of its records.
[MoReq 6.1.1; PRO A.2.1, A.2.4, A.2.6]

2.1.4. *Procedures* SHOULD include quality control mechanics to ensure that accurate records are created.
[Indiana 1.7; Pitt 7a]

2.1.5. *Juridical People* MUST have clearly defined responsibilities for creating and capturing records.
[ISO 6.3]

2.1.6. *Natural People* SHOULD only create records using documented recordkeeping applications and recordkeeping procedures.

[Pitt 3a]

2.1.7. *Natural People* MUST create and receive records as part of their daily work, and should do so in accordance with established policies, procedures, and standards.
[ISO 2.3.2]

2.1.8. An *Application* MUST enable the creation, reception, and keeping of records necessary to support business activities.
[ISO 2.3.1]

2.2. Create and Capture Integrity
This subsection covers the creation and capture of records in a recordkeeping system in a manner that preserves their integrity.

2.2.1. An *Application* MUST create and capture records in a manner that maintains the integrity and identity of the records.
[Pitt 7a1; InterPARES B.1]

2.2.2. An *Application* SHOULD verify the integrity of the records it creates and captures.
[MoReq 6.2.1]

2.2.3. *Procedures* MUST articulate steps that maintain an unbroken custody of records during capture.
[InterPARES B.1.a]

2.3. Create and Capture Recordness
This subsection covers the creation and capture of the essential aspects of records in a recordkeeping system.

2.3.1. An *Application* MUST be able to create and capture a record's context, structure, and content that together documents the institution's decisions, actions, or communications.
[Pitt 7b, 7b1-4; MoReq 6.1.2; PRO A.2.8]

2.3.2. *Procedures* MUST provide for the creation and capture of records in a manner that allows them to correctly reflect the decisions, actions, or communications it documents.
[Pitt 7c, 7c1-3; InterPARES A.1.a.i-v, A.1.b.i-iv, A.5]

2.4. Support of Format Types
This subsection covers the creation and capture of records of various formats.

2.4.1. An *Institution* MUST have recordkeeping applications that together are able to create and capture all of the record formats the institution generates in the course of its business.

[MoReq 6.1.1]

2.4.2. An *Application* SHOULD be able to create and capture records with a variety of format types and structures.
[Indiana 1.2.10; MoReq 6.1, 6.3, 6.3.1-2]

2.5. Create and Capture Complex Documents
This subsection covers the creation and capture of complex records.

2.5.1. An *Application* MUST, if it is used to manage complex records, be able to create and capture records in a manner that captures the structural integrity of its component parts.
[MoReq 6.1.13, 6.3.1, 6.3.2; PRO A.2.5, A.2.8; ISO 7.2.1.a]

2.5.2. An *Application* MAY allow for the creation and capture of complex records, a single compound record, or as a series of linked simple records.
[Indiana 1.2.7; MoReq 6.3.6]

2.6. Create and Capture Relationships Between Records
This subsection covers the capture of the relationships between records.

2.6.1. An *Application* MUST be able to capture the relationships between records.
[PRO A.8.17]

2.7. Create and Capture Information About Records
This section covers the creation and capture of metadata associated with records a recordkeeping system creates and captures.

2.7.1. An *Application* SHOULD be capable of automatically extracting metadata from the records it creates and captures.
[Indiana 1.6.1; MoReq 6.1.6, 6.1.14]

2.7.2. An *Application* MUST allow people to manually enter metadata that cannot be automatically extracted from the records created and captured by the recordkeeping application.
[Indiana 1.6.3; MoReq 6.1.9; PRO A.2.38]

2.7.3. *Procedures* MUST provide for the creation of necessary metadata during the creation and capture process that did not exist before creation or capture.
[MoReq 6.1.9; PRO A.2.38]

2.7.4. An *Application* SHOULD be able to technically validate the metadata it creates or captures.
[Indiana 1.6.4; MoReq 6.1.1]

2.7.5. *Procedures* SHOULD provide for the intellectual validation of the metadata (data content standard of the metadata is met) the recordkeeping system creates or captures during the creation or capture process.
[Indiana 1.6.4; MoReq 6.1.1]

2.7.6. An *Application* SHOULD be able to create and capture descriptive, contextual, and technical metadata.
[PERM 12]

2.7.7. *Procedures* SHOULD provide for the creation and capture of descriptive, contextual, and technical metadata.
[Indiana 1.2.3; Pitt 8a; MoReq 6.1.2, 6.1.3; ISO 7.2.1.b]

2.7.8. An *Application* MUST create and capture records and their metadata in a manner that allows them to be persistently linked.
[Indiana 1.2.3; MoReq 6.1.3; ISO 7.1.c]

2.7.9. An *Application* MUST assign unique identifiers to the records it creates and captures.
[Indiana 1.2.5; MoReq 7.1.5]

2.8. System Interaction

This subsection covers the ability of a recordkeeping application to communicate and integrate with other recordkeeping and various record creating applications.

2.8.1. An *Application* SHOULD be capable of communication with all of the Institution's other recordkeeping and record creating applications.
[Indiana 1.6.2; MoReq 6.2.1; PRO A.2.2]

2.8.2. An *Application* SHOULD provide an application programming interface to enable integration with other business applications.
[PRO A.2.3]

2.9. Normalization

This subsection covers capture of standard format versions of records in a recordkeeping system captured in other formats. This section does not cover migration, which is covered in Section 7, Preservation. This deals specifically with normalization during the capture process.

2.9.1. An *Application* SHOULD be able to capture a standard format version of records it captures in its native format.
[PRO A.2.12]

2.9.2. An *Application* MUST persistently link the format versions of the same records together.

[PRO A.2.12]

3. **Classification**

This section covers the development and management of classification schemes, which include records retention schedules, in recordkeeping systems. It also covers the assigning of records to classes within a classification scheme or multiple schemes and the institutional context of these schemes. Although assigning a record to a scheme assigns meaning and prescribes actions to that record, the execution of those actions is not covered in this section.

3.1. Manage Scheme

This subsection covers the creation, management, and modification of classification scheme(s) within a recordkeeping system. A classification scheme is a logical system used to arrange records. Usually, classes are related component parts that compose a scheme. This section does not cover the act of classifying records.

3.1.1. An *Application* MUST allow the creation and defining of a classification scheme.
[MoReq 3.1.5; PRO A.1.3, A.4.1; ISO 9.3.A; DoD c2.2.1.1]

3.1.2. An *Application* MAY allow the creation and defining of multiple classification schemes.
[MoReq 3.1.8; PRO A.1.10]

3.1.3. An *Application* MAY allow the creation and defining of a vital records classification scheme.
[DoD c2.2.6.7]

3.1.4. An *Application* MUST allow the changing, amending, deleting and adding to a classification scheme.
[Indiana 1.8.7; MoReq 3.1.6, 3.4.1; PRO A.1.4, A.1.6, A.1.8, A.4.4, A.4.6]

3.1.5. An *Application* MUST ensure that classification names are unique.
[PRO A.1.18]

3.1.6. An *Application* SHOULD allow the closing of classes within a scheme so that no new records can be added to a closed class.
[PRO A.1.7, A.1.41]

3.1.7. An *Application* MUST NOT allow the deletion of classes that contain records.
[PRO A.1.9]

3.1.8. An *Application* SHOULD NOT impose any practical limits on the number of classes or class levels that exits within a scheme.
[MoReq 3.1.3, 3.2.9; PRO A.1.28]

3.1.9. An *Application* SHOULD report its classes, schemes, and records in a logical, usable fashion.

[MoReq 3.2.10; ISO 9.3.6]

3.2. Retention Schedules

This subsection covers the management and modification of retention schedules along with the act of assigning record(s) to a retention schedule(s). Retention schedules prescribe a record's required length of retention and its disposition. Retention schedules are a type of classification scheme. This subsection does not cover the execution of a record's disposition, see subsection 1.1.

3.2.1. An *Application* MUST be able to assign a retention schedule to a record.
[Indiana 1.8.3; MoReq 5.1.4; PRO A.4.14; ISO 8.1.f]

3.2.2. An *Application* MUST be able to reassign a retention schedule to a record.
[PRO A.4.21]

3.2.3. An *Institution* MUST associate retention schedules with dispositions and retention periods and the reasons and sources for these determinations.
[Pitt 1b; MoReq 5.1.3, 5.1.11, 5.17, 5.10; PRO A.4.7, A.4.9, A.4.10, A.4.12; ISO 8.1.f, 9.2.c.1-3]

3.2.4. An *Institution* MUST be able to change the dispositions and retention periods of the retention schedules.
[Indiana 1.8.7; MoReq 5.1.15-16; PRO A.4.6, A.4.1]

3.2.5. An *Institution* MUST assign retention schedules to all of its records.
[Indiana 1.8.6]

3.3. Naming

This subsection covers the naming of a classification scheme and its classes within a recordkeeping system.

3.3.1. An *Application* SHOULD support a naming scheme for classification taxonomies.
[MoReq 3.1.4]

3.3.2. An *Application* MAY support user-defined naming schemes for classification taxonomies.
[MoReq 3.1.4]

3.3.3. An *Application* MAY support the use of controlled vocabulary terms to support the creation of naming schemes.
[MoReq 3.2.6, 3.2.8; PRO A.1.24; ISO 9.5.3]

3.3.4. An *Application* MAY use one of two strategies for creating naming schemes: a structured alpha/numeric system or a human understandable textual system.
[MoReq 3.2.2; PRO A.1.14-15]

3.3.5. An *Application* MAY support the mandatory use of a naming scheme.
[PRO A.1.20, A.1.36]

3.4. Assign Classification
This subsection covers the assigning of a record(s) to a class(es) within a classification scheme in a recordkeeping system. Although assigning a record to a scheme assigns meaning and prescribes actions to that record, the execution of those actions is not covered in this subsection. See subsection 1.1.

3.4.1. An *Institution* MUST classify records, assigning them to a pre-established class in a classification scheme.
[ISO 4.2.1-4.2.2]

3.4.2. An *Application* MUST assign all of the records it maintains to a class or multiple classes of a classification scheme.
[Indiana 1.8.3, 1.2.4; MoReq 6.1.1; PRO A.2.19, A.2.21, A.4.55]

3.4.3. An *Application* MUST be able to assign a classification to a particular record that overrides the classification of the group of records that individual record is assigned to.
[MoReq 5.1.14]

3.4.4. An *Application* MUST be able to reassign a record to a different class.
[MoReq 3.4.2, 5.1.16; PRO A.1.47, A.2.50, A.4.21]

3.4.5. An *Application* MAY support the use of controlled vocabulary terms to support the classification of records.
[PRO A.1.37]

3.4.6. An *Application* MAY support records being classified as vital records.
[MoReq 4.3.6]

3.5. Institution and Context
This subsection covers the institutional context into which a classification scheme within a recordkeeping system should fit.

3.5.1. An *Institution* SHOULD ensure that its recordkeeping applications are compatible with the institution's classification scheme(s).
[Indiana 1.3.1; MoReq 3.1.1]

3.5.2. An *Institution* SHOULD ensure its classification scheme(s) reflect its business processes.
[ISO 8.2.2.b, 9.5.2]

4. **Storage and Handling**
This section covers the institution's identification and management of records in

recordkeeping systems which includes location tracking, versioning management, and unique identifier management. This section also discusses the integration of the recordkeeping systems into the business process and workflow of the institution. This section also covers the tracking of a record during its maintenance in a recordkeeping system. This section covers the management of versions of records while they are maintained in a recordkeeping system. This section covers the unique identification of a record, the maintenance of its logical relationships and the identification of its custodian(s) during its maintenance in a recordkeeping system. This section covers the preservation of the context, content, structure, and functionality of records in a recordkeeping system.

4.1. Maintain Integrity

This subsection covers the creation and capture of records in a recordkeeping system in a manner that maintains their integrity.

4.1.1. An *Application* MUST enforce data integrity at all times.
[MoReq 3.4.12; PRO A.9.2; ISO 8.3.6]

4.1.2. An *Application* MUST be able to maintain a record's fixity.
[PRO A.2.14, A.2.18; InterPARES B.1.C; DoD c2.2.3.8]

4.2. Maintain Recordness

This subsection covers the preservation of a record's recordness during its maintenance in a recordkeeping system.

4.2.1. An *Institution* MUST maintain records in a manner that allows them to correctly reflect the decision, action, or communication it documents.
[ISO 7.2.1]

4.2.2. An *Application* MUST maintain a record's content, structure, and context that document the institution's decisions, actions, and communications.
[Pitt 7]

4.2.3. *Procedures* MUST maintain the context, structure, and content of records throughout all recordkeeping activities.
[Pitt 9; PERM non dod5, 2]

4.2.4. *Procedures* MUST maintain the chain of custody of records throughout all recordkeeping activities.
[InterPARES B.1]

4.2.5. *Procedures* MUST maintain the logical boundaries and the relationships between records throughout all recordkeeping activities.
[Pitt 9b1, 9b2]

4.3. Location Tracking

This subsection covers the tracking of a record during its maintenance in a

recordkeeping system.

4.3.1. An *Application* MUST be able to track the location of records in a recordkeeping system.
[MoReq 4.4.1]

4.3.2. An *Application* MUST track a record's unique identifier, current location, time of movements, the person responsible for the movements, and the custodian of the record.
[MoReq 4.4.3; ISO 9.8.3]

4.3.3. *Procedures* MUST articulate steps that govern the receipt, removal, and movement of hardware and media that store electronic records.
[HIPAA 45CFR164.310]

4.4. Versioning

This subsection covers the management of versions of records while they are maintained in a recordkeeping system.

4.4.1. An *Application* SHOULD support versioning of the records it manages.
[Indiana 1.2.9]

4.4.2. An *Application* MUST, if it supports versioning, manage the relationship between the versions of the same record in a recordkeeping system.
[Indiana 1.2.8; DoD c2.2.3.18, c2.2.3.20]

4.4.3. An *Application* SHOULD, if it supports versioning, be able to identify the authoritative version of a record in a recordkeeping system that has multiple versions.
[InterPARES A.7]

4.4.4. An *Application* MUST, if it supports versioning, document the version changes of a record since its creation.
[InterPARES B.3]

4.5. Additional Records Attributes

This subsection covers the ability of a recordkeeping application to interoperate with other record creating and keeping applications while it maintains records.

4.5.1. An *Application* MUST uniquely identify the records it maintains.
[Pitt 6c; MoReq 7.1; PRO A.9.3; DoD c2.2.1.4, c2.2.4.1; PERM 15]

4.5.2. An *Application* MUST maintain the logical relationships between records in a recordkeeping system.
[MoReq 3.4.11; PRO A.2.24; DoD c2.2.3.17]

4.5.3. An *Application* MUST maintain the logical relationships between multiple versions of the same record.
[DoD c2.2.3.19]

4.5.4. An *Application* SHOULD identify the responsible custodian(s) of the records it maintains.
[PRO A.5.41]

4.6. Respond to Data Failure

This subsection covers the planning for and response to disasters that have an impact on the creation, capture, management, and use of records in a recordkeeping system.

4.6.1. *Institution* SHOULD create backup and failure mode procedures for its records and vital records.
[Indiana 1.9, 1.9.4; Pitt 2d; MoReq 4.3.7; InterPARES A.3; ISO 8.3.3]

4.6.2. *Procedures* SHOULD provide for the automated backup of the institution's records, metadata, audit trails, and configuration settings.
[MoReq 4.3, 4.3.1, 9.1.2-3; PRO A.9.11, A.9.17; DoD c2.2.9.1]

4.6.3. An *Application* MUST NOT hinder automated backup of the institution's records.
[DoD c2.2.9.1, MoReq 4.3.1]

4.6.4. *Procedures* SHOULD articulate the actions needed to be undertaken during primary system failure.
[Pitt 2d; MoReq 4.3.5; HIPAA 45CFR164.308]

4.6.5. *Infrastructure* SHOULD allow for backups to be stored at geographically distant locations.
[PRO A.9.12; DoD c2.2.9.2]

4.6.6. An *Application* SHOULD provide facilities for restoring data from backup data and returning the data stores to a consistent state.
[Pitt 4d; MoReq 11.3.5, 4.3.3, 4.3.4; PRO A.9.14-16; DoD c2.2.9.3, c2.2.2.9.3.1-2, c2.2.9.4-5; HIPAA 45CFR164.308]

4.6.7. *Institutions* SHOULD test and review backup and failure mode procedures.
[HIPAA 45CFR164.308, 45CFR164.310]

4.7. Intrusion Detection and Response

This subsection covers the detection and response to unauthorized access to and tampering of records in a recordkeeping system.

4.7.1. An *Institution* SHOULD create and maintain policies and procedures to detect, contain, and correct security violations.

[HIPAA 45CFR164.308, 45CFR164.310, 45CFR164.312]

4.7.2. *Procedures* MUST provide a reasonable guarantee that records are protected from tampering.
[Pitt 9a; PRO A.2.15; ISO 7.1.1, 8.2.2.c; HIPAA 45CFR164.306]

4.7.3. *Procedures* MUST prescribe periodic software security updates.
[HIPAA 45CFR164.308]

4.7.4. An *Institution* SHOULD perform a periodic review of its security procedures.
[InterPARES B.1.b; HIPAA 45CFR164.308]

4.7.5. An *Application* SHOULD be able to facilitate reconstruction, review, and examination of the events surrounding or leading to mishandling of records, or possible compromise of sensitive information.
[DoD c2.2.8.3.2]

4.7.6. An *Institution* SHOULD create and maintain policies and procedures to perform regular reviews of audit logs and log-in attempts.
[HIPAA 45CFR164.308]

5. Access

This section covers the institution's management of users' rights to view and/or receive records, including the development, management, and review of records and user security profiles and the management of access controls and authentication of users. This section also covers the recordkeeping system enabling users to search and discover records along with the system disseminating meaningful and functional records to users, including the management of searching mechanisms and query techniques. In addition it covers services to allow browsing and the proper rendering of complex records, a record's recordness, and redacted records.

5.1. Define Access Controls

This subsection covers the definition of access controls, or the assigning responsibility for the creation, modification, annotation, relocation, and destruction of records on the basis of a person's authority and capacity to carry out an administrative activity.

5.1.1. An *Institution* MUST explicitly assign responsibility for the creation, modification, annotation, relocation, and destruction of records on the basis of a person's authority and capacity to carry out an administrative activity.
[Indiana 1.7.2, InterPARES A.2]

5.1.2. An *Application* MUST confer exclusive capabilities upon people to exercise the responsibility for creation, modification, annotation, relocation, and destruction of records as defined by an institution.
[Indiana 1.4.1; DoD c2.2.5.2, C2.2.7.4; ISO 8.3.6; HIPAA 45CFR164.308;

InterPARES A.2]

5.1.3. An *Application* SHOULD manage the security level of the records it maintains.
[MoReq 9.3.3, 9.3.5]

5.1.4. An *Application* MUST NOT allow unauthorized changes to the records it maintains.
[Indiana 1.7.1; MoReq 3.2.1, 4.5.4, 6.1.4; PRO A.2.41; DoD c2.2.5.4; ISO 9.7.d]

5.1.5. An *Application* MUST NOT allow unauthorized creation of records.
[Pitt 8]

5.1.6. An *Application* MAY tailor its user interface to the user's appropriate access level.
[PRO A.8.9]

5.1.7. *Infrastructure* MUST NOT allow unauthorized access to the workstations and hardware that contain or provide access to records.
[HIPAA 45CFR164.310]

5.1.8. An *Institution* SHOULD demonstrate it has created and maintains a reasonable access criteria and it has successfully implemented the criteria.
[InterPARES A.2; ISO 8.3.6]

5.2. Management of Access Controls

This section covers the institution's management of users' rights to view and/or receive records. This includes the development, management, and review of records and user security profiles. It also includes the management of access controls and authentication of users.

5.2.1. An *Institution* MUST develop and implement access control rules for its records.
[MoReq 4.6.5; ISO 9.7; PERM 25; HIPAA 45CFR 164.308, 45CFR 164.312]

5.2.2. *Procedures* MUST insure that only authorized users gain access to records.
[MoReq 4.1.1; PRO A.5.25, A.5.42, A.5.46-50]

5.2.3. An *Institution* MAY designate people as custodians of records and the custodians are responsible for implementing the access control rules governing their records.
[PRO A.5.41, A.5.43-44; ISO 9.7.e]

5.2.4. An *Application* MUST limit search results to the records the user has rights to access.
[MoReq 4.1.10, 4.1.12, 8.1.28; PRO A.3.18, A.5.51-52, B.3.18]

5.3. Security Profiles and Authentication

This subsection covers the creation, management, assigning, reviewing, and modifying

of security profiles for records in a recordkeeping system.

5.3.1. An *Institution* MUST create and modify records security profiles. [ISO 4.3.5]

5.3.2. An *Application* MUST allow records security profiles to be created and modified. [MoReq 9.3.5; PRO A.5.36]

5.3.3. An *Application* MUST allow record security profiles to be assigned to records. [MoReq 4.6.1; PRO A.2.26, A.5.5, A.5.26, A.5.27; ISO 9.7.2]

5.3.4. An *Application* SHOULD allow time sensitive records profiles that are valid for a limited time period to be assigned to records and should automatically be switched to another records security profile when their valid time period expires. [PRO A.5.38-39]

5.3.5. An *Application* MUST allow user security profiles to be created and modified. [MoReq 9.1.8; PRO A.5.11, A.5.17-18, A.5.20-22]

5.3.6. An *Application* MUST assign or reassign user security profiles to people. [MoReq 4.1.2, 4.1.5, 4.6.7, 9.1.7; PRO A.5.5, A.5.10, A.5.13, A.5.16, A.5.24; DoD c2.2.7.3]

5.3.7. *Infrastructure* SHOULD provide services for secure authentication. [PRO A.5, A.5.1, A.5.11; DoD c2.2.7.1; HIPAA 45CFR164.312]

5.3.8. An *Application* MUST authenticate users before providing services. [PRO A.5, A.5.1, A.5.11; DoD c2.2.7.1; HIPAA 45CFR164.312]

5.3.9. *Procedures* SHOULD allow for the periodic review of access control rules, records security profiles, and user security profiles. [MoReq 4.6.12; PRO A.5.40; ISO 9.7; HIPAA 45CFR164.308]

5.3.10. *Procedures* SHOULD allow for the modification of access control rules, records security profiles, and user security profiles based on the findings of a review. [HIPAA 45CFR164.308]

5.4. Searching

This subsection covers the capabilities of a recordkeeping system to search the records it maintains.

5.4.1. An *Application* MUST ensure all of its records and metadata are discoverable. [Indiana 1.10.2-3; MoReq 8.1.4-5, 8.1.7; PRO A.3.4, A.3.6, A.3.8, A.3.17; PERM 18]

5.4.2. An *Application* SHOULD provide an integrated search interface.
[MoReq 8.1.2; PRO A.3.7]

5.4.3. An *Application* SHOULD support external search engines in addition to any integrated search interface.
[PRO A.3.19]

5.4.4. An *Application* MUST, if it has an integrated search interface, present search results.
[PRO A.3.15; DoD c2.2.6.8.5]

5.4.5. An *Application* MUST be able to render all records returned in a search results list.
[MoReq 8.2.1; PRO A.3.20; DoD c2.2.6.8.10]

5.4.6. An *Application* SHOULD provide capabilities to manage a search results list including, but not limited to, order, number of hits per page, filter results files, and save search results.
[MoReq 8.1.17, 8.1.24-25; DoD c2.2.6.8.5]

5.4.7. An *Application* MUST support searching by records' identifiers.
[MoReq 8.1.16, 8.1.23]

5.4.8. An *Application* SHOULD be able to save and reuse queries.
[MoReq 8.1.20; PRO A.3.11-12]

5.5. Query Techniques

This subsection covers the querying techniques a recordkeeping system employs to search the records it maintains.

5.5.1. An *Application* SHOULD support the full text search of the records and metadata it maintains.
[MoReq 8.1.8; DoD c3.2.9]

5.5.2. An *Application* SHOULD support searching metadata fields containing controlled vocabulary terms managed by thesauri.
[MoReq 8.1.10; PRO A.3.5; DoD c3.2.9]

5.5.3. An *Application* SHOULD support searching multiple metadata fields and/or full text of records.
[MoReq 8.1.6; PRO A.3.9; DoD c2.2.6.8.2]

5.5.4. An *Application* SHOULD support the use of Boolean and/or relational search operators such as "and" "or" "not" "less than" "greater than" "equal to."
[MoReq 9.1.8; PRO A.3.13; DoD c2.2.6.8.4]

5.5.5. An *Application* SHOULD support wild card and/or pattern matching searches.
[MoReq 8.1.11; PRO A.3.13; DoD c2.2.6.8.3]

5.5.6. An *Application* SHOULD support the iterative refinement of a search by adding search conditions to a previously run search—i.e. narrow a search.
[MoReq 8.1.21]

5.5.7. An *Application* MAY support word proximity searching.
[MoReq 8.1.12]

5.5.8. An *Application* MAY support searching null values.
[DoD c2.2.6.8.6]

5.5.9. An *Application* MAY support searching time intervals.
[MoReq 8.1.22]

5.6. Rendering Complex Objects

This subsection covers the ability of a recordkeeping system to deliver a record it maintains to a user in a manner that fully maintains the record's context, structure, and content.

5.6.1. An *Application* MUST render all of the components of a record and its metadata in a logical manner.
[Indiana 1.10.4; MoReq 8.2.3; PRO A.3.21]

5.6.2. An *Application* MUST be able to render records together with their associated metadata.
[MoReq 8.1.15; PRO A.3.24; DoD c2.2.3.21; PERM 23]

5.6.3. An *Application* MUST be able to render records on to appropriate output mediums which should at least include graphical display and printer output.
[MoReq 8.2, 8.3, 8.4.1; Pro A.3.25-26, A.3.28-29; PERM 3, 10, 14, 16, 17, 24, non DoD 2]

5.6.4. An *Application* SHOULD be able to render records into an open export format.
[PRO A.3.31]

5.6.5. An *Application* SHOULD be able to render records independently of their creating environments.
[MoReq 8.2.2; PRO A.3.22; DoD c3.2.14]

5.6.6. An *Application* SHOULD be able to render a record simultaneously for multiple users.
[PRO A.3.23; DoD c2.2.7.5]

5.6.7. An *Application* SHOULD be able to render all versions of a record.
[DoD c2.2.6.8.9]

5.7. Rendering Recordness

This subsection covers the ability of a recordkeeping system to deliver a record it maintains to a user in a manner that fully maintains the record's context, structure, and content.

5.7.1. An *Application* MUST render a record's content.
[Pitt 11, 12; MoReq 8.2.3; PRO A.3.21; PERM 2]

5.7.2. An *Application* MUST render a record's structure.
[Pitt 12, 12b, 12b1; PRO A.3.21; PERM 2]

5.7.3. An *Application* MUST render a record's context.
[Pitt 12, 12b1, 12c; ISO 7.25; PERM 2]

5.7.4. An *Application* MUST render a record's functionality.
[Pitt 11b; DoD c2.2.5.3]

5.8. Availability

This subsection covers the availability of needed records in a recordkeeping system.

5.8.1. An *Application* MUST ensure that records needed for their primary business functions are available.
[Indiana 1.10, 1.10.1; Pitt 12a; ISO 8.3.6]

5.8.2. An *Application* SHOULD ensure that records needed for secondary use are available.
[Indiana 1.10, 1.10.1; Pitt 12a]

5.8.3. An *Application* MUST ensure that its records are available in a timely manner.
[Indiana 1.10.1; Pitt 12a; ISO 8.3.6]

5.9. Browsing

This subsection covers the ability of a recordkeeping application to provide users the capability to browse records.

5.9.1. An *Application* SHOULD support the browsing of its classification schemes, including any hierarchical structure in which the records are managed.
[MoReq 8.1.13, 8.1.27, 3.1.7; PRO A.3.3; DoD c2.2.1.6]

5.10. Redaction

This section covers the management and execution of redacting records and the delivery of redacted versions of records to users.

5.10.1. *Procedures* SHOULD provide for the redaction of restricted content from records delivered to users that do not have the right to see the restricted output.
[Pitt 13; MoReq 9.3.10; PRO A.2.56]

5.10.2. An *Application* SHOULD be able to create redacted versions of textual, audio, and moving image records.
[MoReq 9.3.10]

5.10.3. An *Application* MUST NOT, if it can redact records, alter the content of a record while creating a redacted version of that record.
[Pitt 13a; PRO A.2.56]

6. **Design and Performance**

This section covers the software and hardware design and performance of the recordkeeping application, including system maintenance, scalability, design constraints, and testing and verification. This section also covers the application's usability.

6.1. Testing and Verification

This subsection covers the testing and verification of a recordkeeping application's and the infrastructure's performance.

6.1.1. An *Institution* SHOULD determine an appropriate suite of tests against which the recordkeeping infrastructure and recordkeeping application will be measured and set acceptable ranges for system performance.
[Indiana 1.12; MoReq 11.2, 11.2.5]

6.1.2. *Procedures* SHOULD include provisions for regular execution of application and infrastructure tests.
[Indiana 1.12; PRO A.9.22]

6.1.3. *Infrastructure* SHOULD reliably pass all tests and perform within stated acceptable ranges.
[Indiana 1.12; Moreq 11.2]

6.1.4. An *Application* SHOULD reliably pass all tests and perform within stated acceptable ranges.
[Indiana 1.13; PRO A.9.22; MoReq 11.2, 11.2.1-4]

6.1.5. An *Application* SHOULD undergo formal verification and be provably correct.
[Pitt 4b, 4c]

6.2. System Maintenance

This subsection covers the maintenance of the recordkeeping application and infrastructure.

6.2.1. *Procedures* SHOULD contain provisions for all routine maintenance tasks which fall in line with industry best practices.
[Pitt 2c; CTG System]

6.2.2. An *Application* MUST allow convenient access to and the ability to modify any configuration parameters.
[MoReq 11.2.7, 9.1.1]

6.2.3. *Infrastructure* SHOULD provide the ability to monitor available storage capacity.
[MoReq 9.14; PRO A.9.21]

6.2.4. An *Institution* SHOULD determine the acceptable ranges for downtime and minimum numbers of simultaneous users.
[MoReq 11.3; DoD c3.1.3]

6.2.5. *Infrastructure* SHOULD be capable of fulfilling downtime and simultaneous user requirements laid out by the institution.
[MoReq 11.3]

6.3. User Interface

This subsection covers the user interfaces of a recordkeeping application.

6.3.1. An *Application* SHOULD provide a user interface which is easy to use.
[MoReq 11.1; PRO A.8.11; DoD c2.2.5.1]

6.3.2. An *Application* SHOULD follow generally accepted user interface guidelines by providing a consistent look and feel.
[PRO 8.1-3]

6.3.3. An *Application* MAY provide a remote login facility.
[MoReq A.9.7]

6.3.4. An *Application* SHOULD facilitate use by persons with disabilities by including accessibility features.
[PRO A.8.16]

6.3.5. An *Application* SHOULD provide meaningful error messages in the event of an error, and attempt to guide the user to an appropriate resolution.
[PRO A.8.7-8]

6.4. Scalability

This subsection covers scalability issues concerning the recordkeeping system.

6.4.1. An *Application* SHOULD be able to both scale up to large organizations, and scale down for smaller organizations.
[MoReq 11.2.6, 11.2.8]

6.4.2. *Institutions* SHOULD estimate its medium and long-term scalability requirements and determine acceptable ranges for various scalability metrics.
[PRO A.9.23]

6.4.3. An *Application* SHOULD be capable of fulfilling its institution's scalability requirements, and of operating within acceptable ranges.
[PRO A.9.23]

6.4.4. An *Application* SHOULD NOT impose any practical limit on the number of records which can be managed by the application.
[MoReq 6.3.5; PRO A.2.20]

6.4.5. An *Application* SHOULD provide the ability to synchronize multiple instances of all underlying data stores.
[DoD c2.2.3.24]

6.4.6. An *Application* SHOULD, when it offers remote or distributed services, use efficient network protocols which minimize the amount of data exchange required.
[PRO A.9.20]

6.5. Sustainability

This subsection covers the design constraints that affect sustainability of the recordkeeping application.

6.5.1. An *Application* SHOULD be designed around a flexible architecture which can evolve as the institution's needs change.
[PRO A.9.1]

6.5.2. An *Application* MAY support a distributed repository with multi-site service.
[PRO A.9.18]

6.5.3. An *Application* SHOULD provide at least one version of backward compatibility.
[DoD c2.1.4]

V. Requirements for the Preservation of University Electronic Records

This chapter describes the features, behaviors, and qualities that are necessary in order to preserve and ensure the continued accessibility and authenticity over time of electronic records at a college or university, written as requirements. Preservation is the whole of the activities and processes involved in the technical stabilization and physical and intellectual protection of resources through time. Those sources are:

- Indiana University, *Requirements for Electronic Records Management Systems (ERMS)*, Bloomington, IL: 2002 <http://www.indiana.edu/~libarch/ER/requirementsforrk.doc>.
  In this document referred to as: Indiana

- University of Pittsburgh, *Functional Requirements for Evidence in Recordkeeping*, Pittsburgh, PA: 1996
  <http://web.archive.org/web/20001024112939/www.sis.pitt.edu/~nhprc/prog1.html>.
  In this document referred to as: Pitt

- Alan Kowlowitz and Kristine L. Kelly, *Functional Requirements to Ensure the Creation, Maintenance, and Preservation of Electronic Records*, Albany, NY: Center for Technology in Government, State Univeristy of New York, 1998
  <http://www.ctg.albany.edu/publications/reports/functional/functional.pdf>.
  In this document referred to as: CTG

- IDA Programme of the European Commission, *Model Requirements for the Management of Electronic Records*, 2001 < http://ec.europa.eu/idabc/servlets/Doc?id=16847>.
  In this document referred to as: MoReq

- Public Records Office, *Functional Requirements for Electronic Records Management Systems*, Surrey, UK: 2002
  <http://www.nationalarchives.gov.uk/electronicrecords/reqs2002/pdf/requirementsfinal.pdf>.
  In this document referred to as: PRO

- InterPARES I Project, "Requirements for Assessing and Maintaining the Authenticity of Electronic Records," *in The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project*, San Miniato, Italy: Archilab, 2005
  <http://www.interpares.org/book/interpares_book_k_app02.pdf>
  In this document referred to as: InterPARES

- U.S. Department of Defense, *Design Criteria Standard for Electronic Records Management Software Applications* (DoD 5015.2-STD), Arlington, VA: 2002
  <http://www.dtic.mil/whs/directives/corres/pdf/50152std_061902/p50152s.pdf>
  In this document referred to as: DoD

- International Organization for Standardization, *ISO 15489-I: Information and*

documentation—Records management
In this document referred to as: ISO

- San Diego Super Computer Center at the University of California, San Diego, *Preserving the Electronic Records Stored in a Records Management Application* (PERM Project), San Diego, CA: 2002 <http://www.sdsc.edu/PERM/Final-Report-December-20-2002.pdf>
In this document referred to as: PERM

Health Insurance Portability and Accountability Act, 45 C.F.R. § 160, 162, 164 (2005) <http://ecfr.gpoaccess.gov/cgi/t/text/text-idx?c=ecfr&tpl=/ecfrbrowse/Title45/45cfrv1_02.tpl>.
In this document referred to as: HIPAA

A small number of additional requirements have been identified as a result of this research project and are attributed to The Fedora and the Preservation of University Electronic Records Project.
In this document referred to as: Tufts-Yale

In addition, the project team surveyed the preservation system requirements literature and selected what it believes to be the seven most applicable to the university archives community. This literature is much less standardized than the recordkeeping system requirements literature, consisting mostly of more general research best practices statements. In the last two years, the situation has improved because of the landmark work completed by the National Archives and Records Administration's Electronic Records Archives Program Management Office. This grand and detailed statement of responsibilities must be considered the most significant work of its kind. The full list of the preservation system requirements documents includes:

- National Archives and Records Administration Electronic Records Archives Program Management Office, *Electronic Records Archives Requirements Document* (RD), Version 2.0, prepared by Integrated Computer Engineering (ICE), College Park, MD: 2003 <http://www.archives.gov/era/about/requirements.csv>.
In this document referred to as: NARA

- *ISO 14721:2003: Space data and information transfer systems—Open archival information system—Reference model* (Geneva: International Organization for Standardization, 2003).
In this document referred to as: ISO

- Research Libraries Group, *Trusted Digital Repositories: Attributes and Responsibilities*, Mountain View, CA: RLG, 2002 <http://www.rlg.org/legacy/longterm/repositories.pdf#search=%22Trusted%20Digital%20Repositories%3A%20Attributes%20and%20Responsibilities%22>
In this document referred to as: TDR

- Research Library Group and National Archives and Records Administration, *An Audit*

Checklist for the Certification of Trusted Digital Repositories, Draft for Public Comment, Mountain View, CA: RLG, 2005 <http://www.rlg.org/en/pdfs/rlgnara-repositorieschecklist.pdf>.
In this document referred to as: CTDR

- van Diessen, Dr. R.J., *Preservation Requirements in a Deposit System: IBM/KB Long Term Preservation Study*, The Hague: IBM and Koninklijke Bibliotheek, 2002 <http://www.kb.nl/hrd/dd/dd_onderzoek/reports/3-preservation.pdf>.
In this document referred to as: KB

- Yale University Library, *Requirements Document for the Rescue Repository*, New Haven, CT: Yale, 2004 <http://www.library.yale.edu/iac/documents/RescueRepositoryRequirements.pdf#search=%22Requirements%20Document%20for%20the%20Rescue%20Repository%22>.
In this document referred to as: Yale

1. Common Services

This section covers information technology services that are necessary to support the preservation system, but not necessarily unique to that system, including operating systems, networking, and security. It is most likely that most common services would be offered by central information technology units or contractors. While the service activities these requirements call for would not be unique or unusual for IT units or contractors, the activities may be more be a more extensive and expensive undertaking than IT units or contractors normally carry out.[5]

1.1. Operating System Services

This subsection covers the core services needed to operate and administer the application platform, and provide an interface between application software and the platform.

1.1.1. The *Application* SHOULD function on well-supported operating systems and other core infrastructural software.
[CTDR D1.1]

1.1.2. The *Infrastructure* SHOULD provide tools to support system level testing.
[NARA 26.1]

1.1.3. The *Application* SHOULD generate notices to users.
[NARA 23.6]

1.1.4. The *Application* SHOULD support logging of all system events.
[NARA 24.1]

1.1.5. The *Application* SHOULD comply with relevant *de facto* and *de jure* operating systems standards.
[MoReq 11.4]

1.1.6. The *Institution* SHOULD have a process to stay current with the latest operating system security fixes.
[CTDR D1.10]

1.2. Network Services

This subsection covers the capabilities and mechanisms to support distributed applications requiring data access and applications interoperability in heterogeneous networked environments.

1.2.1. The *Infrastructure* SHOULD provide for networked access to records.
[NARA 19]

1.2.2. The networking *Infrastructure* SHOULD be appropriate to the access services

[5] See 3.1 Maintain Guide. The existing requirements literature on digital preservation is particularly sparse on these services, particularly when such services seem necessary for any computer application to operate effectively.

provided and the designated community.
[CTDR D2.1]

1.2.3. If data storage is outsourced or administered externally, there MUST be sufficient network *Infrastructure* to support this service.
[MoReq 11.6]

1.2.4. A network *Application* MUST be able to provide metadata necessary for preservation.
[MoReq 12.1.22]

1.2.5. The networking *Infrastructure* MUST support the security requirements of the Institution.
[ERA 13.6]

1.2.6. The networking *Infrastructure* SHOULD meet or exceed specified performance reliability requirements
[ERA 31.1–31.4]

1.3. Security Services
This subsection covers the capabilities and mechanisms to protect the records in the system. This includes intrusion detection and response and authentication of users. It does not cover all of the security requirements that are covered in the recordkeeping requirements.

1.3.1. An *Application* MUST enable the use of user security profiles.
[MoReq 9.1.8; PRO A.5.11, A.5.17-18, A.5.20-22]

1.3.2. An *Application* MUST enable the use of record security profiles.
[Indiana 1.2.8; DoD c2.2.3.18–c2.2.3.20; NARA 15.2.1]

1.3.3. *Procedures* MUST provide a reasonable guarantee that records are protected from tampering.
[Pitt 9a; PRO A.2.15; ISO 7.1.1, 8.2.2.c; HIPAA 45CFR164.306; NARA 13–14]

1.3.4. *Procedures* MUST prescribe periodic software security updates.
[HIPAA 45CFR164.308]

1.3.5. An *Application* MUST confer exclusive capabilities upon authorized people to exercise the responsibility for creation, modification, annotation, relocation, and destruction of records as defined by an institution.
[Indiana 1.4.1; DoD c2.2.5.2, c2.2.7.4; ISO 8.3.6; HIPAA 45CFR164.308; InterPARES A.2]

1.3.6. An *Application* SHOULD manage the security level(s) of the records it

maintains.
[MoReq 9.3.3, 9.3.5]

1.3.7. *Infrastructure* SHOULD provide services for secure authentication.
[PRO A.5, A.5.1, A.5.11; DoD c2.2.7.1; HIPAA 45CFR164.312]

1.3.8. An *Application* MUST authenticate users before providing services.
[PRO A.5, A.5.1, A.5.11; DoD c2.2.7.1; HIPAA 45CFR164.312; Yale A.4]

1.3.9. An *Application* MUST NOT allow unauthorized changes to the records it maintains.
[Indiana 1.7.1; MoReq 3.2.1, 4.5.4, 6.1.4; PRO A.2.41; DoD c2.2.5.4; ISO 9.7.d; Yale A.4, A.7; NARA 13]

1.3.10. An *Institution* SHOULD undertake a periodic system security analysis of its data systems and identify security risks and needs.
[CTDR D3.1]

1.3.11. An *Institution* SHOULD implement mechanisms to address each of the security needs identified in a system security analysis.
[CTDR D3.2]

1.3.12. *Natural People* MUST have delineated roles, responsibilities, and authorizations.
[CTDR D3.3]

2. **Ingest**
This section includes requirements describing accepting Submission Information Packages from Producers, preparing Archival Information Packages for storage, and ensuring that Archival Information Packages and their supporting Descriptive Information are stored within the Preservation System

2.1. Receive Submission
This subsection includes taking in records transferred from a Producer, a SIP, or an updated SIP produced through some internal processes. This includes accepting all types of records.

2.1.1. An *Institution* MUST confirm that a transfer is authorized.
[NARA 1.2.1; CTDR B1.4]

2.1.2. An *Application* MUST be able to ingest data files in the digital formats in which they were received, as specified by submission agreements.
[NARA 6.1–7.2]

2.1.3. An *Application* SHOULD be capable of accepting transfers via physical media.
[NARA 16.1]

2.1.4. An *Application* SHOULD be capable of accepting transfers electronically.
[NARA 16.2]

2.1.5. An *Application* SHOULD accept electronic records that are composed of more than one digital component.
[NARA 15.2]

2.1.6. An *Application* SHOULD be capable of interacting with all of the institution's recordkeeping applications.
[Indiana 1.6.2; MoReq 6.2.1; PRO A.2.2; NARA 1.10]

2.1.7. An *Application* SHOULD allow Producers to describe the link between record security profiles and records.
[MoReq 4.6.1; PRO A.2.26, A.5.5, A.5.26, A.5.27; ISO 9.7.2; NARA 8.9.5]

2.1.8. An *Institution* SHOULD provide the Producer with progress reports at specific predetermined points throughout the Ingest process.
[CTDR B1.7]

2.1.9. An *Institution* SHOULD mark the formal acceptance of preservation responsibility.
[CTDR B1.9]

2.2. Quality Assurance
This subsection include checking electronic records contained in a transfer, the SIP, in order to verify the success of that transfer.

2.2.1. An *Application* MUST confirm the success of a file transfer (verification of completeness and correctness).
[NARA 16.3; CTDR B1.6]

2.2.2. An *Application* SHOULD be able to technically validate that records components conform to technical file format standards.
[Yale B.4; NARA 5.8]

2.2.3. *Procedures* SHOULD provide for the intellectual validation (data content standard of the metadata is met) of the metadata the records preservation system creates or captures during ingest.
[Indiana 1.6.4; MoReq 6.1.1]

2.2.4. An *Institution* MUST actively produce benchmarks during Ingest in order to monitor the integrity of Archival Information Packages (AIPs).
[CTDRB3.7]

2.2.5. An *Institution* SHOULD confirm that the determinations of the feasibility of

preservation made during the process of appraisal are still valid.
[Tufts-Yale]

2.3. Generate AIP
This subsection includes requirements for transforming a SIP into an AIP that conforms to the data formatting and documentation standards.

2.3.1. An *Application* MUST uniquely identify the records it maintains.
[Pitt 6c; MoReq 7.1; PRO A.9.3; DoD c2.2.1.4, c2.2.4.1; PERM 15; Yale A.5; NARA 1.1.2.1, 19.1.14; CTDR B2.4–B2.5]

2.3.2. An *Application* MUST be able to bind records components together as part of an AIP.
[Tufts-Yale]

2.3.3. An *Institution* MUST define how AIPs are derived from SIPs.
[CTDR B2.1–B2.3]

2.3.4. An *Application* SHOULD facilitate the transformation of record components according to a format transformation plan.
[Tufts-Yale]

2.4. Generate Descriptive Information
This subsection includes extracting metadata from the records and also attaching metadata collected from other sources.

2.4.1. An *Institution* MUST identify the properties of the records it will preserve.
[CTDR B1.1]

2.4.2. An *Application* SHOULD be capable of automatically extracting metadata from the records it captures from a recordkeeping application (including representation information).
[Indiana 1.6.1; MoReq 6.1.6, 6.1.14; Yale B.5; NARA 3.3–3.5; CTDR 3.3, B3.4, B4.1]

2.4.3. An *Application* MUST allow people to manually enter metadata that cannot be automatically extracted from the records captured from a recordkeeping application.
[Indiana 1.6.3; MoReq 6.1.9; PRO A.2.38, PERM 12; NARA 3.3.1.2; CTDR B4.1]

2.4.4. *Procedures* MUST provide for the creation of necessary metadata during the capture process that did not exist before capture (including descriptive, technical, and contextual metadata necessary to document ingest).
[MoReq 6.1.9; PRO A.2.38, PERM 12, Indiana 1.2.3; Pitt 8a; MoReq 6.1.2, 6.1.3; ISO 7.2.1.b; NARA 3.3; CTDR B4.1]

2.4.5. An *Institution* MUST register records with a unique identifier.
[Yale A.5–A.6]

2.4.6. An *Application* MUST be able to technically validate the metadata it creates or captures.
[Indiana 1.6.4; MoReq 6.1.1]

2.4.7. An *Institution* SHOULD use representation information from appropriate community registries.
[CTDR B3.3]

2.4.8. An *Institution* SHOULD generate or acquire preservation metadata.
[CTDR B3.6]

2.5. Coordinate Updates
This subsection includes the transfer of AIPs to Archival Storage and descriptive information to Data Management. This may be the nexus between separate applications for Ingest, Archival Storage, and Data Management.

2.5.1. An *Application* MUST facilitate the transfer of records from Ingest into the record components store.
[Tufts-Yale]

2.5.2. An *Institution* MUST deposit AIPs into its preservation system according to its preservation system rules.
[Tufts-Yale]

2.5.3. An *Institution* MUST update information on preservation actions applied to acquired records.
[Tufts-Yale]

3. Archival Storage
This section includes requirements for the storage, management, and retrieval of Archival Information Packages.

3.1. Receive Data
This subsection includes the storage of AIPs that have successfully completed the Ingest process and now must be stored and maintained by the Archive.

3.1.1. An *Application* MUST generate storage identifiers and document them in the appropriate AIPs.
[Tufts-Yale]

3.1.2. An *Institution* SHOULD gauge anticipated frequency of utilization of AIPs in order to select the most appropriate storage devices or media.

[Tufts-Yale]

3.1.3. An *Application* MUST NOT modify electronic records to accommodate physical storage media.
[NARA 12.2]

3.1.4. An *Application* MAY be capable of adding an "object accession" event to the PDI history.
[Tufts-Yale]

3.2. Manage Storage Hierarchy
This subsection includes administering the record components storage media.

3.2.1. *Procedures* MUST allow for storage media to be maintained in an appropriate physical environment.
[MoReq 11.7.1; ISO 8.3.3; NARA 12.6]

3.2.2. An *Application* SHOULD support high-reliability and redundancy features such as clustering and hot spares.
[Tufts-Yale]

3.2.3. *Infrastructure* MUST be able to support migration to new Storage Hardware Environments.
[Tufts-Yale]

3.2.4. *Procedures* SHOULD describe backup and failure mode activities.
[HIPAA 45CFR164.308, 45CFR164.310; NARA 27.1; CTDR D1.2, D3.5]

3.3. Replace Media
This subsection describes when one of the primary media elements of the records components or administrative metadata store is refreshed or replaced. This can occur preventively or because errors have been detected on the media.

3.3.1. An *Application* MAY provide the automated capability to move electronic records to different media to accommodate new technology.
[NARA 12.1]

3.3.2. *Procedures* MUST allow for the migration of records from one storage media to another in a manner that preserves the recordness of the records.
[Indiana 1.9.1; MoReq 4.4; NARA 12.1, 28.2.4, 28.2.5; CTDR D1.7]

3.3.3. An *Instituion* SHOULD test new media for manufacturing defects before replacing media.
[Tufts-Yale]

3.3.4. An *Application* MAY automatically update PDI for all records affected by a

media replacement with a "media refresh" event.
[Tufts-Yale]

3.4. Error Checking
The AIPs for each record contain fixity information about records components (perhaps in the form of cryptographic checksums, message authentication codes, integrity check-values, modification detection codes, or message integrity codes). This subsection describes the periodic calculation of these fixity information values from the records components and the verification against the existing fixity information values.

3.4.1. *Procedures* SHOULD allow for periodic checks for media deterioration or loss.
[MoReq 11.7.2, 9.1.5; NARA 12.6–12; CTDR D1.5]

3.4.2. An *Institution* MUST actively monitor the integrity of AIPs.
[CTDR B3.7]

3.5. Disaster Recovery
Disaster preparation is listed more than once in these requirements, both here as part of the Archival Storage activity and later (See 4.1.1) as part of the Administration activity. This section refers to requirements necessary to ensure that records components are stored reliably. In 4.1.1 Disaster Preparation describes the development and maintenance of policies and procedures regarding disaster preparation.

3.5.1. An *Application* MUST NOT hinder automated backup of the institution's records.
[Yale-Tufts]

3.5.2. *Procedures* SHOULD require backups to be stored at geographically distant locations.
[PRO A.9.12; DoD c2.2.9.2; Yale C.2; NARA 10.1.4]

3.5.3. An *Application* MUST provide facilities for restoring data from backup data and returning the data stores to a state prior to disaster.
[Pitt 4d; MoReq 11.3.5, 4.3.3, 4.3.4; PRO A.9.14-16; DoD c2.2.9.3, c2.2.2.9.3.1-2, c2.2.9.4-5; HIPAA 45CFR164.308; NARA 10.2.3]

3.5.4. *Infrastructure* SHOULD include tools for recovery of electronic records from failed media.
[NARA 12].

3.5.5. *Procedures* SHOULD provide for the automated backup of the preserved records and preservation metadata.
[MoReq 4.3, 4.3.1, 9.1.2-3; PRO A.9.11, A.9.17; DoD c2.2.9.1]

3.5.6. *Procedures* SHOULD articulate the actions needed to be undertaken during primary system failure.
[Pitt 2d; MoReq 4.3.5; HIPAA 45CFR164.308]

3.5.7. *Juridical People* SHOULD have clearly defined responsibilities to maintain service continuity and to recover from disasters.
[CTDR D3.6]

3.6. Provide Data
This subsection includes providing record components from storage in order to fulfill a request for a DIP from the Access system. This does not include providing copies of record components for internal functions of the Archive.

3.6.1. An *Application* MUST be able to retrieve all the records components of a record.
[Tufts-Yale]

3.6.2. An *Institution* MUST To gather the information required, from descriptive instruments and other preservation information, to satisfy requests for records and/or information about records.
[Tufts-Yale]

3.6.3. An *Application* MAY automatically update retrieval statistics when providing data.
[Tufts-Yale]

4. **Data Management**
This section includes requirements for populating, maintaining, and accessing data that refers to the operation of an Archive.

4.1. Administer Database
This subsection concerns maintaining the integrity of the Data Management database.

4.1.1. An *Application* MUST maintain any links established between ingested records and their metadata (and demonstrate referential integrity).
[Indiana 1.2.3; MoReq 6.1.3; ISO 7.1.c; CTDR B4.2]

4.1.2. An *Application* MUST be able to maintain a record's Preservation Description Information, which documents all events which affect the record.
[Tufts-Yale]

4.1.3. An *Institution* MUST ensure that all actions taken which affect records cause a Preservation Description Information event to be generated.
[Tufts-Yale]

4.1.4. An *Application* MUST manage the relationship between the copies of records components in the system.[6]
[Indiana 1.2.8; DoD c2.2.3.18–c2.2.3.20; NARA 15.2.1]

4.1.5. An *Application* MUST manage the relationship between all copies of records components to their corresponding records.
[NARA 7.4]

4.1.6. An *Application* MAY support identification of the authoritative version (master copy or preservation copy) of a record component in the system.
[InterPARES A.7; NARA 18.5.1]

4.1.7. An *Application* MUST document any changes of a record component from the point of ingest.
[InterPARES B.3; NARA 8.1.5]

4.1.8. An *Application* MUST document items removed from the preservation system, including filenames, timestamps, and person identifiers.
[Yale D.3; NARA 15.8.1]

4.1.9. An *Application* MUST be able to track the logical location of its records copies.
[MoReq 4.4.1; NARA 10.2.4, 10.2.6; CTDR B2.4–B2.5, D1.3]

4.1.10. An *Application* MUST track a record's unique identifier, current location, time of movements, and the natural person responsible for the movements.
[MoReq 4.4.3; ISO 9.8.3; NARA 10.2.6; CTDR B2.4–B2.5]

4.1.11. An *Institution* SHOULD document the security level of the records it maintains.
[MoReq 9.3.3, 9.3.5]

4.1.12. An *Application* MAY provide for management of templates within a template repository.
[NARA 7.2]

4.1.13. An *Application* MAY provide the capability to associate templates with sets of records.
[NARA 7.7.1–7.7.2]

4.1.14. An *Application* MUST allow records security profiles to be stored and

[6] The term "copies" could mean different things in different contexts. In a Fedora repository, different datastreams, surrogates, or derivatives of the same "original" could be considered copies. Record components, the different digital files that combine to form an electronic record, together could also be considered a copy of the record which they are derived from. The reassembly of those digital components would be considered to be a copy of the record from which the components derived.

modified.
[MoReq 9.3.5; PRO A.5.36; NARA 8.9.5, 16.6.2]

4.1.15. An *Application* MUST allow for the linkage between record security profiles and records.
[MoReq 4.6.1; PRO A.2.26, A.5.5, A.5.26, A.5.27; ISO 9.7.2; NARA 8.9.5]

4.1.16. An *Application* SHOULD allow time sensitive records profiles that are valid for a limited time period to be automatically switched to another records security profile when the time period expires.
[PRO A.5.38-39; NARA 13.13.2]

4.1.17. An *Application* MUST allow user security profiles to be stored and modified.
[MoReq 9.1.8; PRO A.5.11, A.5.17-18, A.5.20-22]

4.1.18. An *Application* MUST allow user security profiles to be linked to natural people.
[MoReq 4.1.2, 4.1.5, 4.6.7, 9.1.7; PRO A.5.5, A.5.10, A.5.13, A.5.16, A.5.24; DoD c2.2.7.3]

4.1.19. A *Juridical Person* SHOULD review records before a time-sensitive change is made in the records security profile.
[Tufts-Yale]

4.2. Perform Queries
This subsection includes queries performed by the Archive against records metadata. This might be for a Consumer search, a regular report for Administration, or a maintenance operation.

4.2.1. An *Application* MUST ensure all of its records metadata are discoverable.
[Indiana 1.10.2-3; MoReq 8.1.4-5, 8.1.7; PRO A.3.4, A.3.6, A.3.8, A.3.17; PERM 18; NARA 19]

4.2.2. An *Application* MAY provide integration with external discovery services.
[Tufts-Yale]

4.3. Generate Report
This subsection includes generating reports for internal administrative purposes.

4.3.1. An *Application* MUST report any unauthorized changes to the records it maintains.
[Indiana 1.7.1; MoReq 3.2.1, 4.5.4, 6.1.4; PRO A.2.4.1; DoD c2.2.5.4; ISO 9.7.d; Yale A.4, A.7; NARA 13]

4.3.2. An *Application* SHOULD be able to produce reports for Administration to document any system activity, including failure.

[MoReq 3.4.14; NARA 26.1, 26.3.1, 27.2.4; CTDR B5.2]

4.3.3. An *Application* MUST provide the capability to produce documentation of any reproduction or copy process and its effects, including the dates of the records' reproduction and the name of the responsible person and the impact of the reproduction process on the form of the records components (any changes the records components have undergone).
[InterPARES B.2]

4.4. Receive Database Updates
This subsection describes whenever record metadata is added, modified, or deleted. This can occur when a member of the Archive creates or modifies descriptive metadata, when technical metadata is created or derived from the records, or when supporting records (such as Representation Information (RI), Record Type Records or Producer Records) are updated.

4.4.1. An *Application* MUST enable the addition of new metadata bistreams or the versioning of an existing bitstream as appropriate.
[Tufts-Yale]

4.4.2. An *Application* MUST update an AIP with any new storage identifiers and fixity information.
[Tufts-Yale]

4.4.3. An *Application* MAY automatically update PDI for all affected records with "Metadata Update Event".
[Tufts-Yale]

4.4.4. An *Application* MUST provide the capability to destroy the components of any electronic record.
[NARA 1.5]

5. **Administration**
This section includes requirements for the demonstration of controls over records transfer, maintenance, and reproduction. This refers to auditable documentation proving the existence of such controls but does not include requirements for transfer, maintenance, and reproduction of the records themselves.

5.1. Negotiate Submission Agreement
This subsection includes the actions needed for an Archive and a Producer to generate a Submission Agreement to define the nature and scope of the records to transfer to the preservation system.

5.1.1. An *Institution* MUST specify all appropriate aspects of acquisition, maintenance, preservation, and access issues in written agreements with the Producer.

[CTDR B1.2]

5.1.2. An *Application* MAY automate the implementation of submission agreements.
[NARA 1.6–7]

5.1.3. An *Institution* MUST provide for transfer of custody of records to the Archive.
[NARA 1.3]

5.2. Manage System Configuration
Includes monitoring integrity, reporting Capability and Event Log, and system administration.

5.2.1. An *Institution* SHOULD develop a physical storage media tracking system.
[NARA 11.1]

5.2.2. An *Application* MUST support migration to a new preservation application Hardware Environments.
[Tufts-Yale]

5.2.3. An *Application* SHOULD facilitate the creation, maintenance, and distribution of documentation to support a demonstration of controls over records transfer, maintenance, and reproduction.
[InterPARES B.1; NARA 6; CTDR B3.8]

5.2.4. An *Institution* SHOULD perform a periodic review of its security procedures, including reanalysis of security threats or access management system failures.
[InterPARES B.1.b; HIPAA 45CFR164.308; CTDR B5.2]

5.2.5. An *Application* MUST be able to identify system failures.
[NARA 27.2.1; CTDR B5.2]

5.2.6. An *Application* MAY provide the facility to isolate and resolve failures.
[NARA 27.2.2–27.2.3]

5.2.7. *Procedures* SHOULD contain provisions for all routine maintenance tasks which fall in line with industry best practices.
[Pitt 2c; CTG System; NARA 27; CTDR D1.10]

5.2.8. An *Application* MUST allow convenient access to and the ability to modify any configuration parameters.
[MoReq 11.2.7, 9.1.1; NARA 27.4]

5.2.9. An *Institution* SHOULD identify the necessary hours of Application availability.

[MoReq 11.3; DoD c3.1.3]

5.2.10. *Infrastructure* SHOULD be capable of fulfilling downtime and simultaneous user requirements laid out by the Institution.
[MoReq 11.3]

5.2.11. An *Institution* SHOULD establish a policy to use Representation Information from appropriate international registries.
[CTDR B3.3]

5.2.12. An *Application* MAY provide the capability to monitor the overall system state in a consolidated manner.
[NARA 27.3]

5.2.13. An *Institution* MUST actively monitor the integrity of AIPs.
[CTDR B3.7]

5.2.14. *Infrastructure* SHOULD provide the ability to monitor available storage capacity.
[MoReq 9.14; PRO A.9.21; NARA 27.3.4]

5.2.15. An *Institution* SHOULD determine the maximum number of simultaneous users necessary for the operation of the preservation application.
[MoReq 11.3; DoD c3.1.3]

5.2.16. An *Application* MUST support a stasis mode where no changes are allowed.
[Tufts-Yale]

5.3. Archival Information Update
Includes transformation, and creating copies of records (versions).

5.3.1. An *Institution* MUST ensure that transformations are synchronized across multiple components of records, where appropriate (when content information may be conceived as identical).
[CTDR D1.4]

5.3.2. An *Application* SHOULD provide the capability to transform any ingested data file to a different, more persistent format.
[NARA 8.5-8.6]

5.3.3. An *Application* MUST persistently link the format versions of the same records together.
[PRO A.2.12; NARA 19.8.9]

5.3.4. An *Application* SHOULD automate the synchronization of transformations across multiple copies of records where appropriate.

[CTDR D1.4]

5.4. Physical Access Control
This subsection includes any mechanisms to restrict or allow physical access to elements of the Archive, including doors, locks, guards, etc.

5.4.1. An *Institution* SHOULD create and maintain policies and procedures to detect, contain, and correct security violations.
[HIPAA 45CFR164.308, 45CFR164.310, 45CFR164.312]

5.4.2. *Procedures* MUST provide a reasonable guarantee that records are protected from tampering.
[Pitt 9a; PRO A.2.15; ISO 7.1.1, 8.2.2.c; HIPAA 45CFR164.306]

5.4.3. An *Institution* MUST implement procedures to protect the Archive's facilities and equipment from unauthorized access, tampering, or theft. Such facilities include the physical surroundings of all storage devices and media.
[HIPAA 45CFR164.310]

5.4.4. *Natural People* MUST be authorized to access the Archives' facilities.
[HIPAA 45CFR164.308]

5.4.5. An *Institution* SHOULD implement physical safeguards for all workstations that access electronic records, restricting access to authorized users.
[HIPAA 45CFR164.310]

5.4.6. An *Institution* MUST NOT dispose of records storage media or make it available for re-use without assuring electronic records are removed.
[HIPAA 45CFR164.310]

5.5. Establish Standards and Policies
This subsection concerns the establishment of a variety of standards and policies concerning a variety of issues, including maintain, use rights (institution's management of user's rights).

5.5.1. An *Institution* MUST ensure that all actions taken which affect records cause a Preservation Description Information event to be generated.
[Tufts-Yale]

5.5.2. An *Institution* SHOULD demonstrate it has created and maintains a reasonable access criteria and it has successfully implemented the criteria.
[InterPARES B.1; ISO 8.3.6; NARA 13.2–13.4; CTDR B3.8]

5.5.3. *Procedures* SHOULD exist to redact restricted content from records.
[Pitt 13; MoReq 9.3.10; PRO A.2.56; NARA 20.11.2]

5.5.4. An *Institution* SHOULD establish procedures for the backup and recovery of its records and metadata associated with those records.
[Indiana 1.9, 1.9.4; Pitt 2d; MoReq 4.3.7; InterPARES A.3; ISO 8.3.3; NARA 10.2.3, 14.9; CTDR D1.2, D3.4]

5.5.5. An *Institution* MUST explicitly assign responsibility for the annotation, relocation, and destruction of records on the basis of a person's authority and capacity to carry out the activity (establishing user security profiles).
[Indiana 1.7.2; InterPARES A.2; MoReq 4.6.5, 9.3.5; PRO A.5.36; ISO 9.7; PERM 25; HIPAA 45CFR 164.308, 45CFR164.312; Yale A.5; NARA 13]

5.5.6. An *Institution* SHOULD create policies to manage the security level of the records it maintains.
[MoReq 9.3.3, 9.3.5]

5.5.7. An *Institution* MUST have a policy of not allowing any unauthorized changes to the records it maintains and must have a procedure for documenting any such unauthorized changes.
[Indiana 1.7.1; MoReq 3.2.1, 4.5.4, 6.1.4; PRO A.2.4.1; DoD c2.2.5.4; ISO 9.7.d; Yale A.4, A.7; NARA 13]

5.5.8. An *Institution* SHOULD perform a periodic review of its security procedures, including reanalysis of security threats or access management system failure.
[InterPARES B.1.b; HIPAA 45CFR164.308; CTDR B5.2]

5.5.9. An *Application* MAY provide the capability for Archive staff to create and maintain preservation and access plans, including the ability to alter plans.
[NARA 8.9.1–8.9.6; CTDR B3.10]

5.5.10. An *Application* MAY provide the capability for Archive staff to associate a preservation and access plan with records.
[NARA 8.9.5]

5.5.11. An *Institution* MAY accept all format types in which electronic records are created.
[Tufts-Yale]

5.5.12. *Procedures* MUST be created to guide the Ingest process.
[Tufts-Yale]

5.5.13. An *Institution* MUST have formats standards to guide the terms and conditions of transfer for Ingest.
[Tufts-Yale]

5.5.14. An *Institution* MUST have standards for what metadata it needs to properly document the Ingest process.

[Tufts-Yale]

5.5.15. An *Institution* SHOULD establish policies regarding rendering the functionality of record types.
[Pitt 11b; DoD c2.2.5.3; NARA 8.1.6.5, 20.11.3]

5.5.16. An *Institution* SHOULD establish policies defining the necessary elements of a response to a Consumer request (what is an appropriate response).
[CTDR B5.3]

5.5.17. An *Institution* SHOULD establish policies defining the description necessary to ensure records are discoverable.
[Indiana 1.10.2-3; MoReq 8.1.4-5, 8.1.7; PRO A.3.4, A.3.6, A.3.8, A.3.17; PERM 18; NARA 19]

5.5.18. *Procedures* SHOULD exist for managing record types with templates.
[NARA 7.2]

5.5.19. *Procedures* SHOULD exist for monitoring the available storage.
[Tufts-Yale]

5.5.20. *An Institution* SHOULD establish format transformation policies and plans to implant them.
[Tufts-Yale]

5.6. Audit Submission
This subsection includes verifying that submissions (either SIP or AIP) meet the specifications set out in the Submission Agreement.

5.6.1. An *Application* SHOULD be able to confirm that a transfer is authorized by a submission agreement.
[NARA 1.2; CTDR B1.4]

5.6.2. *Procedures* MUST stipulate validation of a records transfer against its corresponding submission agreement (terms and conditions of transfer).
[NARA 1.2.1–1.2.1.2; CTDR B1.6]

5.6.3. An *Institution* SHOULD provide feedback to the Producer on the success or failure of the transfer.
[Yale B.4; CTDR B1.7]

5.6.4. An *Institution* SHOULD create and maintain policies and procedures to detect, contain, and correct security violations.
[HIPAA 45CFR164.308, 45CFR164.310, 45CFR164.312; NARA 13]

6. **Preservation Planning**
This section includes evaluating the contents of the archive and periodically recommending archival information updates to migrate current archive holdings, developing recommendations for standards and policies, and monitoring changes in the technology environment and in the Designated Community's service requirements and Knowledge Base. This section also includes requirements for developing format transformation plans.

6.1. Monitor Designated Community
This subsection includes tracking the service requirements of Producers and Consumers as well as the state of the art of information technology.

6.1.1. An *Application* SHOULD enable the Archive to track changes in its service requirements and available product technologies to determine when preservation strategies are no longer viable.
[CTDR B3.9]

6.1.2. An *Institution* SHOULD monitor that consumer requests are responded to (see requirement 7.1.18).
[CTDR B5.5]

6.1.3. An *Institution* MUST periodically monitor the acceptability of chosen preservation strategies to existing Consumers and Producers.
[Tufts-Yale]

6.2. Monitor Technology
This includes monitoring preservation strategies standards and best practices against emerging digital technologies, information standards, and computing platforms.

6.2.1. *Juridical People* SHOULD monitor the state of the art of information technology in order to facilitate preservation planning.
[Tufts-Yale]

6.2.2. An *Application* MAY enable the Archive to track emerging digital technologies, information standards and computing platforms (i.e., hardware and software) to determine when preservation strategies are no longer viable.
[CTDR B3.9]

6.2.3. An *Institution* MUST periodically monitor the viability of chosen preservation strategies.
[CTDR B3.9]

6.3. Develop Preservation Strategies and Standards
This subsection includes requirements for developing recommendations for Archive standards and policies to mitigate issues of hardware and software obsolescence and media decay.

6.3.1. *Juridical People* MUST synthesize information about designated communities, technologies, system performance, inventory, and finances in order to recommend preservation strategies and standards.
[Tufts-Yale]

6.3.2. An *Institution* MUST develop plans for preserving records as long as needed and have a written mission statement that reflects a commitment to long-term preservation.
[Indiana 1.9; MoReq 11.7.4; PERM non dod 4; NARA 8.9; CTDR 3.1]

6.3.3. An *Institution* SHOULD develop strategies for ensuring the accessibility and functionality of records components over time.
[InterPARES A.4; ISO 8.3.5, 9.6]

6.3.4. An *Institution* SHOULD develop preservation action plans specifying the preservation actions to be taken to ensure the accessibility and functionality of templates of records components over time.
[InterPARES A.4; ISO 8.3.5, 9.6; NARA 8.9]

6.3.5. An *Institution* SHOULD develop plans for managing preservation metadata and attaching it to records.
[MoReq 5.3.10, 11.7.7; PERM 5, 6]

6.3.6. An *Institution* SHOULD develop strategies for redaction of restricted content from users.
[Pitt 13; MoReq 9.3.10; PRO A.2.56; NARA 20.11.2]

6.3.7. An *Institution* SHOULD develop strategies for rendering the functionality of types of records.
[Pitt 11b; DoD c2.2.5.3; NARA 8.1.6.5, 20.11.3]

6.3.8. An *Institution* SHOULD develop templates to manage record types.
[NARA 7.2]

6.4. Develop Packaging Designs and Migration Plans
This subsection includes the developing of new information package designs and detailed migration plans and prototypes, to implement Administration policies and directives.

6.4.1. An *Institution* MUST create rules for formulating Submission Information Packages.
[Tufts-Yale]

6.4.2. An *Institution* MUST create rules for formulating Archival Information Packages.
[Tufts-Yale]

6.4.3. An *Institution* MAY define records templates to automate Ingest and processing.
[NARA 7]

7. **Access**

This section includes requirements necessary in order to support Consumers in determining the existence, description, location and availability of records stored in the Archive, as well as applying controls to limit access to specially protected records, generating responses, and delivering the responses to Consumers.

7.1. Coordinate Access Activities

This includes the institution's coordinating of the execution of requests to successful completion.

7.1.1. *Procedures* MUST ensure that only authorized users gain access to records.
[MoReq 4.1.1; PRO A.5.25, A.5.42, A.5.46-50; NARA 13]

7.1.2. An *Application* MUST coordinate the application of user security profiles in order to respond to requests.
[MoReq 4.1.2, 4.1.5, 4.6.7, 9.1.7; PRO A.5.5, A.5.10, A.5.13, A.5.16, A.5.24; DoD c2.2.7.3]

7.1.3. An *Application* MUST ensure all of its records and metadata are discoverable.
[Indiana 1.10.2-3; MoReq 8.1.4-5, 8.1.7; PRO A.3.4, A.3.6, A.3.8, A.3.17; PERM 18; NARA 19]

7.1.4. An *Application* MUST be able to render all records returned in a search results list.
[MoReq 8.2.1; PRO A.3.20; DoD c2.2.6.8.10; NARA 19.8]

7.1.5. An *Application* MUST support searching by records' identifiers.
[MoReq 8.1.16, 8.1.23; NARA 19.1.14]

7.1.6. An *Application* MAY provide an integrated search interface.
[MoReq 8.1.2; PRO A.3.7; NARA 19.1, 21]

7.1.7. An *Application* MAY support resource discovery through external interfaces/mechanisms in addition to any integrated search interface.
[PRO A.3.19]

7.1.8. An *Application* SHOULD limit search results to the records the user has rights to access.
[MoReq 4.1.10, 4.1.12, 8.1.28; PRO A.3.18, A.5.51-52, B.3.18; NARA 19.8.3]

7.1.9. An *Application* SHOULD support the full text search of the records and metadata it maintains.
[MoReq 8.1.8; DoD c3.2.9; NARA 19.1.5–19.1.19

7.1.10. An *Application* SHOULD support searching metadata fields containing controlled vocabulary terms managed by thesauri.
[MoReq 8.1.10; PRO A.3.5; DoD c3.2.9; NARA 19.1.3]

7.1.11. An *Application* SHOULD support searching multiple metadata fields and/or full text of records.
[MoReq 8.1.6; PRO A.3.9; DoD c2.2.6.8.2; NARA 19]

7.1.12. An *Application* SHOULD support the use of Boolean and/or relational search operators such as "and" "or" "not" "less than" "greater than" "equal to."
[MoReq 9.1.8; PRO A.3.13; DoD c2.2.6.8.4; NARA 19.1.19]

7.1.13. An *Application* SHOULD support wild card and/or pattern matching searches.
[MoReq 8.1.11; PRO A.3.13; DoD c2.2.6.8.3; NARA 19.1.24]

7.1.14. An *Application* SHOULD support the iterative refinement of a search by adding search conditions to a previously run search—i.e. narrow a search.
[MoReq 8.1.21; NARA 19.9]

7.1.15. An *Application* MAY support word proximity searching.
[MoReq 8.1.12; NARA 19.1.20]

7.1.16. An *Application* MAY support searching null values.
[DoD c2.2.6.8.6]

7.1.17. An *Application* MAY support searching time intervals.
[MoReq 8.1.22]

7.1.18. An *Institution* SHOULD track Consumer requests in order to determine if requests are responded to (see requirement 6.1.2).
[CTDR B5.5]

7.2. Generate Dissemination Information Package (DIP)
This section includes the retrieval of an AIP from Archival Storage and Data Management and the formation of a DIP in order to fulfill a dissemination request.

7.2.1. An *Application* MUST render all of the components of a record along with their associated metadata in a logical manner.
[Indiana 1.10.4; MoReq 8.1.15, 8.2.3; PRO A.3.21–A3.24; DoD c2.2.3.21; PERM 23; NARA 20.9, 20.11]

7.2.2. An *Application* MUST be able to render records on to appropriate output

media, which should at least include graphical display and printer output.
[MoReq 8.2, 8.3, 8.4.1; Pro A.3.25-26, A.3.28-29; PERM 3, 10, 14, 16, 17, 24, non dod 2; NARA 26.3.3, 26.4.1]

7.2.3. An *Application* SHOULD be able to render records into an open export format.
[PRO A.3.31; NARA 26.4.3]

7.2.4. An *Application* SHOULD be able to render records independently of their creating environments.
[MoReq 8.2.2; PRO A.3.22; DoD c3.2.14]

7.2.5. An *Application* SHOULD be able to render a record simultaneously for multiple users.
[PRO A.3.23, DoD c2.2.7.5]

7.2.6. An *Application* SHOULD be able to render all versions of a record.
[DoD c2.2.6.8.9]

7.2.7. An *Institution* SHOULD redact restricted content from records delivered to users that do not have the right to see the restricted content.
[Pitt 13; MoReq 9.3.10; PRO A.2.56; NARA 20.11.2]

7.3. Deliver Response
This subsection includes the online and off line delivery of responses (DIPs, result sets, reports, and assistance) to Consumers.

7.3.1. An *Application* MUST, if it has an integrated search interface, present search results.
[PRO A.3.15; DoD c2.2.6.8.5; NARA 19.8]

7.3.2. An *Application* MAY provide capabilities to manage a search results list including, but not limited to, order, number of hits per page, filter results files, and saving search results.
[MoReq 8.1.17, 8.1.24-25; DoD c2.2.6.8.5; NARA 19.8.4–19.12.3]

7.3.3. An *Institution* MUST answer a consumer request with an appropriate response, either a DIP fulfilling the entire request, a response denying the request, or a DIP fulfilling part of the request accompanied by a response clarifying why the request is only partially fulfilled.
[CTDR B5.3]

7.3.4. An *Institution* MUST disseminate DIPs that are authentic copies of their corresponding SIPs.
[CTDR B5.6]

7.3.5. An *Application* MUST render a record's content.
[Pitt 11, 12; MoReq 8.2.3; PRO A.3.21; PERM 2; NARA 8.1.6.3, 20.11.1]

7.3.6. An *Application* MUST render a record's structure.
[Pitt 12, 12b, 12b1; PRO A.3.21; PERM 2; NARA 8.1.6.6, 20.11.4]

7.3.7. An *Application* MUST render a record's context.
[Pitt 12, 12b1, 12c; ISO 7.25; PERM 2]

7.3.8. An *Application* MUST render a record's functionality.
[Pitt 11b; DoD c2.2.5.3; NARA 8.1.6.5, 20.11.3]

7.3.9. *Procedures* SHOULD provide for the redaction of restricted content from records delivered to users that do not have the right to see the restricted output.
[Pitt 13; MoReq 9.3.10; PRO A.2.56; NARA 20.11.2]

7.3.10. An *Application* MAY be able to create redacted versions of textual, audio, and moving image records.
[MoReq 9.3.10; NARA 18]

7.3.11. An *Application* MUST NOT, if it can redact records, alter the content of a record while creating a redacted version of that record.
[Pitt 13a; PRO A.2.56; NARA 18]

VI. Conclusion

This report presents the results of an effort to illustrate trustworthy electronic recordkeeping and preservation at a college or university. The Requirements for Trustworthy Recordkeeping and Preservation can assist institutions or university archives evaluate existing recordkeeping systems or preservation programs informally, particularly by giving them an outline of issues to address.

One of the strengths of the recordkeeping requirements chapter lies in the fact that it is based on a number of excellent efforts to describe the functional requirements for recordkeeping, from which the project team was able to select those requirements most applicable to university electronic records. The weakness of this section may be that most of the existing documents are focused on very large, complex, centralized recordkeeping environments with vast resources and the ability to purchase, design, create, and support massive recordkeeping applications. This is not the situation at most colleges and universities. Universities are usually not centralized institutions and are often incapable of controlling the recordkeeping environment in the way that is necessary to fulfill the requirements. Also, the resources necessary to administer and maintain such recordkeeping systems are beyond the means of most university archives. The cost of administering a recordkeeping application is enough to cripple the budgets of most archives or records management programs.

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 fixed, numerical order, the requirement are instead best served by residing in database or other environment that allows users the flexibility to arrange individual requirements to suit their needs. For example, a user may only want to see the mandatory requirements or only the requirements that pertain to applications.

The strength of the records preservation requirements chapter is that it is one of the first efforts to define specific system requirements for each of the sections of the OAIS Reference Model. By distributing the existing literature into the sections and subsections of OAIS, it is easier to judge the status of the digital preservation literature to this point. The project team did not match several subsections of the model with requirements from the existing literature either because they do not exist or because they were hidden in a forest of detailed requirements. In those cases, the project team was forced to include Tufts-Yale requirements for necessary activities. In addition, there were two OAIS subsections the project team entirely skipped because there were no requirements from the existing literature and no obvious requirements came to light for each subsection. These lacunae might be fodder for the current analysis of the OAIS Reference Model, or may say something about the effectiveness of the requirements literature to date. There is still more to be accomplished in this area.

Despite these strengths, the requirements presented here are static; they are still not a true evaluation tool. To be used effectively in an assessment, the Requirements must be customized to fit a particular institutional environment. Institutions will need to identify specific sets of requirements which apply to their circumstances, extract them (by source, degree, or section) and

use them in their own contexts. They may also need to add requirements, or further analyze existing requirements. This would transform the requirements into a living document in the Open Source style, in which users would contribute their developments back to the community.[7] Such an effort is beyond the scope of this project and would require an organization to manage the work and provide some sort of tool allowing others to rearrange the requirements and annotate them to suit individual or institutional needs. It is also clear that this is not the final set of functional requirements for electronic recordkeeping and preservation that will be created. MoReq, ISO 15489, CTDR, and the OAIS Reference Model are all currently undergoing revision. Each of these efforts would benefit from a complete understanding of the work of each of the other projects and from a conception of the requirements as dynamic documents to be applied in different ways in different situations.

Archives may be able to leverage the work of other projects to turn the Requirements for Trustworthy Recordkeeping and Preservation into a true evaluation tool. Possible projects that might be of assistance in this effort include the Center for Research Libraries' Digital Repository Certification project or the PLEDGE Project (PoLicy Enforcement in Data Grid Environments), which is developing tools and mechanisms to enable scalable policy expression in digital repositories.[8]

[7] From internal comments by Nancy Y. McGovern on draft trustworthy electronic recordkeeping project report, February 2006.
[8] 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/>.

APPENDIX A: RECORDKEEPING REQUIREMENTS CROSSWALK

The table below presents the recordkeeping requirements from Section IV along with their corresponding place in the August 2005 draft for public comment version of the requirements and their location in the Trusted Digital Repository framework. The full text of the requirement, along with its appropriate sources, is listed in the "Requirement" column. The "Level" column describes if the corresponding row describes a section of the requirements (denoted by an "S"), a subsection (denoted by an "SS"), or an actual requirement (denoted by an "R").

Final Requirement Number	TDR Requirement Mapping	Aug 2005 Requirement Number	Requirement	Level
X	6.1	1	**Compliance** This section covers the identification of and compliance with laws, regulations, standards, and best practices that govern recordkeeping. This section also deals with an institution's ability to demonstrate its compliance with these laws, regulations, standards, and best practices.	S
x	X	1.1	***Identify Laws, Regulations, Standards, Best Practices, and Professional Ethics*** This subsection covers an institution's identification of the laws, regulations, standards, and best practices that govern its recordkeeping practices.	SS
x	6.1.1	1.1.1	An Institution MUST identify the laws, regulations, standards, best practices and professional ethics that affect its recordkeeping activities. [Pitt 1a, 1a1-3; ISO 5, 5a-e]	R
x	6.1.2	1.1.2	An Institution MUST track changes in the laws, regulations, standards, best practices and professional ethics that affect its recordkeeping activities. [Pitt 1c]	R
x	6.1.3	1.1.3	People MUST understand the laws, regulations, standards, best practices and professional ethics that affect their recordkeeping activities. [ISO 8.2.4]	R
x	X	1.2	***Comply with Laws, Regulations, Standards, Best Practices and Professional Ethics*** This subsection covers an institution's compliance with the laws, regulations, standards, and best practices that govern its recordkeeping practices.	SS
x	6.1.4	1.2.1	An Institution MUST comply with the laws, regulations, standards, best practices and professional ethics that affect its recordkeeping activities. [Indiana 1.l, 1.1.1; Pitt 1; MoReq 11.4, 11.5, 11.5.2-3, 11.5.5; PRO A.10.1, A.10.2; ISO 5, 5a-e, 7.1.h, 8.2.4]	R
x	6.1.5	1.2.2	A Recordkeeping Application MUST NOT include any features that do not comply with the laws, regulations, standards, best practices and professional ethics that affect the recordkeeping activities of the institution that the application serves. [MoReq 11.5.4]	R
x	X	1.3	***Demonstrate Compliance with Laws, Regulations, Standards, Best Practices and Professional Ethics*** This subsection covers the demonstration of its compliance with the laws, regulations, standards, and best practices that govern its recordkeeping practices.	SS
x	6.1.6	1.3.1	An Institution SHOULD be able to demonstrate its compliance with the laws, regulations, standards, best practices and professional ethics that affect its recordkeeping activities. [Pitt 1; ISO 5, 5a-e, 8.2.4]	R

2	1.1	2	**Creation and Capture** This section covers the creation and capture of records through recordkeeping systems. It covers the requirements to create records to document activities. It discusses the creation and capture of a variety of standard document types, complex documents, metadata, and relationships between records, along with the process of assigning unique identifiers and normalization during the creation and capture process.	S
2.1	X	2.1	***Generate Records*** This subsection covers the need to create required records to successfully conduct business activities.	SS
2.1.1			2.1.1. An Institution MUST document its activities by creating or capturing records when those activities commit the institution to action, render the institution accountable, or document an action, decision, or decision-making process. [ISO 9.1]	R
2.1.2	1.1.1	2.1.1	An Institution MUST generate records that document all of its defined functions and activities. [Indiana 1.2.1; ISO 7.1.a, 7.2.1, 8.2.5]	R
2.1.3	1.1.2	2.1.2	An Institution MUST ensure its recordkeeping applications are able to capture all of its records. [MoReq 6.1.1; PRO A.2.1, A.2.4, A.2.6]	R
2.1.4	1.1.3	2.1.3	Procedures SHOULD include quality control mechanics to ensure that accurate records are created. [Indiana 1.7; Pitt 7a]	R
2.1.5	1.1.4	2.1.4	People MUST have clearly defined responsibilities for creating records. [ISO 6.3]	R
2.1.6	1.1.7	2.1.5	People SHOULD only create records using documented recordkeeping applications and recordkeeping procedures. [Pitt 3a]	R
2.1.7	x	x	People MUST create and receive records as part of their daily work, and should do so in accordance with established policies, procedures, and standards. [ISO 2.3.2]	
2.1.8	x	x	An Application MUST enable the creation, reception, and keeping of records necessary to support business activities. [ISO 2.3.1]	
2.2	X	2.2	***Preserve Integrity*** This subsection covers the creation and capture of records in a recordkeeping system in a manner that preserves their integrity.	SS
2.2.1	1.1.8	2.2.1	A Recordkeeping Application MUST create and capture records in a manner that maintains the integrity and identity of the records. [Pitt 7a1; InterPARES B.1]	R
2.2.2	1.1.9	2.2.2	A Recordkeeping Application SHOULD validate the integrity of the records it creates and captures. [MoReq 6.2.1]	R
2.2.3	1.1.10	2.2.3	Procedures MUST articulate steps that maintain an unbroken custody of records during capture. [InterPARES B.1.a]	R
2.3	X	2.3	***Preserve Recordness*** This subsection covers the creation and capture of the essential aspects of a record in a recordkeeping system.	SS
2.3.1	1.1.11	2.3.1	An Application MUST be able to create and capture a record's context, structure, and content that together documents the institution's decisions, actions, or communications. [Pitt 7b, 7b1-4; MoReq 6.1.2; PRO A.2.8]	R
2.3.2	1.1.12	2.3.2	Procedures MUST provide for the creation and capture of records in a manner that allows them to correctly reflect the decisions, actions, or communications it documents. [Pitt 7c, 7c1-3; InterPARES A.1.a.i-v, A.1.b.i-iv, A.5]	R

2.4	4.1	2.4	***Support of Format Types*** This subsection covers the creation and capture of records of various formats.	SS
2.4.1	4.1.1	2.4.1	An Institution MUST have recordkeeping applications that together are able to create and capture all of the record formats the institution generates in the course of its business. [MoReq 6.1.1]	R
2.4.2	4.1.2	2.4.2	A Recordkeeping Application SHOULD be able to create and capture records with a variety of format types and structures. [Indiana 1.2.10; MoReq 6.1, 6.3, 6.3.1-2]	R
2.5	4.2	2.5	***Create and Capture Complex Documents*** This subsection covers the creation and capture of complex records.	SS
2.5.1	4.2.1	2.5.1	A Recordkeeping Application MUST, if it is used to manage complex records, be able to create and capture records in a manner that captures the structural integrity of its component parts. [MoReq 6.1.13, 6.3.1, 6.3.2; PRO A.2.5, A.2.8; ISO 7.2.1.a]	R
2.5.2	4.2.2	2.5.2	A Recordkeeping Application MAY adopt one of the following strategies for creating and capturing complex records: As a single compound record or as a series of linked simple records. [Indiana 1.2.7; MoReq 6.3.6]	R
2.6	4.3	2.6	***Create and Capture Relations between Records*** This subsection covers the creation and capture of the relationships between records.	SS
2.6.1	4.3.1	2.6.1	A Recordkeeping Application MUST be able to capture the relationships between records. [PRO A.8.17]	R
2.7	4.4	2.7	***Create and Capture Metadata*** This section covers the creation and capture of metadata associated with records a recordkeeping system creates and captures.	SS
2.7.1	4.4.1	2.7.1	A Recordkeeping Application SHOULD be capable of automatically extracting metadata for the records it creates and captures. [Indiana 1.6.1; MoReq 6.1.6, 6.1.14]	R
2.7.2	4.4.2	2.7.2	A Recordkeeping Application MUST allow people to manually enter metadata that cannot be automatically extracted from the records created and captured by the recordkeeping application. [Indiana 1.6.3; MoReq 6.1.9; PRO A.2.38]	R
2.7.3	4.4.3	2.7.3	Procedures MUST provide for the creation of necessary metadata during the creation and capture process that did not exist before creation or capture. [MoReq 6.1.9; PRO A.2.38]	R
2.7.4	4.4.4	2.7.4	A Recordkeeping Application MUST be able to technically validate the metadata it creates or captures. [Indiana 1.6.4; MoReq 6.1.1]	R
2.7.5	4.4.5	2.7.5	Procedures SHOULD provide for the intellectual validation of the metadata the recordkeeping system creates or captures during the creation or capture process. [Indiana 1.6.4; MoReq 6.1.1]	R
2.7.6	4.4.6	2.7.6	A Recordkeeping Application MUST be able to create and capture descriptive, technical, and contextual metadata. [PERM 12]	R
2.7.7	4.4.7	2.7.7	Procedures SHOULD provide for the creation and capture of descriptive, contextual, and technical metadata. [Indiana 1.2.3; Pitt 8a; MoReq 6.1.2, 6.1.3; ISO 7.2.1.b]	R
2.7.8	4.4.8	2.7.8	A Recordkeeping Application MUST create and capture records and their metadata in a manner that allows them to be persistently linked. [Indiana 1.2.3; MoReq 6.1.3; ISO 7.1.c]	R
2.8	4.5	2.8	***System Interaction*** This subsection covers the ability of a recordkeeping application to communicate and integrate with other recordkeeping and various record creating applications.	SS
2.8.1	4.5.1	2.8.1	A Recordkeeping Application SHOULD be capable of communication with all of the institution's other recordkeeping and record creating applications. [Indiana 1.6.2; MoReq	R

			6.2.1; PRO A.2.2]	
2.8.2	4.5.2	2.8.2	A Recordkeeping Application SHOULD provide an application programming interface to enable integration with other business applications. [PRO A.2.3]	R
X	X	2.9	***Identifier*** This subsection covers the assigning of unique identifiers to records in a recordkeeping system.	SS
2.7.9	4.4.9	2.9.1	A Recordkeeping Application MUST assign unique identifiers to the records it creates and captures. [Indiana 1.2.5; MoReq 7.1.5]	R
2.9	4.6	2.10	***Normalization*** This subsection covers capture and of standard format versions of records in a recordkeeping system captured in other formats. This section does not cover migration, which is covered in Section 7, Preservation. This deals specifically with normalization during the capture process.	SS
2.9.1	4.6.1	2.10.1	A Recordkeeping Application SHOULD be able to capture a standard format version of records it captures in its native format. [PRO A.2.12]	R
2.9.2	4.6.2	2.10.2	A Recordkeeping Application MUST persistently link the format versions of the same records together. [PRO A.2.12]	R
X Now Storage and Handling in preservation requirements	4.7	3	**Maintenance** This section covers the institution's identification and management of records in recordkeeping systems which includes location tracking, versioning management, and unique identifier management. This section also discusses the integration of the recordkeeping systems into the business process and workflow of the institution.	S
4.2	X	3.1	***Preserve Recordness*** This subsection covers the preservation of a record's recordness during its maintenance in a recordkeeping system. (Now Maintain Recordness)	SS
4.2.1	4.7.1	3.1.1	An Institution MUST maintain records in a manner that allows them to correctly reflect the decisions, action, or communication it documents. [ISO 7.2.1]	R
4.2.2	4.7.2	3.1.2	A Recordkeeping Application MUST maintain a record's content, structure, and context that documents the institution's decisions, actions, and communications. [Pitt 7]	R
4.3	X	3.2	***Location Tracking*** This subsection covers the tracking of a record during its maintenance in a recordkeeping system.	SS
4.3.1	4.7.3	3.2.1	A Recordkeeping Application MUST be able to track the location of records in a recordkeeping system. [MoReq 4.4.1]	R
4.3.2	4.7.4	3.2.2	A Recordkeeping Application MUST track a record's unique identifier, current location, time of movements, the person responsible for the movements, and the custodian of the record. [MoReq 4.4.3; ISO 9.8.3]	R
4.3.3	4.7.5	3.2.3	Procedures MUST articulate steps that govern the receipt, removal, and movement of hardware and media that store electronic records. [HIPAA 45CFR164.310]	R
4.4	X	3.3	***Versioning*** This subsection covers the management of versions of records while they are maintained in a recordkeeping system.	SS
4.4.1	4.7.6	3.3.1	A Recordkeeping Application SHOULD support versioning. [Indiana 1.2.9]	R
4.4.2	4.7.7	3.3.2	A Recordkeeping Application MUST, if it supports versioning, manage the relationship between the versions of the same record in a recordkeeping system. [Indiana 1.2.8; DoD c2.2.3.18, c2.2.3.20]	R

4.4.3	4.7.8	3.3.3	A Recordkeeping Application MAY, if it supports versioning, be able to identify the authoritative version of a record in a recordkeeping system that has multiple versions. [IP A.7]	R
4.4.4	4.7.9	3.3.4	A Recordkeeping Application MUST, if it supports versioning, document the version changes of a record since its creation. [InterPARES B.3]	R
X	X	3.4	***Summary for Management*** This subsection covers the ability of managers to receive reports on the management of records while they are maintained in a recordkeeping system.	SS
	6.2	3.4.1	A Recordkeeping Application SHOULD be able to produce reports for administrators on the activities of the records in a recordkeeping system. [MoReq 3.4.14]	R
	X	3.5	***Application Interoperability*** This subsection covers the ability of a recordkeeping application to interoperate with other record creating and keeping applications while it maintains records.	SS
		3.5.1	A Recordkeeping Application SHOULD be able to interoperate with its institution's other applications. [MoReq 10.8.1-4]	R
4.5	X	3.6	***Additional Records Attributes*** This subsection covers the unique identification of a record, the maintenance of its logical relationships and the identification of its custodian(s) during its maintenance in a recordkeeping system.	SS
4.5.1	4.7.10	3.6.1	A Recordkeeping Application MUST uniquely identify the records it maintains. [Pitt 6c; MoReq 7.1; PRO A.9.3; DoD c2.2.1.4, c2.2.4.1; PERM 15]	R
4.5.2	4.7.11	3.6.2	A Recordkeeping Application MUST maintain the logical relationships between records in a recordkeeping system. [MoReq 3.4.11; PRO A.2.24; DoD c2.2.3.17]	R
4.5.3	4.7.12	3.6.3	A Recordkeeping Application MUST maintain the logical relationships between multiple versions of the same record. [DoD c2.2.3.19]	R
4.5.4	4.7.13	3.6.4	A Recordkeeping Application SHOULD identify the responsible custodian(s) of the records it maintains. [PRO A.5.41]	R
3	2.1	4	**Classification** This section covers the development and management of classification schemes, which include records retention schedules, in recordkeeping systems. It also covers the assigning of records to classes within a classification scheme or multiple schemes and the institutional context of these schemes. Although assigning a record to a scheme assigns meaning and prescribes actions to that record, the execution of those actions is not covered in this section.	S
3.1	X	4.1	***Manage Scheme*** This subsection covers the creation, management, and modification of classification scheme(s) within a recordkeeping system. A classification scheme is a logical system used to arrange records. Usually, classes are related component parts that compose a scheme. This section does not cover the act of classifying records.	SS
3.1.1	2.1.1	4.1.1	A Recordkeeping Application MUST allow the creation and defining of a classification scheme. [MoReq 3.1.5; PRO A.1.3, A.4.1; ISO 9.3.A; DoD c2.2.1.1]	R
3.1.2	2.1.2	4.1.2	A Recordkeeping Application MAY allow the creation and defining of multiple classification schemes. [MoReq 3.1.8; PRO A.1.10]	R
3.1.3	2.1.3	4.1.3	A Recordkeeping Application MAY allow the creation and defining of a vital records classification scheme. [DoD c2.2.6.7]	R
3.1.4	2.1.4	4.1.4	A Recordkeeping Application MUST allow the changing, amending, deleting and adding to a classification scheme.	R

			[Indiana 1.8.7; MoReq 3.1.6, 3.4.1; PRO A.1.4, A.1.6, A.1.8, A.4.4, A.4.6]	
3.1.5	2.1.5	4.1.5	A Recordkeeping Application MUST ensure that classification names are unique. [PRO A.1.18]	R
3.1.6	2.1.6	4.1.6	A Recordkeeping Application SHOULD allow the closing of classes within a scheme so that no new records can be added to a closed class. [PRO A.1.7, A.1.41]	R
3.1.7	2.1.7	4.1.7	A Recordkeeping Application MUST NOT allow the deletion of classes that contain records. [PRO A.1.9]	R
3.1.8	2.1.8	4.1.8	A Recordkeeping Application SHOULD NOT impose any practical limits on the number of classes or class levels that exits within a scheme. [MoReq 3.1.3, 3.2.9; PRO A.1.28]	R
3.1.9	2.1.9	4.1.9	A Recordkeeping Application SHOULD report its classes, schemes, and records in a logical, usable fashion. [MoReq 3.2.10; ISO 9.3.6]	R
3.2	X	4.2	***Retention Schedules*** This subsection covers the management and modification of retention schedules along with act of assigning record(s) to a retention schedule(s). Retention schedules prescribe a record's required length of retention and its disposition. Retention schedules are a type classification scheme. This subsection does not cover the execution of a record's disposition.	SS
3.2.1	2.1.10	4.2.1	A Recordkeeping Application MUST be able to assign a retention schedule to a record. [Indiana 1.8.3; MoReq 5.1.4; PRO A.4.14; ISO 8.1.f]	R
3.2.2	2.1.11	4.2.2	A Recordkeeping Application MUST be able to reassign a retention schedule to a record. [PRO A.4.21]	R
3.2.3	2.1.12	4.2.3	An Institution MUST associate retention schedules with dispositions and retention periods and the reasons and sources for these determinations. [Pitt 1b; MoReq 5.1.3, 5.1.11, 5.17, 5.10; PRO A.4.7, A.4.9, A.4.10, A.4.12; ISO 8.1.f, 9.2.c.1-3]	R
3.2.4	2.1.13	4.2.4	An Institution SHOULD be able to change the dispositions and retention periods of the retention schedules. [Indiana 1.8.7; MoReq 5.1.15-16; PRO A.4.6, A.4.1]	R
3.3	X	4.3	***Naming*** This subsection covers the naming of a classification scheme and its classes within a recordkeeping system.	SS
3.3.1	2.1.14	4.3.1	A Recordkeeping Application SHOULD support a naming scheme for classification taxonomies. [MoReq 3.1.4]	R
3.3.2	2.1.15	4.3.2	A Recordkeeping Application MAY support user-defined naming schemes for classification taxonomies. [MoReq 3.1.4]	R
3.3.3	2.1.16	4.3.3	A Recordkeeping Application MAY support the use of controlled vocabulary terms to support the creation of naming schemes. [MoReq 3.2.6, 3.2.8; PRO A.1.24; ISO 9.5.3]	R
3.3.4	2.1.17	4.3.4	A Recordkeeping Application MAY use one of two strategies for creating naming schemes: a structured alpha/numeric system or a human understandable textual system. [MoReq 3.2.2; PRO A.1.14-15]	R
3.3.5	2.1.18	4.3.5	A Recordkeeping Application MAY support the mandatory use of a naming scheme. [PRO A.1.20, A.1.36]	R
3.4	X	4.4	***Assign Classification*** This subsection covers assigning record(s) to a class(es) within a classification scheme in a recordkeeping system. Although assigning a record to a scheme assigns meaning and prescribes actions to that record, the execution of those actions is not covered in this subsection.	SS
3.4.1	2.1.19	4.4.1	An Institution MUST classify records (assign records to a pre-established class in a classification scheme and, within each class, to the dossiers to which they belong. [ISO 4.2.1-4.2.2]	R

3.4.2	2.1.20	4.4.2	A Recordkeeping Application MUST assign all of the records it maintains to a class or multiple classes of a classification scheme. [Indiana 1.8.3, 1.2.4; MoReq 6.1.1; PRO A.2.19, A.2.21, A.4.55]	R
3.4.3	2.1.21	4.4.3	A Recordkeeping Application MUST be able to assign a classification to a particular record that overrides the classification of the group of records that contains the individual record. [MoReq 5.1.14]	R
3.4.4	2.1.22	4.4.4	A Recordkeeping Application MUST be able to reassign a record to a different class. [MoReq 3.4.2, 5.1.16; PRO A.1.47, A.2.50, A.4.21]	R
3.4.5	2.1.23	4.4.5	A Recordkeeping Application MAY support the use of controlled vocabulary terms to support the classification of records.[PRO A.1.37]	R
3.4.6	2.1.24	4.4.6	A Recordkeeping Application MAY support records being classified as vital records. [MoReq 4.3.6]	R
3.5	X	4.5	***Institution Context*** This subsection covers the institutional context into which a classification scheme within a recordkeeping system should fit.	SS
3.5.1	2.1.25	4.5.1	An Institution SHOULD ensure that its recordkeeping applications are compatible with the institution's classification scheme(s). [Indiana 1.3.1; MoReq 3.1.1]	R
3.5.2	2.1.26	4.5.2	An Institution SHOULD ensure its classification scheme(s) reflect its business processes. [ISO 8.2.2.b, 9.5.2]	R
1	2.2	5	**Retention and Disposition** This section covers the act of executing the disposition of records according to a records retention schedule. This usually means the act of removing records and their metadata from the recordkeeping application for either destruction or for transfer to a preservation application. The work also includes reviewing records before carrying out their disposition and the application of legal holds on records that are involved in a legal action, audit, or review. This section does not cover the creation and assigning of records retention schedules. See Subsection 4.2.	S
1.1	X	5.1	***Execution*** This subsection covers the execution of a record's disposition, which usually means either destruction or transfer to a semi-active, inactive, or preservation application.	SS
1.1.1	2.2.1	5.1.1	An Institution SHOULD dispose of records that no longer have operational value, either by permitting (arranging for) their destruction, or by transferring (arranging for) their transfer to a preservation repository. [ISO, 4.3.9, MoReq 5]	R
x	2.2.2	5.1.2	People MUST make a determination on the disposition of the record reviewed. [MoReq 5.2.10, 5.10]	R
1.1.2	2.2.3	5.1.3	Procedures MUST articulate the management of records disposition, in particular the destruction or transfer of records to a preservation system. [MoReq 5.2.10, 5.3.1; ISO 9.9]	R
1.1.3	2.2.4	5.1.4	Procedures MUST allow for the confidential destruction of all copies and instances of records scheduled for destruction. [MoReq 5.3.9; PRO A.4.74, B.3.26; DoD c2.2.10.6; ISO 9.9]	R
1.1.4	2.2.5	5.1.5	A Recordkeeping Application MUST confidentially destroy records scheduled for destruction in a manner that does not allow their recovery. [Pitt 10; MoReq 5.2.13, 5.3.14, 9.3.2; PRO A.4.67-68; DoD c2.2.6.63; ISO 9.9.a; HIPAA 45CFR164.310]	R
1.1.5	2.2.6	5.1.6	A Recordkeeping Application SHOULD be able to retain metadata about records that it destroys. [Pitt 10C; MoReq 5.2.15-16; DoD c2.2.6.6.4]	R
1.1.6	2.2.7	5.1.7	A Recordkeeping Application MUST be able to successfully transfer records scheduled for long-term retention to a preservation system. [MoReq 5.3.3, 5.3.5, 5.3.7; ISO 9.9.c]	R

1.1.7	2.2.8	5.1.8	A Recordkeeping Application SHOULD be able to retain metadata about records that it transfers to a preservation system. [DoD c2.2.6.5.4]	R
1.1.8	2.2.9	5.1.9	A Recordkeeping Application MAY track the actual time of disposition for a record based on the retention schedule assigned to that record. [PRO A.4.29, A.4.35-36, A.4.49]	R
1.2	X	5.2	***Compliance with Schedules*** The subsection covers the need for the disposition of records to be executed in compliance with appropriate retention schedules.	SS
1.2.1	2.2.10	5.2.1	An Institution MUST base the disposition of its records and audit trails on authorized and approved records retention schedules. [Indiana 1.4.2, 1.8, 1.8.1-2; MoReq 3.4.6; PRO A.1.46; ISO 7.1, 9.9]	R
3.2.5	2.2.11	5.2.2	An Institution MUST assign retention schedules to all of its records. [Indiana 1.8.6]	R
1.2.2	2.2.12	5.2.3	A Recordkeeping Application SHOULD be able to manage a variety of retention period configurations and disposition instructions. [DoD c2.2.2.2, c2.2.2.4, c2.2.2.4.1-3, c2.2.2.5]	R
1.2.3	2.2.13	5.2.4	A Recordkeeping Application SHOULD be able to adjust the scheduled disposition of a record if the content of the retention schedule that governs the record changes. [DoD c2.2.2.6, c2.2.2.7]	R
1.3	2.2.14	5.3	***Review*** This subsection covers the review of records before executing their disposition prescribed by their assigned retention schedule.	SS
1.3.1	2.2.15	5.3.1	Procedures MUST articulate steps for reviewing records before their scheduled disposition is executed. [MoReq 5.1.10; 5.2, ISO 9.9]	R
1.3.2	X	5.3.2	A Recordkeeping Application SHOULD alert people of and present to them for review records, including vital records, that has a pending disposition. [Indiana 1.8.4; MoReq 5.1.10, 5.2.1, 5.2.3-4, 5.2.7-8, 9.3.7; PRO A.4.32, A.4.45-46, A.4.64;]	R
1.4	X	5.4	***Legal Holds*** This subsection covers managing the process of suspending the execution of a record's disposition that is a part of any ongoing or reasonably expected legal action or proceedings, litigation, audit, investigation, or review.	SS
1.4.1	2.2.16	5.4.1	An Institution MUST be aware of ongoing and reasonably expected legal action or proceedings, litigation, audit, investigation, or review that involves or may involve its records and identify any records so affected. [Indiana 1.8.5; ISO 9.9]	R
1.4.2	2.2.17	5.4.2	Procedures MUST allow for the interruption of the scheduled disposition of a legal hold on records that are or are expected to be involved in legal action or proceedings, litigation, audit, investigation, or review. [Indiana 1.8.5; PRO A.4.25-26, A.4.38; DoD c2.2.6.4.1; ISO 9.9]	R
1.4.3	2.2.18	5.4.3	Procedures MUST allow for the appropriate lifting of legal holds on records and the resumption of their scheduled disposition. [PRO A.4.27; DoD c2.2.6.4.3]	R
X	5.1	6	**Protective** This section covers the recordkeeping institution's discovery and/or prevention of unauthorized, accidental, or unwanted deletion, change, or corruption of records. It covers access control, detection and response to unauthorized actions, protecting records from tampering, and disaster preparation. All of these activities are undertaken to maintain the data integrity and fixity of records.	S
5.1	X	6.1	***Define Access Controls*** This subsection covers the definition of access controls, or the assigning responsibility for the creation, modification, annotation, relocation, and destruction of records. A more detailed discussion of implementing access controls is in the Use Rights section.	SS

5.1.1	5.1.1	6.1.1	An Institution MUST explicitly assign responsibility for the creation, modification, annotation, relocation, and destruction of records on the basis of a person's authority and capacity to carry out an administrative activity. [Indiana 1.7.2; IP A.2]	R
5.1.2	5.1.2	6.1.2	An Application MUST confer exclusive capabilities upon people to exercise the responsibility for creation, modification, annotation, relocation, and destruction of records as defined by an institution. [Indiana 1.4.1; DoD c2.2.5.2, c2.2.7.4; ISO 8.3.6; HIPAA 45CFR164.308; IP A.2]	R
5.1.3	5.1.3	6.1.3	An Application SHOULD manage the security level of the records it maintains. [MoReq 9.3.3, 9.3.5]	R
5.1.4	5.1.4	6.1.4	An Application MUST NOT allow unauthorized changes to the records it maintains. [Indiana 1.7.1; MoReq 3.2.1, 4.5.4, 6.1.4; PRO A.2.41; DoD c2.2.5.4; ISO 9.7.d]	R
5.1.5	5.1.5	6.1.5	An Application MUST NOT allow unauthorized creation of records. [Pitt 8]	R
5.1.6	5.1.6	6.1.6	An Application MAY tailor its user interface to the user's appropriate access level. [PRO A.8.9]	R
5.1.7	5.1.7	6.1.7	Infrastructure MUST NOT allow unauthorized access to the workstations and hardware that contain or provide access to records. [HIPAA 45CFR164.310]	R
5.1.8	5.1.8	6.1.8	An Institution SHOULD demonstrate it has created and maintains a reasonable access criteria and it has successfully implemented the criteria. [InterPARES A.2; ISO 8.3.6]	R
4.7	X	6.2	***Intrusion Detection and Response*** This subsection covers the detection of and response to unauthorized access to and tampering of records in a recordkeeping system.	SS
4.7.1	5.1.9	6.2.1	An Institution SHOULD create and maintain policies and procedures to detect, contain, and correct security violations. [HIPAA 45CFR164.308, 45CFR164.310, 45CFR164.312]	R
4.7.2	5.1.10	6.2.2	Procedures MUST provide a reasonable guarantee that records are protected from tampering. [Pitt 9a; PRO A.2.15; ISO 7.1.1, 8.2.2.c; HIPAA 45CFR164.306]	R
4.7.3	5.1.11	6.2.3	Procedures MUST prescribe periodic software security updates. [HIPAA 45CFR164.308]	R
4.7.4	5.1.12	6.2.4	An Institution SHOULD perform a periodic review of its security procedures. [InterPARES B.1.b; HIPAA 45CFR164.308]	R
4.7.5	5.1.13	6.2.5	An Application SHOULD be able to facilitate reconstruction, review, and examination of the events surrounding or leading to mishandling of records, or possible compromise of sensitive information. [DoD c2.2.8.3.2]	R
4.7.6	5.1.14	6.2.6	An Institution SHOULD create and maintain policies and procedures to perform regular reviews of audit logs and log-in attempts. [HIPAA 45CFR164.308]	R
4.6	X	6.3	***Disaster Preparation*** This subsection covers the planning for and response to disasters that have an impact on the creation, capture, management, and use of records in a recordkeeping system.	SS
4.6.1	5.1.15	6.3.1	Institution SHOULD create backup and failure mode procedures for its records and vital records. [Indiana 1.9, 1.9.4; Pitt 2d; MoReq 4.3.7; InterPARES A.3; ISO 8.3.3]	R
4.6.2	5.1.16	6.3.2	Procedures SHOULD provide for the automated backup of the institution's records, metadata, audit trails, and configuration settings. [MoReq 4.3, 4.3.1, 9.1.2-3; PRO A.9.11, A.9.17; DoD c2.2.9.1]	R
4.6.3	5.1.17	6.3.3	An Application MUST NOT hinder automated backup of the institution's records. [DoD c2.2.9.1, MoReq 4.3.1]	R
4.6.4	5.1.18	6.3.4	Procedures SHOULD articulate the actions needed to be undertaken during primary system failure. [Pitt 2d; MoReq	R

			4.3.5; HIPAA 45CFR164.308]	
4.6.5	5.1.19	6.3.5	Infrastructure SHOULD allow for backups to be stored at geographically distant locations. [PRO A.9.12; DoD c2.2.9.2]	R
4.6.6	5.1.20	6.3.6	An Application SHOULD provide facilities for restoring data from backup data and returning the data stores to a consistent state. [Pitt 4d; MoReq 11.3.5, 4.3.3, 4.3.4; PRO A.9.14-16; DoD c2.2.9.3, c2.2.2.9.3.1-2, c2.2.9.4-5; HIPAA 45CFR164.308]	R
4.6.7	5.1.21	6.3.7	Institutions SHOULD test and review backup and failure mode procedures. [HIPAA 45CFR164.308, 45CFR164.310]	R
4.1	X	6.4	***Data Integrity*** This subsection covers the enforcement of the data integrity of records in a recordkeeping system.	SS
4.1.1	5.1.22	6.4.1	A Recordkeeping Application MUST enforce data integrity at all times. [MoReq 3.4.12; PRO A.9.2; ISO 8.3.6]	R
X	X	6.5	***Record Fixity*** This subsection covers the maintenance of the fixity of records in a recordkeeping system.	SS
4.1.2	5.1.23	6.5.1	A Recordkeeping Application MUST be able to maintain a record's fixity. [PRO A.2.14, A.2.18; InterPARES B.1.C; DoD c2.2.3.8]	R
X	4.8	7	**Preservation** This section covers the recordkeeping system's procedures and planning process to mitigate issues of media decay and hardware and software obsolescence and to allow the interoperability and openness of its records. This also concerns the ability of a recordkeeping system to transfer records to a preservation system.	S
X	2.3	7.1	***Planning*** This subsection covers the process of establishing plans for preserving records in a recordkeeping system over time.	SS
X	2.3.1	7.1.1	An Institution SHOULD establish plans for preserving records as long as needed. [Indiana 1.9; MoReq 11.7.4; PERM non dod 4]	R
X	2.3.2	7.1.2	An Institution SHOULD establish plans for ensuring the accessibility and functionality of records over time; these may include migration, emulation, and normalization plans. [InterPARES A.4; ISO 8.3.5, 9.6]	R
X	2.3.3	7.1.3	An Institution SHOULD establish plans for managing preservation metadata and attaching it to records. [MoReq 5.3.10, 11.7.7; PERM 5, 6]	R
X	X	7.2	***Preservation System Integration*** This subsection covers the ability of a recordkeeping application to export records to a preservation system.	SS
1.1.9	4.8.1	7.2.1	A Recordkeeping Application SHOULD be able to export records to a preservation system. [PRO A.4.50, A.4.58; PERM non dod 1]	R
1.1.9	4.8.2	7.2.2	A Recordkeeping Application MUST, if it can export records to a preservation system, export records in a manner that preserves their recordness. [PRO A.4.50-52; InterPARES A.8; DoD c2.2.6.5.3; PERM non dod 1]	R
X	X	7.3	***Media Issues*** This subsection covers the management of storage media and the migration of records in a recordkeeping system from one storage media to another.	SS
X	4.8.3	7.3.1	Procedures MUST allow for storage media to be maintained in an appropriate physical environment. [MoReq 11.7.1; ISO 8.3.3]	R
X	4.8.4	7.3.2	Procedures SHOULD allow for periodic checks for media deterioration. [MoReq 11.7.2, 9.1.5]	R

X	4.8.5	7.3.3	Procedures MUST allow for the migration of records from one storage media to another in a manner that preserves the recordness of the records. [Indiana 1.9.1; MoReq 4.4]	R
X	X	7.4	***Technology Obsolescence*** This subsection covers the prevention of records in a recordkeeping system being stranded on obsolete technologies.	SS
X	2.3.4	7.4.1	An Institution MUST plan for and execute strategies for preserving the recordness of their records as it uses new technologies and discontinues use of old ones. [InterPARES A.4; DoD c2.2.10.3, c2.2.10.3.1-4]	R
X	2.3.5	7.4.2	An Institution SHOULD select open, well-documented, and widely-accepted document formats for its record creation in order to combat technology obsolesce. [MoReq 11.7.5]	R
X	X	7.5	***Preserve Recordness*** This subsection covers the preservation of the context, content, structure, and functionality of records in a recordkeeping system.	SS
4.2.3	4.7.14	7.5.1	Procedures MUST preserve context, structure, and content of records throughout all recordkeeping activities. [Pitt 9; PERM non dod5, 2]	R
	X	7.5.2	Procedures MUST preserve the functionality and essential appearance of records throughout all recordkeeping activities. [DoD c2.2.5.3; ISO 8.3.5]	R
4.2.4	4.7.15	7.5.3	Procedures MUST preserve the chain of custody of records throughout all recordkeeping activities. [InterPARES B.1]	R
4.2.5	4.7.16	7.5.4	Procedures MUST preserve the logical boundaries and the relationships between records throughout all recordkeeping activities. [Pitt 9b1, 9b2]	R
5.2	2.4	8	**Use Rights** This section covers the institution's management of users' rights to view and/or receive records. This includes the development, management, and review of records and user security profiles. It also includes the management of access controls and authentication of users.	S
X	X	8.1	***Access Controls*** This subsection covers the development and management of processes that control the access of records in a recordkeeping system.	SS
5.2.1	2.4.1	8.1.1	An Institution MUST develop and implement access control rules for its records. [MoReq 4.6.5; ISO 9.7; PERM 25; HIPAA 45CFR 164.308, 45CFR 164.312]	R
5.2.2	2.4.2	8.1.2	Procedures MUST insure that only authorized users gain access to records. [MoReq 4.1.1; PRO A.5.25, A.5.42, A.5.46-50]	R
5.2.3	2.4.3	8.1.3	An Institution MAY designate people as custodians of records and the custodians are responsible for implementing the access control rules governing their records. [PRO A.5.41, A.5.43-44; ISO 9.7.e]	R
5.2.4	2.4.4	8.1.4	A Recordkeeping Application SHOULD limit search results to the records the user has rights to access. [MoReq 4.1.10, 4.1.12, 8.1.28; PRO A.3.18, A.5.51-52, B.3.18]	R
5.3	5.2	8.2	***Record Security Profile*** This subsection covers the creation and management of security profiles for records in a recordkeeping system. This subsection also covers the assigning of a security profile to a record in a recordkeeping system.	SS
5.3.1	X	X	An Institution MUST create and modify records security profiles. [[ISO 4.3.5]]	
5.3.2	5.2.1	8.2.1	A Recordkeeping Application MUST allow records security profiles to be created and modified. [MoReq 9.3.5; PRO A.5.36]	R

5.3.3	5.2.2	8.2.2	A Recordkeeping Application MUST allow record security profiles to be assigned to records. [MoReq 4.6.1; PRO A.2.26, A.5.5, A.5.26, A.5.27; ISO 9.7.2]	R
5.3.4	5.2.3	8.2.3	A Recordkeeping Application SHOULD allow time sensitive records profiles that are valid for a limited time period to be assigned to records and should automatically be switched to another records security profile when their valid time period expires. [PRO A.5.38-39]	R
X	X	8.3	***User Security Profile*** This subsection covers the creation and management of security profiles for users who use records in a recordkeeping system. This subsection also covers the assigning of a security profile to a user who uses records in a recordkeeping system.	SS
5.3.5	5.2.4	8.3.1	A Recordkeeping Application MUST allow user security profiles to be created and modified. [MoReq 9.1.8; PRO A.5.11, A.5.17-18, A.5.20-22]	R
5.3.6	5.2.5	8.3.2	A Recordkeeping Application MUST assign or reassign user security profiles to people. [MoReq 4.1.2, 4.1.5, 4.6.7, 9.1.7; PRO A.5.5, A.5.10, A.5.13, A.5.16, A.5.24; DoD c2.2.7.3]	R
X	X	8.4	***Authentication*** This subsection covers the process of authenticating users—verifying a user is who he or she purports to be—who are trying to use records in a recordkeeping system.	SS
5.3.7	5.2.6	8.4.1	Infrastructure SHOULD provide services for secure authentication. [PRO A.5, A.5.1, A.5.11; DoD c2.2.7.1; HIPAA 45CFR164.312]	R
5.3.8	5.2.7	8.4.2	A Recordkeeping Application MUST authenticate users before providing services. [PRO A.5, A.5.1, A.5.11; DoD c2.2.7.1; HIPAA 45CFR164.312]	R
X	X	8.5	***Review Security Profiles*** This subsection covers the process of reviewing and modifying user and record security profiles.	SS
5.3.9	5.2.8	8.5.1	Procedures SHOULD allow for the periodic review of access control rules, records security profiles, and user security profiles. [MoReq 4.6.12; PRO A.5.40; ISO 9.7; HIPAA 45CFR164.308]	R
5.3.10	5.2.9	8.5.2	Procedures SHOULD allow for the modification of access control rules, records security profiles, and user security profiles based on the findings of a review. [HIPAA 45CFR164.308]	R
X	4.9	9	**Discovery and Delivery** This section covers the recordkeeping system enabling users to search and discover records along with the system disseminating meaningful and functional records to users. This includes the management of searching mechanisms and query techniques. In addition it covers services to allow browsing and the proper rendering of complex records, a record's recordness, and redacted records.	S
5.3	X	9.1	***Searching*** This subsection covers the capabilities of a recordkeeping system to search the records it maintains.	SS
5.4.1	4.9.1	9.1.1	A Recordkeeping Application MUST ensure all of its records and metadata are discoverable. [Indiana 1.10.2-3; MoReq 8.1.4-5, 8.1.7; PRO A.3.4, A.3.6, A.3.8, A.3.17; PERM 18]	R
5.4.2	4.9.2	9.1.2	A Recordkeeping Application SHOULD provide an integrated search interface. [MoReq 8.1.2; PRO A.3.7]	R
5.4.3	4.9.3	9.1.3	A Recordkeeping Application SHOULD support external search engines in addition to any integrated search interface. [PRO A.3.19]	R
5.4.4	4.9.4	9.1.4	A Recordkeeping Application MUST, if it has an integrated search interface, present search results. [PRO A.3.15; DoD c2.2.6.8.5]	R

5.4.5	4.9.5	9.1.5	A Recordkeeping Application MUST be able to render all records returned in a search results list. [MoReq 8.2.1; PRO A.3.20; DoD c2.2.6.8.10]	R
5.4.6	4.9.6	9.1.6	A Recordkeeping Application SHOULD provide capabilities to manage a search results list including, but not limited to, order, number of hits per page, filter results files, and saving search results. [MoReq 8.1.17, 8.1.24-25; DoD c2.2.6.8.5]	R
5.4.7	4.9.7	9.1.7	A Recordkeeping Application MUST support searching by records' identifiers. [MoReq 8.1.16, 8.1.23]	R
5.4.8	4.9.8	9.1.8	A Recordkeeping Application SHOULD be able to save and reuse queries. [MoReq 8.1.20; PRO A.3.11-12]	R
5.5	X	9.2	***Query Techniques*** This subsection covers the searching techniques a recordkeeping system employs to search the records it maintains.	SS
5.5.1	4.9.9	9.2.1	A Recordkeeping Application SHOULD support the full text search of the records and metadata it maintains. [MoReq 8.1.8; DoD c3.2.9]	R
5.5.2	4.9.10	9.2.2	A Recordkeeping Application SHOULD support searching metadata fields containing controlled vocabulary terms managed by thesauri. [MoReq 8.1.10; PRO A.3.5; DoD c3.2.9]	R
5.5.3	4.9.11	9.2.3	A Recordkeeping Application SHOULD support searching multiple metadata fields and/or full text of records. [MoReq 8.1.6; PRO A.3.9; DoD c2.2.6.8.2]	R
5.5.4	4.9.12	9.2.4	A Recordkeeping Application SHOULD support the use of Boolean and/or relational search operators such as "and" "or" "not" "less than" "greater than" "equal to." [MoReq 9.1.8; PRO A.3.13; DoD c2.2.6.8.4]	R
5.5.5	4.9.13	9.2.5	A Recordkeeping Application SHOULD support wild card and/or pattern matching searches. [MoReq 8.1.11; PRO A.3.13; DoD c2.2.6.8.3]	R
5.5.6	4.9.14	9.2.6	A Recordkeeping Application SHOULD support the iterative refinement of a search by adding search conditions to a previously run search—i.e. narrow a search. [MoReq 8.1.21]	R
5.5.7	4.9.15	9.2.7	A Recordkeeping Application MAY support word proximity searching. [MoReq 8.1.12]	R
5.5.8	4.9.16	9.2.8	A Recordkeeping Application MAY support searching null values. [DoD c2.2.6.8.6]	R
5.5.9	4.9.17	9.2.9	A Recordkeeping Application MAY support searching time intervals. [MoReq 8.1.22]	R
5.6	X	9.3	***Rendering Complex Objects*** This subsection covers the ability of a recordkeeping system to deliver a complex record it maintains to a user in a manner that maintains the full functionality of that record.	SS
5.6.1	4.9.18	9.3.1	A Recordkeeping Application MUST render all of the components of a record and its metadata in a logical manner. [Indiana 1.10.4; MoReq 8.2.3; PRO A.3.21]	R
5.6.2	4.9.19	9.3.2	A Recordkeeping Application MUST be able to render records together with their associated metadata. [MoReq 8.1.15; PRO A.3.24; DoD c2.2.3.21; PERM 23]	R
5.6.3	4.9.20	9.3.3	A Recordkeeping Application MUST be able to render records on to appropriate output mediums which should at least include graphical display and printer output. [MoReq 8.2, 8.3, 8.4.1; Pro A.3.25-26, A.3.28-29; PERM 3, 10, 14, 16, 17, 24, non dod 2]	R
5.6.4	4.9.21	9.3.4	A Recordkeeping Application SHOULD be able to render records into an open export format. [PRO A.3.31]	R
5.6.5	4.9.22	9.3.5	A Recordkeeping Application SHOULD be able to render records independently of their creating environments. [MoReq 8.2.2; PRO A.3.22; DoD c3.2.14]	R

5.6.6	4.9.23	9.3.6	A Recordkeeping Application SHOULD be able to render a record simultaneously for multiple users. [PRO A.3.23, DoD c2.2.7.5]	R
5.6.7	4.9.24	9.3.7	A Recordkeeping Application SHOULD be able to render all versions of a record. [DoD c2.2.6.8.9]	R
5.7	X	9.4	***Rendering Recordness*** This subsection covers the ability of a recordkeeping system to deliver a record it maintains to a user in a manner that fully maintains the record's context, structure, and content.	SS
5.7.1	4.9.25	9.4.1	A Recordkeeping Application MUST render a record's content. [Pitt 11, 12; MoReq 8.2.3; PRO A.3.21; PERM 2]	R
5.7.2	4.9.26	9.4.2	A Recordkeeping Application MUST render a record's structure. [Pitt 12, 12b, 12b1; PRO A.3.21; PERM 2]	R
5.7.3	4.9.27	9.4.3	A Recordkeeping Application MUST render a record's context. [Pitt 12, 12b1, 12c; ISO 7.25; PERM 2]	R
5.7.4	4.9.28	9.4.4	A Recordkeeping Application MUST render a record's functionality. [Pitt 11b; DoD c2.2.5.3]	R
5.8	X	9.5	***Availability*** This subsection covers the availability of needed records in a recordkeeping system.	SS
5.8.1	4.9.29	9.5.1	A Recordkeeping Application MUST ensure that records needed for their primary business functions are available. [Indiana 1.10, 1.10.1; Pitt 12a; ISO 8.3.6]	R
5.8.2	4.9.30	9.5.2	A Recordkeeping Application SHOULD ensure that records needed for secondary use are available. [Indiana 1.10, 1.10.1; Pitt 12a]	R
5.8.3	4.9.31	9.5.3	A Recordkeeping Application MUST ensure that its records are available in a timely manner. [Indiana 1.10.1; Pitt 12a; ISO 8.3.6]	R
5.9	X	9.6	***Browsing*** This subsection covers the ability of a recordkeeping application to provide users the capability to browse records.	SS
5.9.1	4.9.32	9.6.1	The Recordkeeping Application SHOULD support the browsing of its classification schemes, including any hierarchical structure in which the records are managed. [MoReq 8.1.13, 8.1.27, 3.1.7; PRO A.3.3; DoD c2.2.1.6]	R
5.10	X	9.7	***Redaction*** This subsection covers the management and execution of redacting records and the delivery redacted versions of records to users.	SS
5.10.1	4.9.33	9.7.1	Procedures SHOULD provide for the redaction of restricted content from records delivered to users that do not have the right to see the restricted output. [Pitt 13; MoReq 9.3.10; PRO A.2.56]	R
5.10.2	4.9.34	9.7.2	A Recordkeeping Application SHOULD be able to create redacted versions of textual, audio, and moving image records. [MoReq 9.3.10]	R
5.10.3	4.9.35	9.7.3	A Recordkeeping Application MUST NOT, if it can redact records, alter the content of a record while creating a redacted version of that record. [Pitt 13a; PRO A.2.56]	R
6	4.10	10	**Design and Performance** This section covers the software and hardware design and performance of the recordkeeping application, including system maintenance, scalability, design constraints, and testing and verification. This section also covers the application's usability.	S
6.1	X	10.1	***Testing and Verification*** This subsection covers the testing and verification of the recordkeeping application's and the infrastructure's performance.	SS
6.1.1	4.10.1	10.1.1	An Institution SHOULD determine an appropriate suite of tests against which the recordkeeping infrastructure and recordkeeping application will be measured and set acceptable ranges for system performance. [Indiana 1.12;	R

			MoReq 11.2, 11.2.5]	
6.1.2	4.10.2	10.1.2	Procedures SHOULD include provisions for regular execution of application and infrastructure tests. [Indiana 1.12; PRO A.9.22]	R
6.1.3	4.10.3	10.1.3	Infrastructure SHOULD reliably pass all tests and perform within stated acceptable ranges. [Indiana 1.12; Moreq 11.2]	R
6.1.4	4.10.4	10.1.4	A Recordkeeping Application SHOULD reliably pass all tests and perform within stated acceptable ranges. [Indiana 1.13; PRO A.9.22; MoReq 11.2, 11.2.1-4]	R
6.1.5	4.10.5	10.1.5	A Recordkeeping Application SHOULD undergo formal verification and be provably correct. [Pitt 4b, 4c]	R
6.2	X	10.2	***System Maintenance*** This subsection covers the maintenance of the recordkeeping application and infrastructure.	SS
6.2.1	4.10.6	10.2.1	Procedures SHOULD contain provisions for all routine maintenance tasks which fall in line with industry best practices. [Pitt 2c; CTG System]	R
6.2.2	4.10.7	10.2.2	A Recordkeeping Application MUST allow convenient access to and the ability to modify any configuration parameters. [MoReq 11.2.7, 9.1.1]	R
6.2.3	4.10.8	10.2.3	Infrastructure SHOULD provide the ability to monitor available storage capacity. [MoReq 9.14; PRO A.9.21	R
6.2.4	4.10.9	10.2.4	An Institution SHOULD determine the acceptable ranges for downtime and minimum numbers of simultaneous users. [MoReq 11.3; DoD c3.1.3]	R
6.2.5	4.10.10	10.2.5	Infrastructure SHOULD be capable of fulfilling downtime and simultaneous user requirements laid out by the institution. [MoReq 11.3]	R
6.3	X	10.3	***User Interface*** This subsection covers the user interfaces of a recordkeeping application.	SS
6.3.1	4.10.11	10.3.1	A Recordkeeping Application SHOULD provide a user interface which is easy to use. [MoReq 11.1; PRO A.8.11; DoD c2.2.5.1]	R
6.3.2	4.10.12	10.3.2	A Recordkeeping Application SHOULD follow generally accepted user interface guidelines by providing a consistent look and feel. [PRO 8.1-3]	R
6.3.3	4.10.13	10.3.3	A Recordkeeping Application MAY provide a remote login facility. [MoReq A.9.7]	R
6.3.4	4.10.14	10.3.4	A Recordkeeping Application SHOULD facilitate use by persons with disabilities by including accessibility features. [PRO A.8.16]	R
6.3.5	4.10.15	10.3.5	A Recordkeeping Application SHOULD provide meaningful error messages in the event of an error, and attempt to guide the user to an appropriate resolution. [PRO A.8.7-8]	R
6.4	X	10.4	***Scalability*** This subsection covers scalability of the recordkeeping system..	SS
6.4.1	X	X	An Institution MUST create rules for formulating Submission Information Packages. [Tufts-Yale]	
6.4.2	X	X	An Institution MUST create rules for formulating Submission Information Packages. [Tufts-Yale]	
6.4.3	4.10.16	10.4.1	A Recordkeeping Application SHOULD be able to both scale up to large organizations, and scale down for smaller organizations. [MoReq 11.2.6, 11.2.8]	R
6.4.2	4.10.17	10.4.2	Institutions SHOULD estimate its medium and long-term scalability requirements and determine acceptable ranges for various scalability metrics. [PRO A.9.23]	R

6.4.3	4.10.18	10.4.3	A Recordkeeping Application SHOULD be capable of fulfilling its institution's scalability requirements, and of operating within acceptable ranges. [PRO A.9.23]	R
6.4.4	4.10.19	10.4.4	A Recordkeeping Application SHOULD NOT impose any practical limit on the number of records which can be managed by the application. [MoReq 6.3.5; PRO A.2.20]	R
6.4.5	4.10.20	10.4.5	A Recordkeeping Application SHOULD provide the ability to synchronize multiple instances of all underlying data stores. [DoD c2.2.3.24]	R
6.5	X	10.5	***Design Constraints*** This subsection covers the design constraints of the recordkeeping application.	SS
6.5.1	4.10.21	10.5.1	A Recordkeeping Application SHOULD be designed around a flexible architecture which can evolve as the institution's needs change. [PRO A.9.1]	R
6.5.2	4.10.22	10.5.2	A Recordkeeping Application MAY support a distributed repository with multi-site service. [PRO A.9.18]	R
6.5.3	4.10.23	10.5.3	A Recordkeeping Application SHOULD provide at least one version of backward compatibility. [DoD c2.1.4]	R
6.4.6	4.10.24	10.5.4	A Recordkeeping Application SHOULD, when it offers remote or distributed services, use efficient network protocols which minimize the amount of data exchange required. [PRO A.9.20]	R

APPENDIX B: PRESERVATION REQUIREMENTS CROSSWALK

The table below presents the preservation requirements from Section V along with their corresponding place in the August 2005 draft for public comment version of the requirements. The full text of each requirement, along with its appropriate sources, is listed in the "Requirement" column. The "Level" column describes if the corresponding row describes a section of the requirements (denoted by an "S"), a subsection (denoted by an "SS"), or an actual requirement (denoted by an "R").

Final Requirements Number	Aug 2005 Requirements Number	Requirement	Level
1	x	**Common Services**	S
1.1	x	**Operating system services**	SS
1.1.1	x	The Application SHOULD function on well-supported operating systems and other core infrastructural software. [CTDR D1.1]	R
1.1.2	x	The Infrastructure SHOULD provide tools to support system level testing. [NARA 26.1]	R
1.1.3	x	The Application SHOULD generate notices to users. [NARA 23.6]	R
1.1.4	x	The Application SHOULD support logging of all system events. [NARA 24.1]	R
1.1.5	x	The Application SHOULD comply with relevant de facto and de jure operating systems standards. [MoReq 11.4]	R
1.1.6	x	The Institution SHOULD have a process to stay current with the latest operating system security fixes. [CTDR D1.10]	R
1.2	x	**Network Services**	SS
1.2.1	x	The Infrastructure SHOULD provide for networked access to records.[NARA 19]	R
1.2.2	x	The networking Infrastructure SHOULD be appropriate to the access services provided and the designated community.[CTDR D2.1]	R
1.2.3	x	If data storage is outsourced or administered externally, there MUST be sufficient network Infrastructure to support this service. [MoReq 11.6]	R
1.2.4	x	A network Application MUST be able to provide metadata necessary for preservation. [MoReq 12.1.22]	R
1.2.5	X	The networking Infrastructure MUST support the security requirements of the instution. [ERA13.6]	
1.2.6	X	The networking Infrastructure SHOULD meet or exceed specified performance reliability requirements [ERA 31.1–31.4]	
1.3	x	**Security Services**	SS
1.3.1	6.1.3	An Application MUST enable the use of user security profiles. [MoReq 9.1.8; PRO A.5.11, A.5.17-18, A.5.20-22]	R
1.3.2	6.1.2	An Application MUST enable the use of record security profiles. [Indiana 1.2.8; DoD c2.2.3.18–c2.2.3.20; NARA 15.2.1]	R
1.3.3	x	Procedures MUST provide a reasonable guarantee that records are protected from tampering. [Pitt 9a; PRO A.2.15; ISO 7.1.1, 8.2.2.c; HIPAA 45CFR164.306; NARA 13–14]	R
1.3.4	x	Procedures MUST prescribe periodic software security updates. [HIPAA 45CFR164.308]	R
1.3.5	X	An Application MUST confer exclusive capabilities upon authorized people to exercise the responsibility for creation, modification, annotation, relocation, and destruction of records as defined by an institution. [Indiana 1.4.1; DoD c2.2.5.2, c2.2.7.4; ISO 8.3.6; HIPAA 45CFR164.308; InterPARES A.2]	R

1.3.6	x	An Application SHOULD manage the security level of the records it maintains. [MoReq 9.3.3, 9.3.5]	R
1.3.7	x	Infrastructure SHOULD provide services for secure authentication. [PRO A.5, A.5.1, A.5.11; DoD c2.2.7.1; HIPAA 45CFR164.312]	R
1.3.8	x	An Application MUST authenticate users before providing services. [PRO A.5, A.5.1, A.5.11; DoD c2.2.7.1; HIPAA 45CFR164.312; Yale A.4]	R
1.3.9	x	An Application MUST NOT allow unauthorized changes to the records it maintains. [Indiana 1.7.1; MoReq 3.2.1, 4.5.4, 6.1.4; PRO A.2.41; DoD c2.2.5.4; ISO 9.7.d; Yale A.4, A.7; NARA 13]	R
1.3.10	x	An Institution SHOULD undertake a periodic system security analysis of its data systems and identify security risks and needs. [CTDR D3.1]	R
1.3.11	x	An Institution SHOULD implement mechanisms to address each of the defined security needs. [CTDR D3.2]	R
1.3.12	x	Natural People MUST have delineated roles, responsibilities, and authorizations. [CTDR D3.3]	R
2	1	**Ingest**	S
2.1	x	**Receive Submission**	SS
2.1.9	x	An Institution SHOULD provide the Producer with progress reports at specific predetermined points throughout the Ingest process. [CTDR B1.7]	R
2.1.10	x	An Institution SHOULD mark the formal acceptance of preservation responsibility. [CTDR B1.9]	R
2.2	x	**Quality Assurance**	SS
2.2.5	X	An Institution SHOULD confirm that the determinations of the feasibility of preservation made during the process of appraisal are still valid. [Tufts-Yale]	R
2.3	x	**Generate AIP**	SS
2.3.3	x	An Institution MUST define how AIPs are derived from SIPs. [CTDR B2.1–B2.3	R
2.3.4	x	An Application SHOULD facilitate the transformation of record components according to the format transformation plan. [Tufts-Yale]	R
2.4	x	**Generate Descriptive Information**	SS
2.4.1	x	An Institution MUST identify the properties of the records it will preserve. [CTDR B1.1]	R
2.4.8	x	An Institution SHOULD generate or acquire preservation metadata.[CTDR B3.6]	R
2.5	x	**Coordinate Updates**	SS
2.5.1	x	An Application MUST facilitate the transfer of records from Ingest into the record components store. [Tufts-Yale]	R
2.5.2	x	An Institution SHOULD deposit AIPs into its preservation system according to its preservation system rules. [Tufts-Yale]	R
2.5.3	X	An Institution MUST update information on preservation actions applied to acquired records. Tufts-Yale]	R
X	1.2.	***Manage transfers***	SS
X	1.2.1.	*Accessioning*	SSS
2.4.5	1.2.1.1.	An Institution MUST register transfers with a unique identifier. [Yale A.5–A.6]	R
X	1.2.2.	*Capture Information about Records*	SSS
2.4.2	1.2.2.1.	An Application SHOULD be capable of automatically extracting metadata for the records it captures from a recordkeeping application (including representation information). [Indiana 1.6.1; MoReq 6.1.6, 6.1.14; Yale B.5; NARA 3.3–3.5; CTDR 3.3, B3.4, B4.1]	R

2.4.3	1.2.2.2.	An Application MUST allow people to manually enter metadata that cannot be automatically extracted from the records captured from a recordkeeping application. [Indiana 1.6.3; MoReq 6.1.9; PRO A.2.38, PERM 12; NARA 3.3.1.2; CTDR B4.1]	R
2.4.4	1.2.2.3.	Procedures MUST provide for the creation of necessary metadata during the capture process that did not exist before capture (including descriptive, technical, and contextual metadata necessary to document ingest). [MoReq 6.1.9; PRO A.2.38, PERM 12, Indiana 1.2.3; Pitt 8a; MoReq 6.1.2, 6.1.3; ISO 7.2.1.b; NARA 3.3; CTDR B4.1]	R
4.1.1	1.2.2.4.	An Application MUST maintain any links established between ingested records and their metadata (and demonstrate referential integrity). [Indiana 1.2.3; MoReq 6.1.3; ISO 7.1.c; CTDR B4.2]	R
X	1.3.	***Accept all types of electronic records***	SS
5.5.11	1.3.1.	An Institution MAY accept all format types in which electronic records are written.	R
2.1.2	1.3.2.	An Application MUST be able to ingest data files in the digital formats in which they were received, as specified by submission agreements. [NARA 6.1–7.2]	R
2.1.3	1.3.3.	An Application SHOULD accept transfers via physical media. [NARA 16.1]	R
2.1.4	1.3.4.	An Application SHOULD accept transfers electronically. [NARA 16.2]	R
2.1.5	1.3.5.	An Application SHOULD accept electronic records that are composed of more than one digital component. [NARA 15.2]	R
2.1.6	1.3.6.	An Application SHOULD be capable of interacting with all of the institution's recordkeeping applications. [Indiana 1.6.2; MoReq 6.2.1; PRO A.2.2; NARA 1.10]	R
X	1.4.	***Check electronic records contained in a transfer***	SS
2.1.1	1.4.1.	An Institution MUST confirm that the transfer is authorized. [NARA 1.2.1; CTDR B1.4]	R
5.6.1	1.4.2.	An Application SHOULD be able to confirm that a transfer is authorized by a submission agreement. [NARA 1.2; CTDR B1.4]	R
5.6.2	1.4.3.	Procedures MUST be able to validate that a records transfer complies with the submission agreement (terms and conditions of transfer). [NARA 1.2.1–1.2.1.2; CTDR B1.6]	R
2.2.1	1.4.4.	An Application MUST confirm the success of a file transfer (verification). [NARA 16.3; CTDR B1.6]	R
2.2.2	1.4.5.	An Application SHOULD be able to technically validate that records components conform to technical file format standards. [Yale B.4; NARA 5.8]	R
5.6.3	1.4.6.	An Institution SHOULD provide feedback to the producer on the success or failure of the transfer. [Yale B.4; CTDR B1.7]	R
2.4.6	1.4.7.	An Application MUST be able to technically validate the metadata it creates or captures. [Indiana 1.6.4; MoReq 6.1.1]	R
2.2.3	1.4.8.	Procedures SHOULD provide for the intellectual validation of the metadata the records preservation system creates or captures during ingest. [Indiana 1.6.4; MoReq 6.1.1]	R
3	2	**Archival Storage**	S
3.1	x	**Receive Data**	SS
3.1.1	x	An Application MUST generate storage identifiers and document them in the appropriate AIPs.[Tufts-Yale]	R
3.1.2	x	An Institution SHOULD gauge anticipated frequency of utilization of AIPs in order to select the most appropriate storage devices or media. [Tufts-Yale]	R
		An Application MAY be capable of adding an "object accession" event to the PDI history. [Tufts-Yale]	R
3.2	x	**Manage Storage Hierarchy**	
3.3	x	**Replace Media**	SS

3.3.3	x	Before replacing media, the Instituion SHOULD test new media for manufacturing defects.[Tufts-Yale]	SS
3.3.4	x	An Application MAY automatically update PDI for all records affected by a media replacement with a “media refresh” event. [Tufts-Yale]	R
3.4	x	**Error Checking**	SS
3.5	x	**Disaster Recovery**	SS
3.5.7	x	Juridical People SHOULD have clearly defined responsibilities to maintain service continuity and recovery from disasters.[CTDR D3.6]	R
3.6	x	**Provide Data**	SS
3.6.2	X	An Institution MUST To gather the information required, from descriptive instruments and other preservation information, to satisfy requests for records and/or information about records. [Tufts-Yale]	R
3.6.3	x	An Application MAY automatically update retrieval statistics when providing data. [Tufts-Yale]	R
X	2.1.	***Store records reliably*** (including Protection from Loss or Corruption)	SS
X	2.1.1	***Intrusion Detection and Response***	SSS
1.3.3	2.1.1.1	Procedures MUST provide a reasonable guarantee that records are protected from tampering. [Pitt 9a; PRO A.2.15; ISO 7.1.1, 8.2.2.c; HIPAA 45CFR164.306; NARA 13–14]	R
1.3.3	2.1.1.2	Procedures MUST prescribe periodic software security updates. [HIPAA 45CFR164.308]	R
X	2.1.2	*Disaster Preparation* Disaster preparation is listed more than once in these requirements, both here as part of the Archival Storage activitiy and later (See 4.1.1) as part of the Administration activity. This section refers to requirements necessary to ensure that records components are stored reliably. Later Disaster Prepartion is described in terms of the development and maintenance of policies and procedures regarding disaster preparation.	SSS
3.5.1	2.1.2.1	An Application MUST NOT hinder automated backup of the institution’s records.	R
3.5.2	2.1.2.2	Infrastructure SHOULD allow for backups to be stored at geographically distant locations. [PRO A.9.12; DoD c2.2.9.2; Yale C.2; NARA 10.1.4]	R
3.5.3	2.1.2.3	An Application MUST provide facilities for restoring data from backup data and returning the data stores to a state prior to disaster. [Pitt 4d; MoReq 11.3.5, 4.3.3, 4.3.4; PRO A.9.14-16; DoD c2.2.9.3, c2.2.2.9.3.1-2, c2.2.9.4-5; HIPAA 45CFR164.308; NARA 10.2.3]	R
X	2.1.3	***Manage the Preservation Process***	SSS
X	2.1.3.1	***Media Decay*** (manage media)	SSSS
3.4.1	2.1.3.1.1	Procedures SHOULD allow for periodic checks for media deterioration or loss. [MoReq 11.7.2, 9.1.5; NARA 12.6–12; CTDR D1.5]	R
5.2.1	2.1.3.1.2	An Institution SHOULD develop a physical storage media tracking system. [NARA 11.1]	R
3.2.1	2.1.3.1.3	Procedures MUST allow for storage media to be maintained in an appropriate physical environment. [MoReq 11.7.1; ISO 8.3.3; NARA 12.6]	R
3.3.2	2.1.3.1.4	Procedures MUST allow for the migration of records from one storage media to another in a manner that preserves the recordness of the records. [Indiana 1.9.1; MoReq 4.4; NARA 12.1, 28.2.4, 28.2.5; CTDR D1.7]	R
3.1.3	2.1.3.1.5	An Application MUST NOT modify electronic records to accommodate physical storage media. [NARA 12.2]	R
3.3.1	2.1.3.1.6	An Application MAY provide the automated capability to move electronic records to different media to accommodate new technology. [NARA 12.1]	R
3.5.4	2.1.3.1.7	An Institution SHOULD possess tools for recovery of electronic records from failed media. [NARA 12.4]	R

X	2.1.3.2	***Hardware and Software Obsolescence***	SSSS
X	2.1.3.2.1	An Application MUST have the capability to store copies of electronic records [NARA 10.1]	R
X	2.1.3.3	***Transformation***	SSSS
5.3.1	2.1.3.3.1	An Institution MUST ensure that transformations are synchronized across multiple copies of records, where appropriate (when content information may be conceived as identical). [CTDR D1.4]	R
5.3.2	2.1.3.3.2	An Application SHOULD provide the capability to transform any ingested data file to a different, more persistent format. [NARA 8.5-8.6]	R
5.3.3	2.1.3.3.3	An Application MUST persistently link the format versions of the same records together. [PRO A.2.12; NARA 19.8.9]	R
X	2.1.3.3.4	Upon any transformation, an Application SHOULD NOT involve an irreversible conversion from one data format to another. [Yale A.9; NARA 8]	R
5.3.4	2.1.3.3.5	An Application SHOULD automate the synchronization of transformations across multiple copies of records where appropriate. [CTDR D1.4]	R
X	2.1.4	***Destruction of Records***	SSS
4.4.4	2.1.4.1	An Application MUST provide the capability to destroy the components of any electronic record [NARA 1.5]	R
X	2.1.5	***Monitor Integrity***	SSS
2.2.4	2.1.5.1	An Institution MUST actively monitor the integrity of AIPs [CTDRB3.7]	
5.2.15	2.1.5.1	An Institution MUST actively monitor the integrity of AIPs [CTDRB3.7]	R
X	2.1.5.2	An Application SHOULD be able to check records components using stored signatures or checksums	R
X	2.1.5.3	An Application MUST be able to discover records components which are not enclosed in an AIP (eg, after data loss)	R
X	2.2	***Hardware Replacement and Maintenance***	SS
5.2.2	2.2.1	An Application MUST support migration to new Preservation Application Hardware Environments	R
3.2.2	2.2.2	An Application SHOULD support high-reliability and redundancy features such as clustering and hot spares	R
3.2.3	2.2.3	Infrastructure MUST be able to support migration to new Storage Hardware Environments	R
X	2.3	***Basic Functions***	SS
2.3.2	2.3.1	An Application MUST be able to store records components and bind records components together with an AIP	R
3.6.1	2.3.2	An Application MUST be able to retrieve all the records components of a record	R
4.1.2	2.3.3	An Application MUST be able to maintain a record's Preservation Description Information, which documents all events which affect the record	R
5.5.1	2.3.4	An Institution MUST ensure that all actions taken which affect records cause a Preservation Description Information event to be generated	R
4	3	**Data Management**	S
4.1	x	**Administer Database**	SS
4.2	x	**Perform Queries**	SS
4.2.1	x	An Application MUST ensure all of its records metadata are discoverable. [Indiana 1.10.2-3; MoReq 8.1.4-5, 8.1.7; PRO A.3.4, A.3.6, A.3.8, A.3.17; PERM 18; NARA 19]	R

4.2.2	x	An Application MAY provide integration with external discovery services. [Tufts-Yale]	R
4.3	x	**Generate Report**	SS
4.4	x	**Receive Database Updates**	SS
4.4.1	x	An Application MUST enable the addition of new metadata bistreams or the versioning of an existing bitstream as appropriate. [Tufts-Yale]	R
4.4.2	x	An Application MUST update an AIP with any new storage identifiers and fixity information. [Tufts-Yale]	R
4.4.3	x	An Application MAY automatically update PDI for all affected records with “Metadata Update Event”. [Tufts-Yale]	R
X	3.1	***Information About Records***	SS
X	3.1.1	***Representation Information***	SSS
4.1.3	3.1.1.1	An Institution MUST maintain representation information for all records types stored. [CTDR B3.4]	R
X	3.1.2	***Unique Identifier***	SSS
2.3.5	3.1.2.1	An Application MUST uniquely identify the records it maintains. [Pitt 6c; MoReq 7.1; PRO A.9.3; DoD c2.2.1.4, c2.2.4.1; PERM 15; Yale A.5; NARA 1.1.2.1, 19.1.14; CTDR B2.5]	R
X	3.1.3	***Copies of Records (Versions)***	SSS
4.1.4	3.1.3.1	An Application MUST manage the relationship between the copies of records components in the system. [Indiana 1.2.8; DoD c2.2.3.18–c2.2.3.20; NARA 15.2.1]	R
4.1.5	3.1.3.2	An Application MUST manage the relationship between all copies of records components to their corresponding records. [NARA 7.4]	R
4.1.6	3.1.3.3	An Application MAY support identification of the authoritative version (master copy or preservation copy) of a record component in the system. [InterPARES A.7; NARA 18.5.1]	R
4.1.7	3.1.3.4	An Application MUST document any changes of a record component from the point of ingest. [InterPARES B.3; NARA 8.1.5]	R
4.1.8	3.1.3.5	An Application MUST document items removed from the preservation system, including filenames, timestamps and a person identifier. [Yale D.3; NARA 15.8.1]	R
X	3.1.4	***Location Tracking***	SSS
4.1.9	3.1.4.1	An Application MUST be able to track the location of its records copies. [MoReq 4.4.1; NARA 10.2.4, 10.2.6; CTDR B2.4–B2.5, D1.3]	R
4.1.10	3.1.4.2	An Application MUST track a record’s unique identifier, current location, time of movements, and the person responsible for the movements. [MoReq 4.4.3; ISO 9.8.3; NARA 10.2.6; CTDR B2.4–B2.5]	R
X	3.1.5	***Demonstration of Controls over Records Transfer, Maintenance, and Reproduction.*** This refers to auditable documentation proving the existence of such controls. This does not include requirements for transfer, maintenance, and reproduction of the records themselves.	SSS
5.5.2	3.1.5.1	An Institution SHOULD demonstrate it has created and maintains a reasonable access criteria and it has successfully implemented the criteria. [InterPARES B.1; ISO 8.3.6; NARA 13.2–13.4; CTDR B3.8]	R
5.2.3	3.1.5.2	An Application SHOULD facilitate the creation, maintenance, and distribution of documentation to support a demonstration of controls over records transfer, maintenance, and reproduction. [InterPARES B.1; NARA 6; CTDR B3.8]	R
5	4	**Administration**	S
5.1	1.1.	***Negotiate Submission Agreement***	SS
5.1.1	1.1.1.		R

5.1.3	1.2.1.2.	An Institution MUST provide for transfer of legal custody of records to the archives. [NARA 1.3]	R
5.1.2	1.2.1.3.	An Application MAY automate the implementation of submission agreements. [NARA 1.6–7]	R
5.2	x	**Manage System Configuration**	SS
5.3	x	**Archival Information Update**	SS
5.4	x	**Physical Access Control**	SS
5.4.1	x	An Institution SHOULD create and maintain policies and procedures to detect, contain, and correct security violations. [HIPAA 45CFR164.308, 45CFR164.310, 45CFR164.312]	R
5.4.2	x	Procedures MUST provide a reasonable guarantee that records are protected from tampering. [Pitt 9a; PRO A.2.15; ISO 7.1.1, 8.2.2.c; HIPAA 45CFR164.306]	R
5.4.3	x	An Institution MUST implement procedures to protect the Archive's facilities and equipment from unauthorized access, tampering, or theft. Such facilities include the physical surroundings of all storage devices and media. [HIPAA 45CFR164.310]	R
5.4.4	x	Natural People MUST be authorized to access the Archives' facilities. [HIPAA 45CFR164.308]	R
5.4.5	x	An Institution SHOULD implement physical safeguards for all workstations that access electronic records, to restrict access to authorized users. [HIPAA 45CFR164.310]	R
5.4.6	x	An Instituion MUST NOT dispose of records storage media or make it available for re-use without assuring electronic records are removed. [HIPAA 45CFR164.310]	R
5.5	4.1	**Establish and Maintain Standards and Policies**	SS
5.5.3	6.2.6	Procedures SHOULD exist to redact restricted content from records. [Pitt 13, MoReq 9.3.10; PRO A.2.56; NARA 20.11.2]	R
5.5.15	6.2.6	An Institution SHOULD establish policies regarding rendering the functionality of record types. [Pitt 11b, DoD c2.2.5.3; NARA 8.1.6.5, 20.11.3]	R
5.5.16	6.2.3	An Institution SHOULD establish policies defining the necessary elements of a response to a Consumer request (what is an appropriate response). [CTDR B5.3]	R
5.5.17	6.2.1	An Institution SHOULD establish policies defining the description necessary to ensure records are discoverable. [Indiana 1.10.2-3; MoReq 8.1.4-5, 8.1.7; PRO A.3.4, A.3.6, A.3.8, A.3.17; PERM 18; NARA 19]	R
5.5.18	5.3	Procedures SHOULD exist for managing record types with templates. [NARA 7.2]	R
5.5.19	4.4.5	Procedures SHOULD exist for monitoring the available storage. [Tufts-Yale]	R
5.5.20	x	An Institution SHOULD establish format transformation policies and plans to implant them. [Tufts-Yale]	R
X	4.2	**Protection from Loss or Corruption**	SS
5.6	x	**Audit Submission**	SS
X	4.2.1	**Disaster Preparation**	SSS
5.5.4	4.2.1.1	An Institution SHOULD create backup and failure mode procedures for its records and metadata associated with those records. [Indiana 1.9, 1.9.4; Pitt 2d; MoReq 4.3.7; InterPARES A.3; ISO 8.3.3; NARA 10.2.3, 14.9; CTDR D1.2, D3.4]	R
3.2.4	4.2.1.2	An Institution SHOULD test and review backup and failure mode procedures. [HIPAA 45CFR164.308, 45CFR164.310; NARA 27.1; CTDR D1.2, D3.5]	R
3.5.5	4.2.1.3	Procedures SHOULD provide for the automated backup of the preserved records and preservation metadata. [MoReq 4.3, 4.3.1, 9.1.2-3; PRO A.9.11, A.9.17; Dod c2.2.9.1]	R
3.5.6	4.2.1.4	Procedures SHOULD articulate the actions needed to be undertaken during primary system failure. [Pitt 2d; MoReq 4.3.5; HIPAA 45CFR164.308]	R

x	4.2.2	***Access Control***	SSS
5.5.5	4.2.2.1	An Institution MUST explicitly assign responsibility for the annotation, relocation, and destruction of records on the basis of a person's authority and capacity to carry out the activity (establishing user security profiles). [Indiana 1.7.2, IP A.2, MoReq 4.6.5, 9.3.5; PRO A.5.36; ISO 9.7; PERM 25; HIPAA 45CFR 164.308, 45CFR164.312; Yale A.5; NARA 13]	R
1.3.3	4.2.2.2	An Application MUST confer exclusive capabilities upon people to exercise the responsibility for annotation, relocation, and destruction of records as defined by an institution. [Indiana 1.4.1; DoD c2.2.5.2, c2.2.7.4; ISO 8.3.6; HIPAA 45CFR164.308; IP A.2; NARA 13]	R
1.3.4	4.2.2.3	Procedures MUST prescribe periodic software security updates. [HIPAA 45CFR164.308]	R
5.5.6	4.2.2.3		R
4.1.11	4.2.2.3		R
4.3.1	4.2.2.4	An Application MUST NOT allow unauthorized changes to the records it maintains. [Indiana 1.7.1; MoReq 3.2.1, 4.5.4, 6.1.4; PRO A.2.41; DoD c2.2.5.4; ISO 9.7.d; Yale A.4, A.7; NARA 13]	R
5.5.7	4.2.2.4		R
1.3.8	4.2.2.4		R
5.6.4	4.2.2.5	An Institution SHOULD create and maintain policies and procedures to detect, contain, and correct security violations. [HIPAA 45CFR164.308, 45CFR164.310, 45CFR164.312; NARA 13]	R
5.5.8	4.2.2.6	An Institution SHOULD perform a periodic review of its security procedures (including reanalysis of security threats or access management system failure). [InterPARES B.1.b; HIPAA 45CFR164.308; CTDR B5.2]	R
5.2.4	4.2.2.6		R
x	4.2.2.7	Procedures SHOULD allow for the periodic review of access control rules, records security profiles, and user security profiles. [MoReq 4.6.12; PRO A.5.40; ISO 9.7; HIPAA 45CFR164.308]	R
x	4.2.2.8	Procedures SHOULD allow for the modification of access control rules, records security profiles, and user security profiles based on the findings of a review. [HIPAA 45CFR164.308; NARA 8.9]	R
x	4.3	***Reporting Capability and Event Log***	SS
5.2.5	4.3.1	An Application MUST be able to identify system failures [NARA 27.2.1; CTDR B5.2]	R
x	4.3.2	An Application SHOULD be able to produce reports for administrators to document any system activity, including failure. [MoReq 3.4.14, NARA 26.1, 26.3.1, 27.2.4; CTDR B5.2]	R
4.3.3	4.3.3	An Application MUST provide the capability to produce documentation of any reproduction or copy process and its effects, including the dates of the records' reproduction and the name of the responsible person and the impact of the reproduction process on the form of the records components (any changes the records components have undergone). [IP B.2]	R
5.2.6	4.3.4	An Application MAY provide the facility to isolate and resolve failures, provided any activity is documented and any changes to affected records components checked and documented. [NARA 27.2.2–27.2.3]	R
x	4.4	***System Administration***	SS
x	4.4.1	An Application MUST function on well-supported operating systems and other core infrastructure software. [CTDR D1.1]	R
5.2.7	4.4.2	Procedures SHOULD contain provisions for all routine maintenance tasks which fall in line with industry best practices. [Pitt 2c; CTG System; NARA 27; CTDR D1.10]	R
5.2.16	4.4.3	An Application MUST support a stasis mode where no changes are allowed. [Tufts-Yale]	R
2.5.2	4.4.3	An Application MUST support a stasis mode where no changes to records are allowed.	R

5.2.8	4.4.4	An Application MUST allow convenient access to and the ability to modify any configuration parameters. [MoReq 11.2.7, 9.1.1; NARA 27.4]	R
5.2.14	4.4.5	Infrastructure SHOULD provide the ability to monitor available storage capacity. [MoReq 9.14; PRO A.9.21; NARA 27.3.4]	R
5.2.14	4.4.6	An Institution SHOULD determine the maximum number of simultaneous users necessary. [MoReq 11.3; DoD c3.1.3]	R
5.2.9	4.4.7	An Institution SHOULD identify the necessary hours of Application availability. [MoReq 11.3; DoD c3.1.3]	R
5.2.10	4.4.8	Infrastructure SHOULD be capable of fulfilling downtime and simultaneous user requirements laid out by the institution. [MoReq 11.3]	R
5.2.11	4.4.9	An Application MAY provide the capability to monitor overall system state in a consolidated manner. [NARA 27.3]	R
6	5	**Preservation Planning**	S
6.1	x	**Monitor Designated Community**	SS
6.1.3	x	An Institution MUST periodically monitor the acceptability of chosen preservation strategies to existing Consumers and Producers. [Tufts-Yale]	R
6.2	x	**Monitor Technology**	SS
x	5.1	**Preservation Planning framework**	SS
6.3	5.1.1	**Develop Preservation Strategies & Standards**	SSS
6.3.1	x	Juridical People MUST synthesize information about designated communities, technologies, system performance, inventory, and finances in order to recommend preservation strategies and standards. [Tufts-Yale]	
6.3.2	5.1.1.1	An Institution MUST establish plans for preserving records as long as needed and have a written mission statement that reflects a commitment to long-term preservation. [Indiana 1.9; MoReq 11.7.4; PERM non dod 4; NARA 8.9; CTDR 3.1]	R
6.3.3	5.1.1.2	An Institution SHOULD establish strategies for ensuring the accessibility and functionality of records components over time. [InterPARES A.4; ISO 8.3.5, 9.6]	R
6.3.4	5.1.1.3	An Institution SHOULD establish preservation action plans specifying preservation actions to be taken in ensuring the accessibility and functionality of templates of records components over time. [InterPARES A.4; ISO 8.3.5, 9.6; NARA 8.9]	R
6.3.5	5.1.1.4	An Institution SHOULD establish plans for managing preservation metadata and attaching it to records. [MoReq 5.3.10, 11.7.7; PERM 5, 6]	R
5.5.9	5.1.1.5	An Application MAY provide the capability for users to create and maintain preservation and access plans (including the ability to alter plans) [NARA 8.9.1–8.9.6; CTDR B3.10]	R
5.5.10	5.1.1.6	An Application MAY provide the capability for users to associate a preservation and access plan with electronic records [NARA 8.9.5]	R
6.3.6	6.2.6	An Institution SHOULD develop strategies for redaction of restricted content from users. [Pitt 13, MoReq 9.3.10; PRO A.2.56; NARA 20.11.2]	
6.3.7	6.2.5	An Institution SHOULD develop strategies for rendering the functionality of types of records. [Pitt 11b, DoD c2.2.5.3; NARA 8.1.6.5, 20.11.3]	
6.3.8	5.3	An Institution SHOULD develop templates to manage record types. [NARA 7.2]	
x	5.2	**Monitor preservation strategies standards and best practices**	SS
6.2.1	5.2.1	Juridical People SHOULD monitor the state of the art of information technology in order to facilitate preservation planning.	R
5.2.13	5.2.2	5.2.13. An Institution MUST actively monitor the integrity of AIPs. [CTDRB3.7]	

2.4.7	5.2.2	An Institution SHOULD use representation information from appropriate international registries.[CTDR B3.3]	R
x	5.3	***Manage record types.*** This subsection describes the management of sets of specifications about records. Every set of record stored in the system should conform to a type registered for that type of set of records, either at the time it is ingested into the system or through a subsequent transformation.	SS
6.4.1	5.3.1	An Institution SHOULD define records templates to automate preservation planning and processing. [NARA 7]	R
6.1.1	5.3.2		
6.2.2	5.3.2	An Application SHOULD enable monitoring and notification when preservation strategies are no longer viable. [CTDR B3.9]	R
6.2.3	5.3.3	An Institution MUST periodically monitor the viability of chosen preservation strategies. [CTDR B3.9]	R
4.1.12	5.3.4	An Application MAY provide for management of templates within a template repository. [NARA 7.2]	R
4.1.13	5.3.5	An Application MAY provide the capability to associate templates with sets of records. [NARA 7.7.1–7.7.2]	R
7	6	**Access**	S
7.1	x	**Coordinate Access Activities**	
x	6.1	***Use Rights*** This subsection covers the institution's management of users' rights to view and/or receive records. This includes the development, management, and review of records and user security profiles. It also includes the management of access controls and authentication of users.	SS
x	6.1.1	***Access Controls*** This subsection covers the management of processes that control the access of records in a preservation system.	SSS
7.1.1	6.1.1.1	Procedures MUST ensure that only authorized users gain access to records. [MoReq 4.1.1; PRO A.5.25, A.5.42, A.5.46-50; NARA 13]	R
x	6.1.2	***Record Security Profile***	SSS
2.1.7	6.1.2.1		
4.1.14	6.1.2.1	An Application MUST allow records security profiles to be created and modified. [MoReq 9.3.5; PRO A.5.36; NARA 8.9.5, 16.6.2]	R
2.1.8	6.1.2.2		
4.1.15	6.1.2.2	An Application MUST allow record security profiles to be assigned to records. [MoReq 4.6.1; PRO A.2.26, A.5.5, A.5.26, A.5.27; ISO 9.7.2; NARA 8.9.5]	R
4.1.16	6.1.2.3	An Application SHOULD allow time sensitive records profiles that are valid for a limited time period to be assigned to records and should automatically be switched to another records security profile when their valid time period expires. [PRO A.5.38-39; NARA 13.13.2]	R
x	6.1.3	***User Security Profile***	SSS
4.1.17	6.1.3.1	An Application MUST allow user security profiles to be created and modified. [MoReq 9.1.8; PRO A.5.11, A.5.17-18, A.5.20-22]	R
4.1.18	X	An Application MUST allow user security profiles to be linked to natural people. [MoReq 4.1.2, 4.1.5, 4.6.7, 9.1.7; PRO A.5.5, A.5.10, A.5.13, A.5.16, A.5.24; DoD c2.2.7.3]	
4.1.19	X	A Juridical Person SHOULD review records before a time-sensative change is made in the records security profile. [Tufts-Yale]	
7.1.2	6.1.3.2	An Application MUST assign or reassign user security profiles to people. [MoReq 4.1.2, 4.1.5, 4.6.7, 9.1.7; PRO A.5.5, A.5.10, A.5.13, A.5.16, A.5.24; DoD c2.2.7.3]	R
x	6.1.4	***Authentication of Users***	SSS

1.3.7	6.1.4.1	Infrastructure SHOULD provide services for secure authentication. [PRO A.5, A.5.1, A.5.11; DoD c2.2.7.1; HIPAA 45CFR164.312]	R
1.3.8	6.1.4.2	An Application MUST authenticate users before providing services. [PRO A.5, A.5.1, A.5.11; DoD c2.2.7.1; HIPAA 45CFR164.312; Yale A.4]	R
x	6.1.5	***Access Review***	SSS
x	6.2	***Discovery and Delivery***	SS
x	6.2.1	***Searching***	SSS
7.1.3	6.2.1.1	An Application MUST ensure all of its records and metadata are discoverable. [Indiana 1.10.2-3; MoReq 8.1.4-5, 8.1.7; PRO A.3.4, A.3.6, A.3.8, A.3.17; PERM 18; NARA 19]	R
7.1.4	6.2.1.2	An Application MUST be able to render all records returned in a search results list. [MoReq 8.2.1; PRO A.3.20; DoD c2.2.6.8.10; NARA 19.8]	R
7.1.5	6.2.1.3	An Application MUST support searching by records' identifiers. [MoReq 8.1.16, 8.1.23; NARA 19.1.14]	R
7.1.6	6.2.1.4	An Application SHOULD provide an integrated search interface. [MoReq 8.1.2; PRO A.3.7; NARA 19.1, 21]	R
7.3.1	6.2.1.5	An Application MUST, if it has an integrated search interface, present search results. [PRO A.3.15; DoD c2.2.6.8.5; NARA 19.8]	R
7.1.7	6.2.1.6	An Application MAY support resource discovery through external interfaces/mechanisms in addition to any integrated search interface. [PRO A.3.19]	R
7.1.8	6.2.1.7	An Application SHOULD limit search results to the records the user has rights to access. [MoReq 4.1.10, 4.1.12, 8.1.28; PRO A.3.18, A.5.51-52, B.3.18; NARA 19.8.3]	R
7.3.2	6.2.1.8	An Application MAY provide capabilities to manage a search results list including, but not limited to, order, number of hits per page, filter results files, and saving search results. [MoReq 8.1.17; 8.1.24-25; DoD c2.2.6.8.5; NARA 19.8.4–19.12.3]	R
x	6.2.2	***Query Techniques***	SSS
7.1.9	6.2.2.1	An Application SHOULD support the full text search of the records and metadata it maintains. [MoReq 8.1.8; DoD c3.2.9; NARA 19.1.5–19.1.19	R
7.1.10	6.2.2.2	An Application SHOULD support searching metadata fields containing controlled vocabulary terms managed by thesauri. [MoReq 8.1.10; PRO A.3.5; DoD c3.2.9; NARA 19.1.3]	R
7.1.11	6.2.2.3	An Application SHOULD support searching multiple metadata fields and/or full text of records. [MoReq 8.1.6; PRO A.3.9; DoD c2.2.6.8.2; NARA 19]	R
7.1.12	6.2.2.4	An Application SHOULD support the use of Boolean and/or relational search operators such as "and" "or" "not" "less than" "greater than" "equal to." [MoReq 9.1.8; PRO A.3.13; DoD c2.2.6.8.4; NARA 19.1.19]	R
7.1.13	6.2.2.5	An Application SHOULD support wild card and/or pattern matching searches. [MoReq 8.1.11; PRO A.3.13; DoD c2.2.6.8.3; NARA 19.1.24]	R
7.1.14	6.2.2.6	An Application SHOULD support the iterative refinement of a search by adding search conditions to a previously run search—i.e. narrow a search. [MoReq 8.1.21; NARA 19.9]	R
7.1.15	6.2.2.7	An Application MAY support word proximity searching. [MoReq 8.1.12; NARA 19.1.20]	R
7.1.16	6.2.2.8	An Application MAY support searching null values. [DoD c2.2.6.8.6]	R
7.1.17	6.2.2.9	An Application MAY support searching time intervals. [MoReq 8.1.22]	R
x	6.2.3	***Responding to Requests***	SSS

7.3.3	6.2.3.1	An Institution MUST answer a consumer request with an appropriate response (a DIP fulfilling the entire request, a response denying the request, or a DIP fulfilling part of the request accompanied by a response clarifying why the request is only partially fulfilled). [CTDR B5.3]	R
6.1.2	6.2.3.2		
7.1.18	6.2.3.2	An Institution SHOULD document that consumer requests are responded to. [CTDR B5.5]	R
7.3.4	6.2.3.3	An Institution MUST disseminate DIPs that are authentic copies of their corresponding SIPs. [CTDR B5.6]	R
x	6.2.4	***Rendering Complex Objects***	SSS
7.2	x	**Generate DIP**	
7.2.1	6.2.4.1	An Application MUST render all of the components of a record along with their associated metadata in a logical manner. [Indiana 1.10.4; MoReq 8.1.15, 8.2.3; PRO A.3.21–A3.24; DoD c2.2.3.21; PERM 23; NARA 20.9, 20.11]	R
7.2.2	6.2.4.2	An Application MUST be able to render records on to appropriate output media, which should at least include graphical display and printer output. [MoReq 8.2, 8.3, 8.4.1; Pro A.3.25-26, A.3.28-29; PERM 3, 10, 14, 16, 17, 24, non dod 2; NARA 26.3.3, 26.4.1]	R
7.2.3	6.2.4.3	An Application SHOULD be able to render records into an open export format. [PRO A.3.31; NARA 26.4.3]	R
7.2.4	6.2.4.4	An Application SHOULD be able to render records independently of their creating environments. [MoReq 8.2.2; PRO A.3.22; DoD c3.2.14]	R
7.2.5	6.2.4.5	An Application SHOULD be able to render a record simultaneously for multiple users. [PRO A.3.23, DoD c2.2.7.5]	R
7.2.6	6.2.4.6	An Application SHOULD be able to render all versions of a record. [DoD c2.2.6.8.9]	R
x	6.2.5	***Rendering Recordness***	SSS
7.3.5	6.2.5.1	An Application MUST render a record's content. [Pitt 11, 12; MoReq 8.2.3; PRO A.3.21; PERM 2; NARA 8.1.6.3, 20.11.1]	R
7.3.6	6.2.5.2	An Application MUST render a record's structure. [Pitt 12, 12b, 12b1; PRO A.3.21; PERM 2; NARA 8.1.6.6, 20.11.4]	R
7.3.7	6.2.5.3	An Application MUST render a record's context. [Pitt 12, 12b1, 12c; ISO 7.25; PERM 2]	R
7.3.8	6.2.5.4	An Application MUST render a record's functionality. [Pitt 11b, DoD c2.2.5.3; NARA 8.1.6.5, 20.11.3]	R
x	6.2.6	***Redaction***	SSS
7.2.7	6.2.6.1		
7.3.9	6.2.6.1	Procedures SHOULD provide for the redaction of restricted content from records delivered to users that do not have the right to see the restricted output. [Pitt 13, MoReq 9.3.10; PRO A.2.56; NARA 20.11.2]	R
7.3.10	6.2.6.2	An Application SHOULD be able to create redacted versions of textual, audio, and moving image records. [MoReq 9.3.10; NARA 18]	R
7.3.11	6.2.6.3	An Application MUST NOT, if it can redact records, alter the content of a record while creating a redacted version of that record. [Pitt 13a; PRO A.2.56; NARA 18]	R

Fedora and the Preservation of University Records Project

2.1 Ingest Guide

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/

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:UA069.004.001.00006
(September 2006)
and
http://dca.tufts.edu/features/nhprc/reports/ingest/index.html
(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

TABLE OF CONTENTS

OVERVIEW

One of the key challenges to preserving electronic records in a meaningful way is preserving the authenticity and integrity of records during their movement from a recordkeeping system to a preservation system. This Ingest Guide describes the actions needed for a trustworthy ingest process. This process enables an Archive and Producer to move records from a recordkeeping system to a preservation system in a manner that allows a presumption of authenticity.

This Ingest Guide refers to ingest broadly, defining it as the entire process involved in moving records from a recordkeeping system to a preservation system. This process consists of the Producer and Archive agreeing to and defining what records will be transferred and the manner of the transfer, validation, and transformation. Following the Guide should help an Archive and Producer ensure the functional, not just byte-stream preservation, of records. Not only does the Guide articulate steps for ensuring that records are properly tracked and have maintained their structural integrity during ingest, it also provides a way for the Archive to ensure that records remain renderable, functional, and meaningful. Following the Guide should enable an Archive to have a trustworthy ingest process, which would allow a reasonable person to presume that a record has maintained its level of authenticity during ingest.

This guide does not describe the functional or technical requirements for building either a recordkeeping or a preservation system. Instead, this guide presents a detailed description of the complex ingest workflow step by step. For more on authenticity and trustworthy recordkeeping systems and the preservation of records see "Requirements for Trustworthy Recordkeeping Systems and the Preservation of Electronic Records in a University Setting."[1]

The Ingest Guide contains two main sections. Section A, Negotiate Submission Agreement, details how the Producer and the Archive create and arrange a Submission Agreement that defines the terms and conditions of the transfer of records from the Producer to the Archive, and it details the scope of the records along with the nature of their validation and transformation. Section B, Transfer and Validation, details the actual transfer, validation, and transformation of records. Section A contains eleven parts and Section B has six parts. Each part in Section A and B is composed of a number of steps.

Each part includes a narrative summary, a flowchart illustrating all of its steps, and a description of each step. Each description includes an Overview, a list of Components, Resources, Products, and Documentation that each step utilizes and/or produces, and a thumbnail flowchart.

The Ingest Guide also includes a separate section on Components, Resources, Products, and Documentation that describes each of these roles in the ingest process and refers to the steps that use and produce them. The Ingest Guide also has a Submission Agreement section that explains

[1] Fedora and the Preservation of University Records (NHPRC 2004-083), "Requirements for Trustworthy Recordkeeping Systems and the Preservation of Electronic Records in a University Setting," <http://dca.tufts.edu/features/nhprc/reports/1_5final.pdf>.

the Agreement in further detail. Finally, the Guide includes crosswalks between its own steps of the *Producer-Archive Interface Methodology Abstract Standard*.[2]

Although the Ingest Guide is a prescriptive guide for a trustworthy ingest process, it is not a detailed manual of procedures. The implementation of the Guide can produce a wide variety of procedures and policies from archive to archive. The Guide describes the actions that must be undertaken to trust the ingest process and prescribes how to undertake these steps at a high level, but it does not prescribe how to proceed in full detail. For example, Step A5.5 calls for the Archive to choose a preservation format for records it chooses to transform, but it does not dictate what those preservation formats should be. An Archive following the Ingest Guide will still have to determine what preservation formats best serve its needs. The Guide points out many tasks that Archives must undertake to have a trustworthy ingest processes, without discussing those tasks in detail. The most prominent of these tasks include the details of the appraisal process in Part A3, the creation of Submission Information Packages in Part B1, and the creation of Resources. For more on the implementation of the Ingest Guide, see Appendix A: Using the Ingest Guide.

The Guide uses the *Open Archival Information System Reference Model (OAIS)* definition of Archive: "An organization that intends to preserve information for access and use."[3] Therefore, while an Archive may be an archives in the sense used by the archival community, it does not necessarily have to be an archives.[4] In the context of the Ingest Guide, an Archive is any type of office or juridical body that has the responsibility of providing long-term preservation and access to records. Like *OAIS*, the Guide refers to a single ingest or a single set of recurring ingests as an Ingest Project.

The Ingest Guide also uses the *OAIS* definition of Producer: "The role played by those persons, or client systems, who provide the information to be preserved. This can include other OAISs or internal OAIS persons or systems."[5] This means that Producer will normally be the custodian of the records—or an entity the Producer has authorized to act on its behalf—that has the authority to transfer the records to the Archive. The Producer may or may not be the individual, group, or organization that is responsible for the creation, production, accumulation, or formation of the records it transfers to the Archive.

The Ingest Guide is based upon the work of the Consultative Committee for Space Data Systems (CCSDS) and builds upon its OAIS framework. In particular, Section A of the Ingest Guide is based on the CCSDS's *Producer-Archive Interface Methodology Abstract Standard*, which is composed of four phases: Preliminary, Formal, Transfer, and Validation. The *Producer-Archive Interface* is a follow-up document to *OAIS*. The Preliminary and Formal phases greatly expand on the Negotiate Submission Agreement activity in the Administration function of *OAIS*. The

[2] 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>.
[3] ISO 14721:2003, Space data and information transfer systems -- Open Archival Information System -- Reference model.
[4] For the "archival community" definition of the term archives see Richard Pearce-Moses, *A Glossary of Archival and Records Terminology* (Archival Fundamentals Series II), (Chicago: Society of American Archivists, 2005).
[5] ISO 14721:2003, p. 1–12.

Transfer and Validation phases reiterate the Receive Submission and Quality Assurance activities respectively, both of which are in the Ingest function of *OAIS*.

As a product of the Fedora and the Preservation of University Records grant project (NHPRC 2004-083), the Ingest Guide is designed primarily for a university setting. However, its general nature may also make it applicable in other environments. The university orientation of the Ingest Guide differs from the *Producer-Archive Interface*'s orientation. While the PAI treats the creation of a Submission Agreement formally and in two different phases ("Preliminary" and "Formal"), the Ingest Guide does not distinguish preliminary and formal phases, because this level of formality is unnecessary. Such formality would impose unrealistic implementation expectations in a university setting.

The Ingest Guide is based largely on the conceptual underpinnings of the records lifecycle model, presuming that a Producer will create, acquire, utilize, and manage records in a Recordkeeping System to suit its current business needs, and later the Archive will ingest some of those records into a separate Preservation System that the Archive administers. In this model, the Archive acts as a neutral third party to the recordkeeping process acting on behalf of broader societal needs rather than on behalf of the Producer. As a neutral third party the Archive has no stake in the content of the records and no reasons to alter records under its custody, and it should not allow anybody to alter the records either accidentally or on purpose. Many archivists have rejected the lifecycle model in favor of the records continuum concept, where recordkeeping is seen as a continuous process that is not time-based, separated into a series of clearly defined steps, or administered by completely separate juridical entities. Many Producers and Archives operate in a mixed world between these two models. For example, many Archives operate separately from a Producer but are part of same organization as the Producer and do not act as a neutral third party. The Ingest Guide should be useful to most Archives operating in a mixed lifecycle/continuum environment, particularly ones where separate Recordkeeping and Preservation Applications are maintained.[6]

The Ingest Guide assumes that a Producer is submitting managed records to an Archive. Traditionally, an archives might accept boxes of unorganized paper records from a faculty member, for example, with the idea that the archives could add these records to its processing backlog and later impose some sort of order, or arrangement, long after the transfer. It is the assumption of this Guide, and the corresponding preservation requirements, that such delayed arrangement of electronic records is neither scalable nor sustainable. A box of unlabeled disks sent from a faculty member illustrates this point. The Ingest Guide places the activity of imposing order on electronic records outside of the ingest and preservation activities. The work of preparing organized and managed records for transfer to the Archive is the Producer's responsibility and, in the case of the box of unorganized diskettes, the Archive is doing the Producer's job, imposing order on the records after the fact. In a situation like this, during part A of the Ingest Guide, the Archive would either require the Producer to organize the disks and the records they hold before the transfer takes place, or artificially organize the records after accepting them from the Producer but still before the transfer to the preservation repository. By

[6] The authors do not wish to express any opinion of the relative merit of either the lifecycle or continuum models, but instead simply disclose their inherent bias towards the lifecycle model based on educational background and work experience.

imposing this artificial arrangement the Archive has become a Producer and thus plays both roles in the Ingest process.

SECTION A: NEGOTIATE SUBMISSION AGREEMENT

Overview

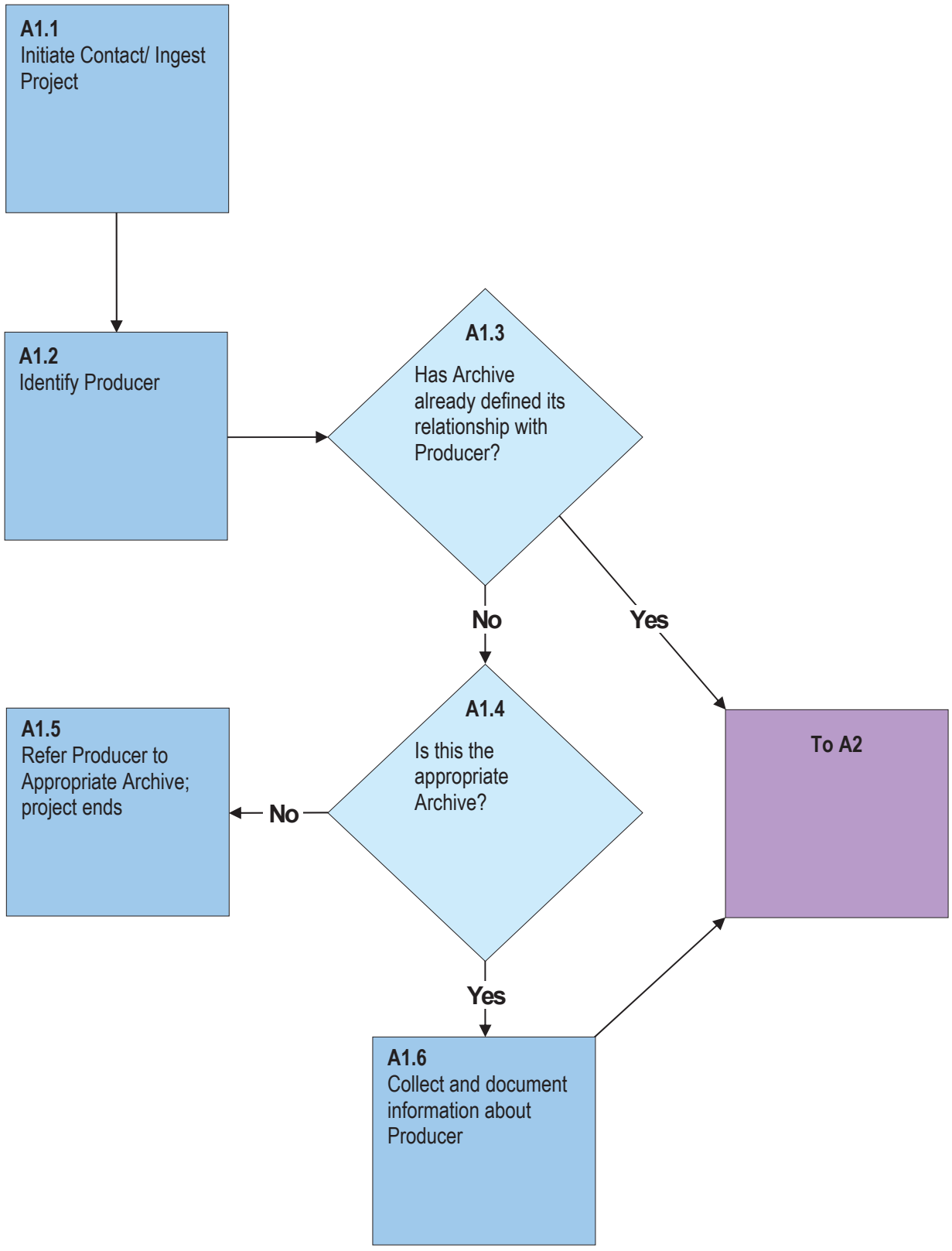
The Negotiate Submission Agreement Section of the Ingest Guide describes the actions needed for an Archive and a Producer to generate a Submission Agreement. These agreements define the nature and scope of the records to transfer to the Preservation System and how the Archive will execute transfer, validation, and transformation of these records. All of the work in Section A is undertaken for the production of a Submission Agreement. Actual transfer, validation, and transformation work only occurs in Section B of the Ingest Guide.

This section is composed of eleven parts. During first three parts, Establish Relationships, Define Project, and Collect Information and Assess Value of Records, the Archive conducts an intellectual appraisal of the records under consideration and determines if it should accession them into the Preservation System. During these three parts, the Archive also gathers a variety of information about the records it should accession. This information includes the creator, record type, format, date, and extent of the records and any identifiers associated with them. During the next seven parts, Assess Record Types, Assess Formats, Assess Identifier Rules, Assess Copyright, Assess Access Rights, Assess Recordkeeping System, and Assess Feasibility, the Archive determines if its existing resources for preservation formats, record types, identifier rules, creator records, security procedures, transfer procedures, and system capabilities—referred to as Resources and described in the Components, Resources, Products, and Documentation section of this guide—meets the needs of the records identified in Part A3. The resulting feasibility report should present a gap analysis if the Resources do not reflect the continuing value of any records assessed in A3. The Archive must then determine if it should modify or add to its Resources to meet those assessments or if it should reject or modify the scope of the records involved in the Ingest Project. In Part A11, Finalize Submission Agreement, the Archive and the Producer finalize and agree to a Submission Agreement that is based on the scope of the records defined in Parts A1 through A3 and the decisions made in Parts A4 through A10.

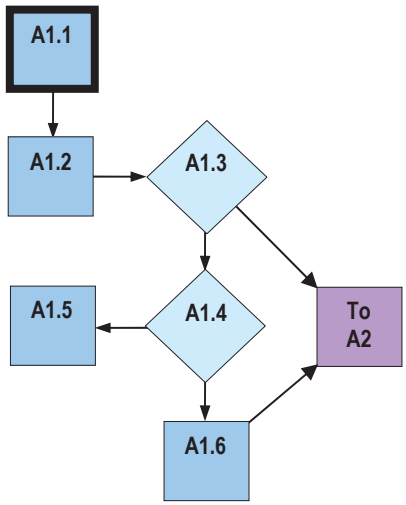
For more information on Submission Agreements, see the Submission Agreement section of this Guide.

SECTION A: NEGOTIATE SUBMISSION AGREEMENT
Part A1: Establish Relationship

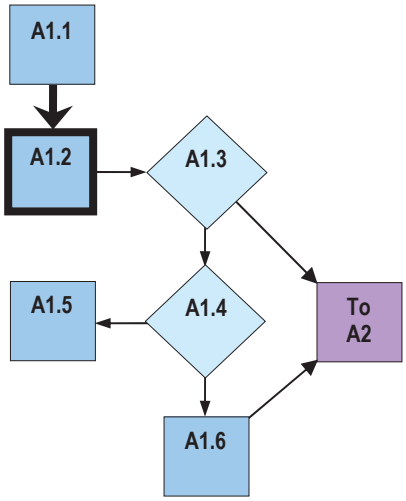
Overview
During this Part, either the Archive or the Producer will initiate contact with the other. If the Archive does not already have a relationship with the Producer, the Archive will define its administrative, legal, and/or collecting relationship with the Producer and generate metadata about the Producer. Not every relationship that is established will result in a negotiated Submission Agreement, or even a project definition (Part A2). It is also possible that there will be a time delay between establishing a relationship and defining a potential accession. The Archive should document contacts with Producers to mange potential accession opportunities. How the Archive manages this documentation will depend on the size of the Archive, the number of relationships, the frequency of contact, and the length of the delays.


A1.1
Initiate Contact/ Ingest Project
A1.2
Identify Producer
A1.3
Has Archive already defined its relationship with Producer?
No
Yes
A1.4
Is this the appropriate Archive?
No
A1.5
Refer Producer to Appropriate Archive; project ends
Yes
To A2
A1.6
Collect and document information about Producer

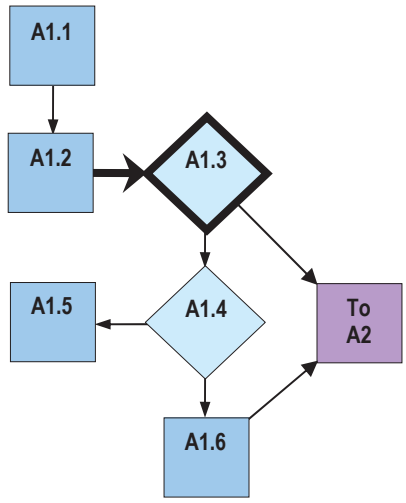
A1.1
Description Either the Archive or the Producer will initiate contact with the other. This is the first step of an Ingest Project. Generally this contact is made informally, although it can be made formally if necessary. Document contact in appropriate Activity Log.
Uses None
Produces/Modifies Activity Log



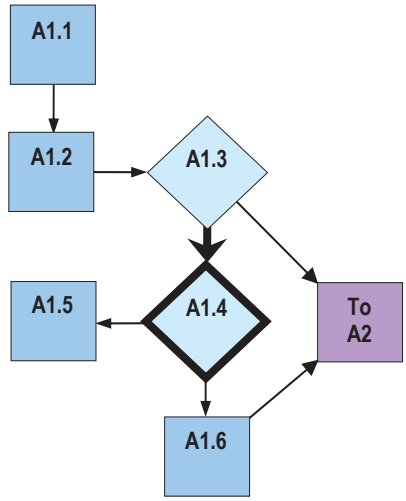
A1.2
Description The Archive determines who the Producer is, in particular determining who he/she/it is as a juridical body.
Uses Institutional Identity Management System
Produces/Modifies None



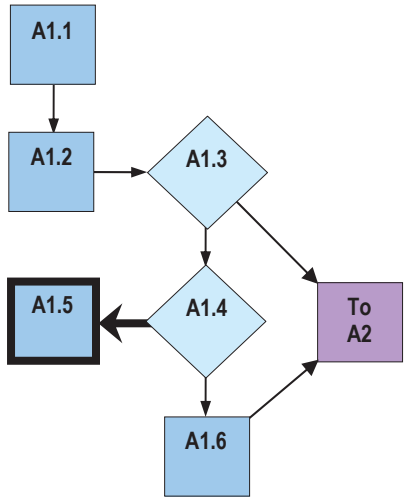
A1.3
Description Based on identifying who or what the Producer is in Step A1.2, the Archive determines if it has already established a Producer-Archive relationship with the Producer in which it can serve as the Archive for at least some of the Producer's records. If this is the case, the Archive documents who the Producer is in the Submission Agreement and goes on to Part A2 of the Ingest Guide. The Archive makes this determination by reviewing Producer Records, Accession Logs; or Activity Logs that describe Producers, their relationships to the Archive, and document its past interactions between the Archive and Producers.
Uses Accession Log, Activity Log, Producer Record
Produces/Modifies Producer Entry



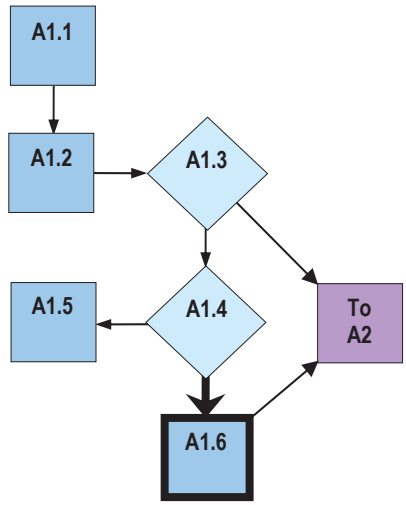
A1.4
Description If the Archive has not already established a relationship with the Producer as the Archive for at least some of its records, the Archive needs to determine if in fact it has the authority to serve as the Archive for at least some of the Producer's records.
Uses Records Authority Statement, Collection Policy
Produces/Modifies None



A1.5
Description If the Archive does not have the authority to serve as the Archive for at least some of the Producer's records, it should recommend the Producer to an appropriate Archive. The Ingest Project ends.
Uses Archives Directory
Produces/Modifies Ingest Project Termination Notice

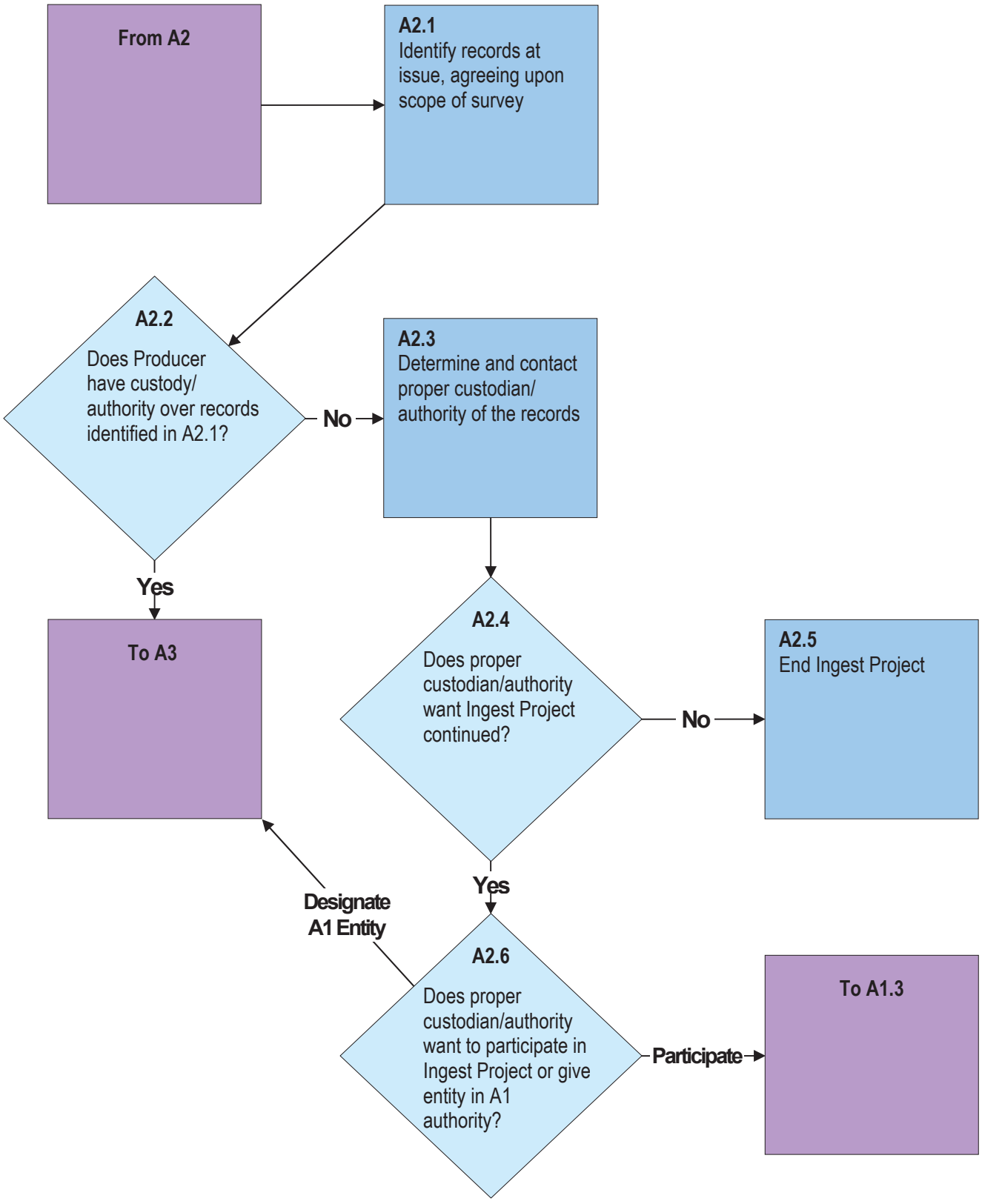


A1.6
Description If the Archive does have the authority to serve as the Archive for at least some of the Producer's records, it needs to collect information about the Producer and create a Producer Record and document who the Producer is in the Submission Agreement.
Uses Institutional Identity Management System
Produces/Modifies Producer Record, Producer Entry

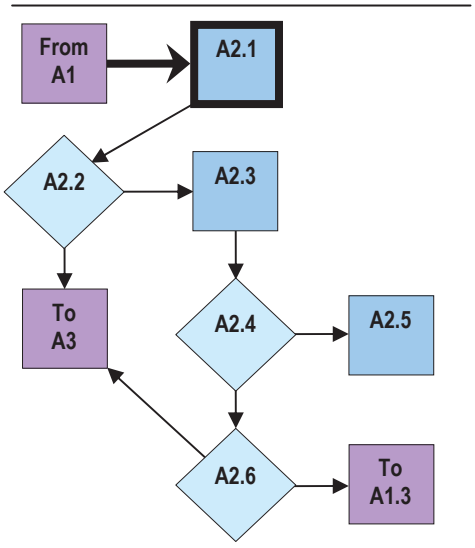


SECTION A: NEGOTIATE SUBMISSION AGREEMENT
Part A2: Define Project

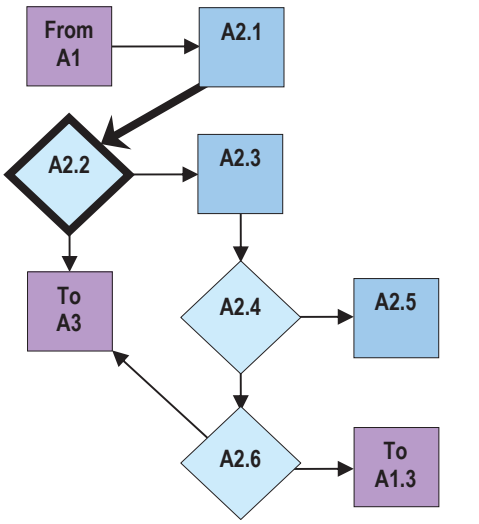
Overview
During this Part, the Archive and the Producer come to an agreement regarding which records the Archive will consider for accession, essentially defining the scope of the Ingest Project. The Archive then verifies that the Producer has proper custody of the records under consideration.


From A2
A2.1
Identify records at issue, agreeing upon scope of survey
A2.2
Does Producer have custody/ authority over records identified in A2.1?
No
A2.3
Determine and contact proper custodian/ authority of the records
Yes
To A3
A2.4
Does proper custodian/authority want Ingest Project continued?
No
A2.5
End Ingest Project
Yes
Designate A1 Entity
A2.6
Does proper custodian/authority want to participate in Ingest Project or give entity in A1 authority?
Participate
To A1.3

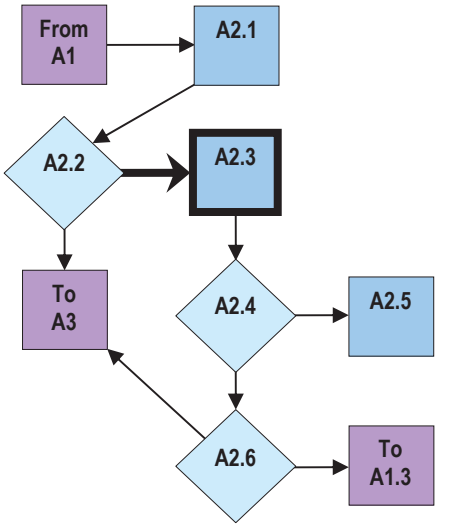
A2.1
Description The Archive and the Producer agree upon the scope of the records that the Archive will survey. The Archive should document this scope of records in what will ultimately become a Survey Report.
Uses None
Produces/Modifies Survey Report



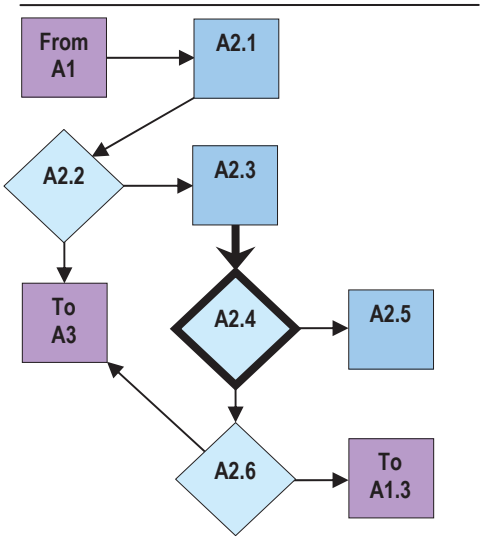
A2.2
Description The Archive determines if the Producer actually has custody or authority over the records the Archive agreed to survey in Step A2.1.
Uses Producer Record
Produces/Modifies None



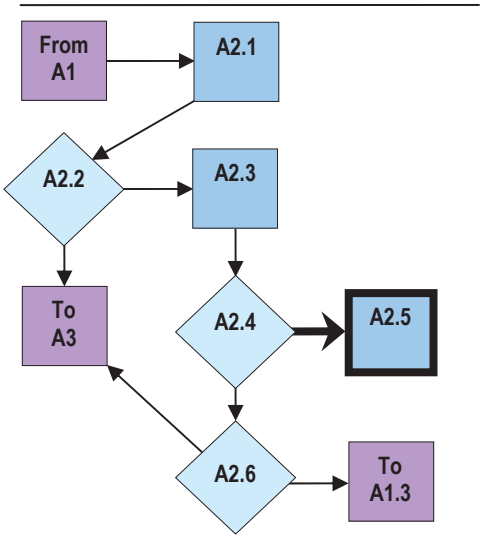
A2.3
Description If the Producer that approached the Archive in Step A1.1 does not have custody or authority over the records, the Archive determines what Producer has proper authority or custody over the records and then contacts that Producer.
Uses Institutional Identity Management System, Producer Record, Activity Logs
Produces/Modifies None



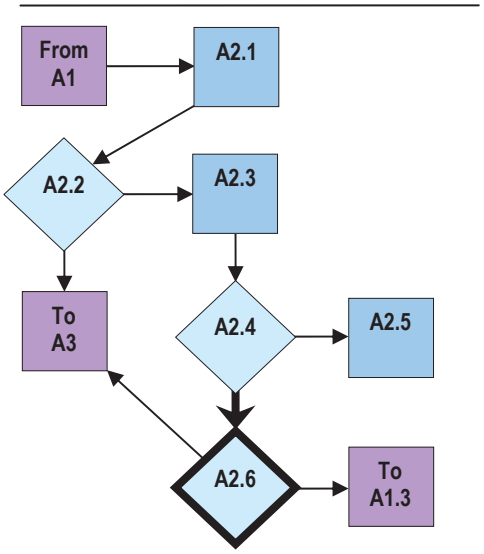
A2.4
Description The Producer who is the proper custodian or authority of the records determines if it wants the Ingest Project to continue.
Uses None
Produces/Modifies None



A2.5
Description If the proper custodian or authority of the records does not want the Ingest Project to continue, the Ingest Project ends.
Uses None
Produces/Modifies Ingest Project Termination Notice

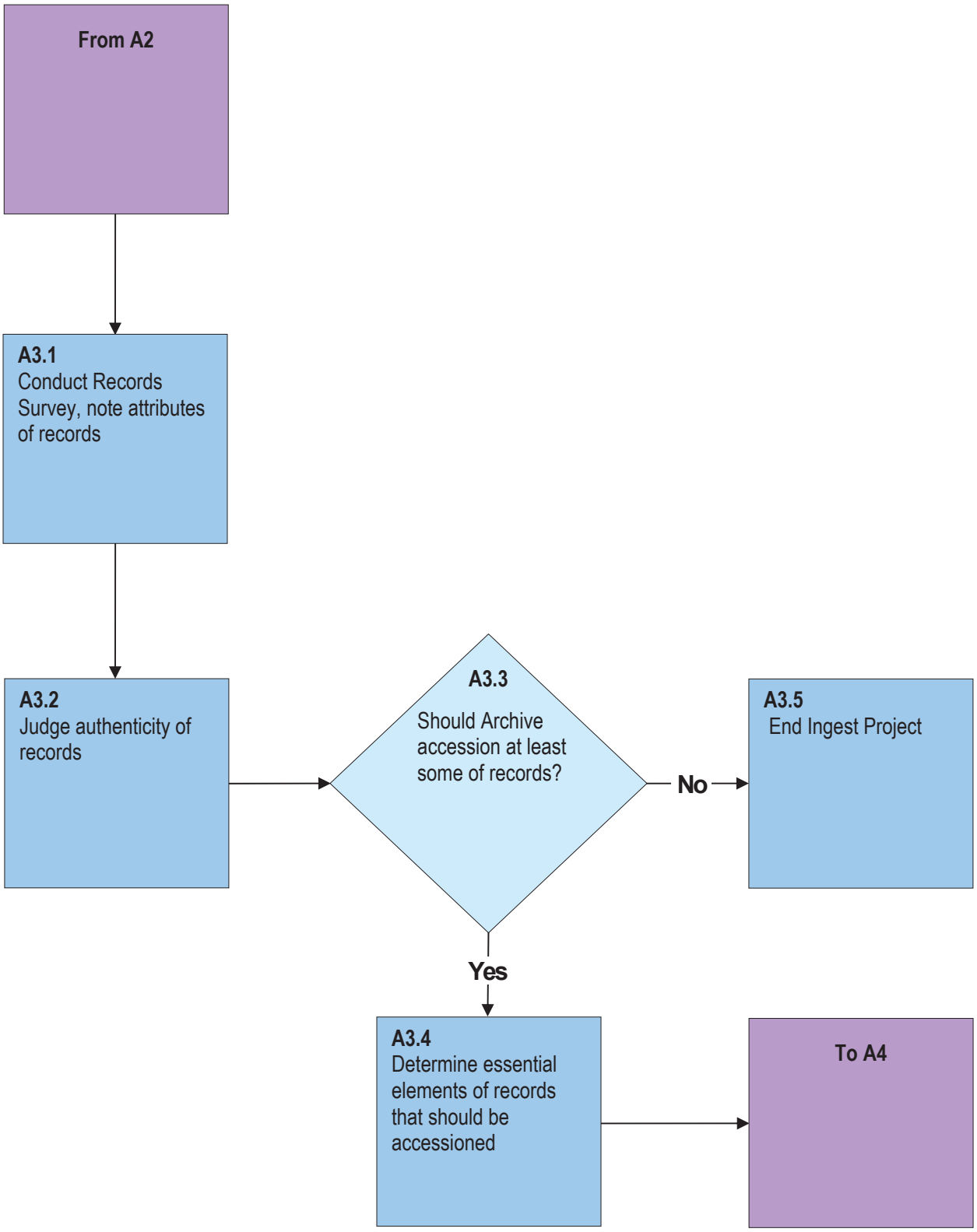


A2.6
Description The Producer who is the proper custodian or authority of the records determines if it wants to participate in the Ingest Project or designate the Entity from Step A1.1 as the Producer to continue the Ingest Project. If the Producer who is the proper custodian or authority of the records wants to participate in the Ingest Project, the Archive should return to Step A1.3 to determine if it has already established a Producer-Archive relationship with the Producer.
Uses None
Produces/Modifies None

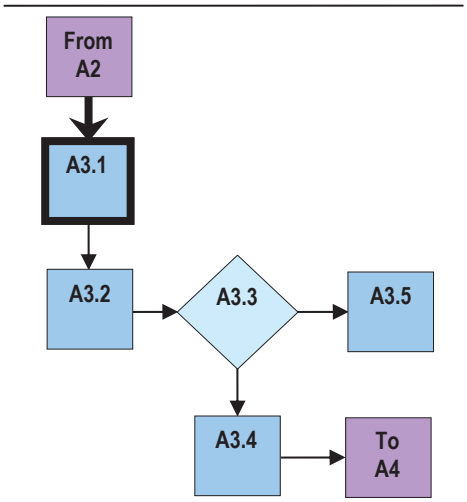


SECTION A: NEGOTIATE SUBMISSION AGREEMENT
Part A3: Collect Information and Assess Value of Records

Overview
Once the Archive has confirmed it is working with the proper custodian, it will conduct a survey of the records identified in Part A2 in order to collect the information needed to undertake the assessments described throughout the rest of Section A. In addition to collecting information the Archive will analyze the records in order to assess their continuing value and authenticity. This assessment will be combined with an assessment of the feasibility of preservation (Parts A4 through A10) in making a final appraisal decision. The appraisal work in this Part focuses on whether the records intellectually belong in the Archive, asking if it *should* accession the records.

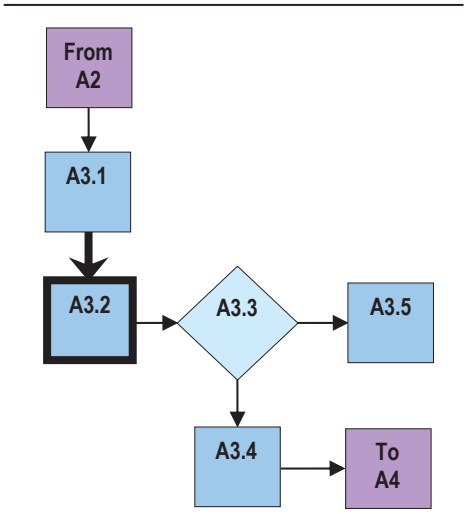

From A2
A3.1
Conduct Records Survey, note attributes of records
A3.2
Judge authenticity of records
A3.3
Should Archive accession at least some of records?
No
A3.5
End Ingest Project
Yes
A3.4
Determine essential elements of records that should be accessioned
To A4

A3.1



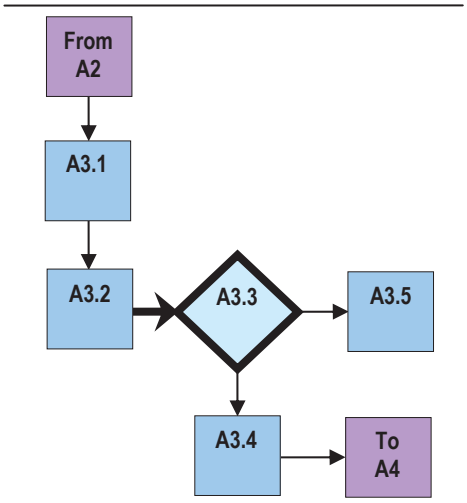
Description The Archive conducts a survey of the records identified in Step A2.1, delineating the descriptive data and other information that must be collected in order to make the assessments required in Parts A3 through A10. The Archive will note the record type of the records, the function(s) the records play for the Producer, their recordkeeping environment, format types, file size, Producer-created identifiers, confidentiality requirements, and copyright status in the Survey Report. Such a survey might be undertaken through any combination of a number of different methods, including the Archive interviewing the Producer, the Archiving querying the Producer through some sort of questionnaire, or even the Archive querying the electronic records themselves. The Survey Report should capture all of the information that the Archive will need to complete Parts A4 through A10.
Uses Survey Procedures, Survey Instrument, Survey Report
Produces/Modifies Survey Report

A3.2

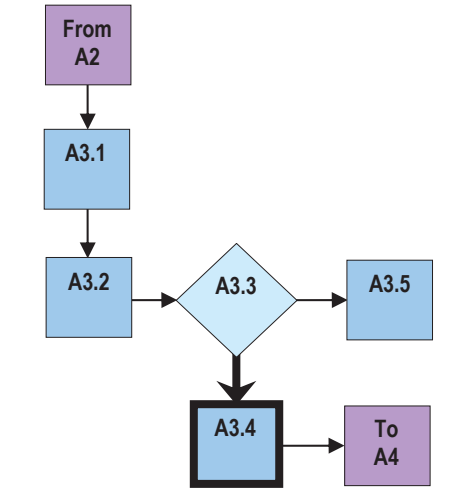


Description The Archive analyzes and judges the grounds for presuming the authenticity of the records identified in Step A2.1. This includes determining if the recordkeeping system that maintains the records has the qualities that allows one to presume the records' authenticity. The Archive should base this determination on the "Requirements for Trustworthy Recordkeeping Systems and the Preservation of Electronic Records in a University Setting" or some other set of requirements for trustworthy recordkeeping systems. The Archive may forgo this evaluation if it has already identified the recordkeeping system as a trustworthy system in the Recordkeeping System Report. If the Archive evaluates a new or modified system, it should document the system in a new or updated Recordkeeping System Report. Judging the grounds for presuming the authenticity of the records in the Ingest Project also includes checking if the records are managed according to the rules of the recordkeeping system, and sometimes examining the extrinsic and intrinsic qualities of the records themselves.
Uses Recordkeeping System Report, Recordkeeping System Evaluation Tool
Produces/Modifies Survey Report, Recordkeeping System Report

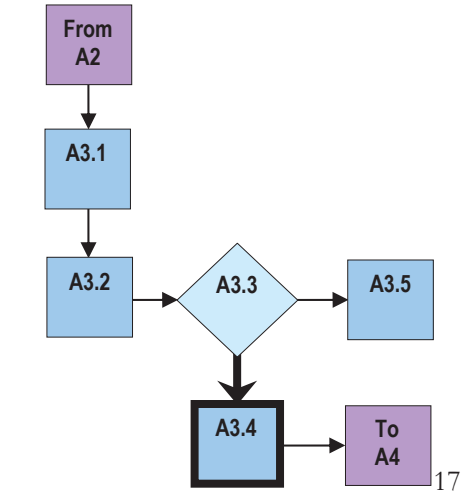
A3.3
Description The Archive determines if it should accession at least some of the records it surveyed. In this Step, this determination is based entirely on policy considerations with no regard to an Archive's capacity or technical considerations, which will come later during the Ingest Project. The Archive makes this appraisal decision by considering the Producer, the record type of the records, the function(s) the records play for the Producer, and the records' authenticity against the Archive's Collection Policy and Records Retention Schedule.
Uses Survey Report, Collection Policy, Records Retention Policy
Produces/Modifies Survey Report



A3.4
Description The Archive determines the essential elements of the records—documentary components, elements of form, and digital or physical components—it needs to preserve in order to preserve the recordness and authenticity of the records. This Step will guide the Archive's later decisions concerning format transformation.
Uses Survey Report, Collection Policy, Records Retention Policy
Produces/Modifies Survey Report

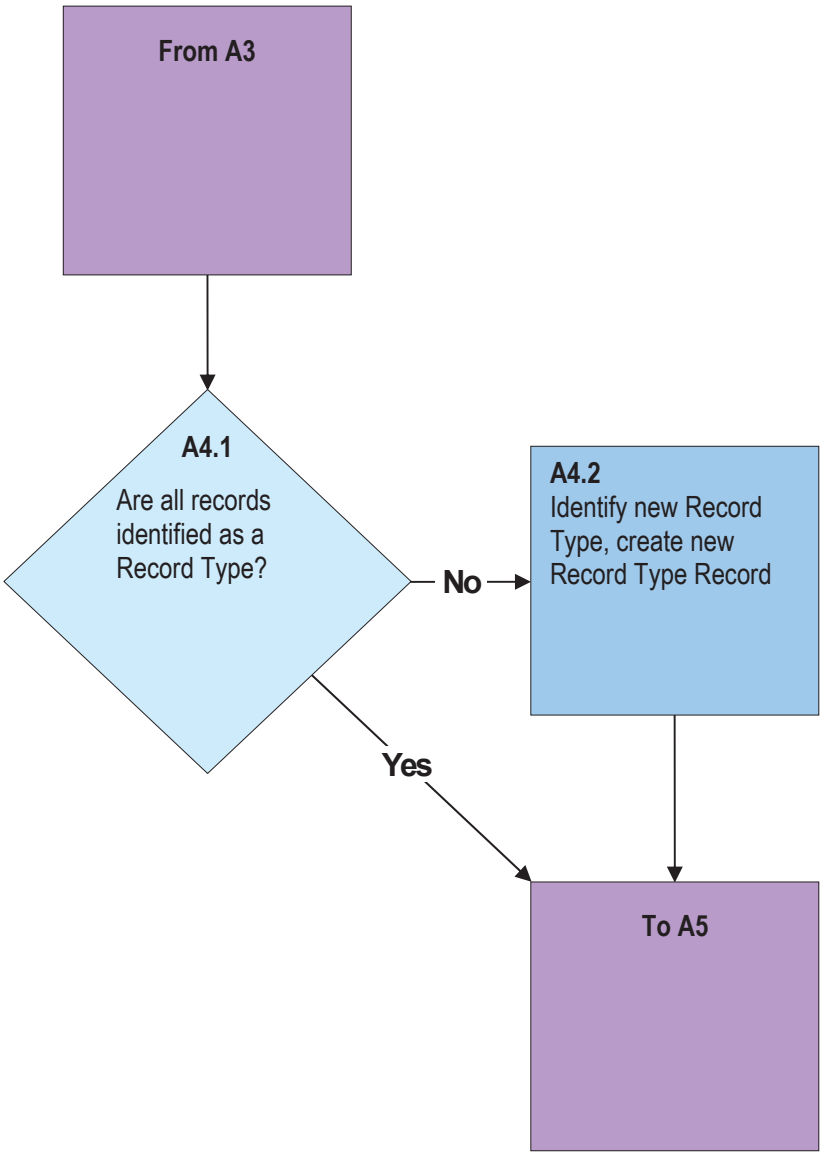


A3.5
Description If an Archive decides that it should not accession any of the records identified in Step A2.1, the Ingest Project ends.
Uses None
Produces/Modifies Ingest Project Termination Notice



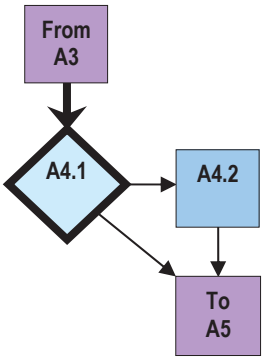
SECTION A: NEGOTIATE SUBMISSION AGREEMENT
Part A4: Assess Record Types

Overview
In this part the Archive determines if any of the records that should be accessioned are record types that the Archive has not previously defined or dealt with. If that is the case, the Archive will define the new record type. Record types define the nature of a class of records. They usually define retention and disposition of records, and, sometimes, their confidentiality status. Definitions of record types usually guide an Archive's appraisal and format transformation decisions.


From A3
A4.1
Are all records identified as a Record Type?
No
A4.2
Identify new Record Type, create new Record Type Record
Yes
To A5

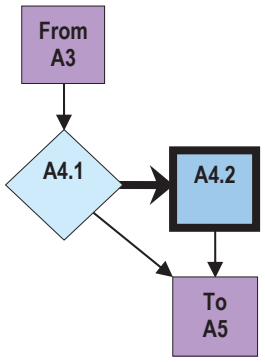
A4.1

Description Based on the information gathered in the Survey Report, the Archive determines if any of the records it should accession are record types that are not one of its established record types.
Uses Survey Report, Record Type Record
Produces/Modifies Record Type List



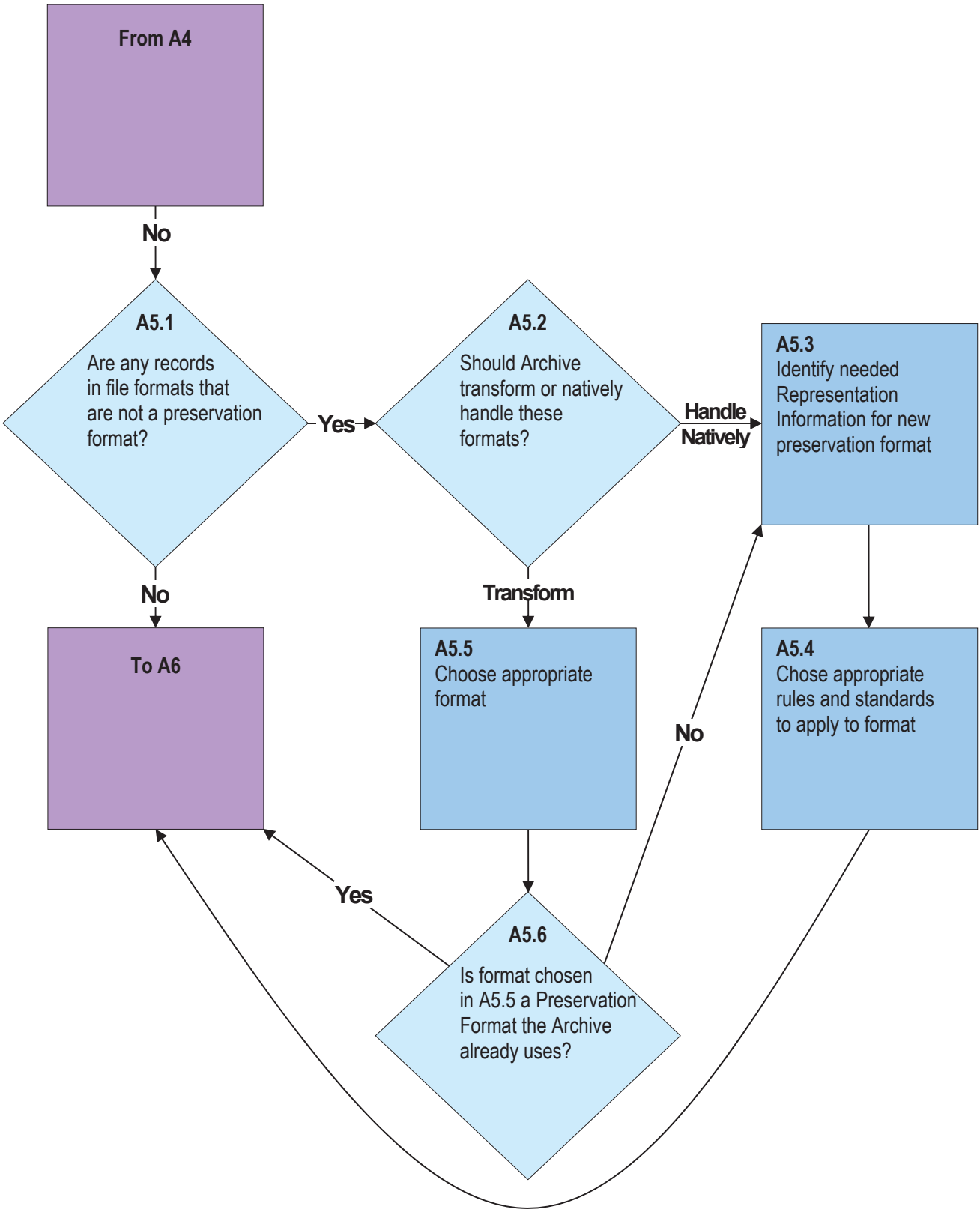
A4.2

Description If there are records in an Ingest Project that are a record type that is not one of the Archive's established record types, then it identifies this new record type and creates a new Record Type Record.
Uses Survey Report
Produces/Modifies Record Type Record



SECTION A: NEGOTIATE SUBMISSION AGREEMENT
Part A5: Assess Formats

Overview
In this Part the Archives appraises the formats of the records that it should accession into the Preservation System and determines if any of these records are in formats that are not one of the formats that the Preservation System supports. The Archives must determine if it will: 1) transform the records into one of the existing preservation formats, 2) transform the records into a new preservation format, or 3) keep the records in their existing formats and make that format a new preservation format. In order to create a new preservation format, the Archive will have to select a variety of rules and standards that apply to the format and identify the Representation Information that is needed for the new format.

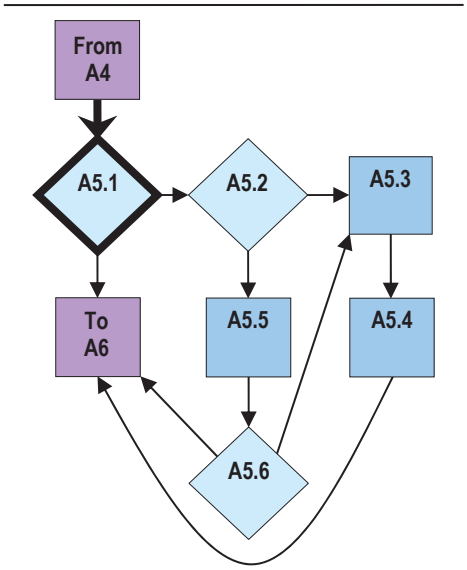

From A4
No
A5.1
Are any records in file formats that are not a preservation format?
Yes
A5.2
Should Archive transform or natively handle these formats?
Handle Natively
A5.3
Identify needed Representation Information for new preservation format
No
To A6
Transform
A5.5
Choose appropriate format
A5.4
Chose appropriate rules and standards to apply to format
No
Yes
A5.6
Is format chosen in A5.5 a Preservation Format the Archive already uses?

A5.1

Description Based on the information gathered in the Survey Report, the Archive determines if any of the digital components of records earmarked for preservation are in formats that meet the Archive's Format Standards Policy. If all of the digital components comply with the Format Standards Policy, the Archive proceeds to Part A6.

Uses Survey Report, Format Standards Policy, Format Representation Information System

Produces/Modifies Transformation Plan

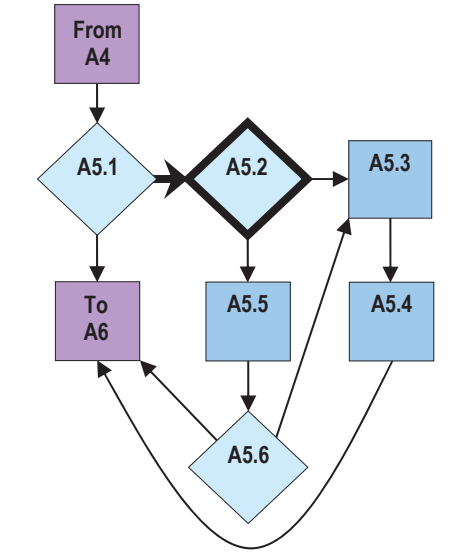


A5.2

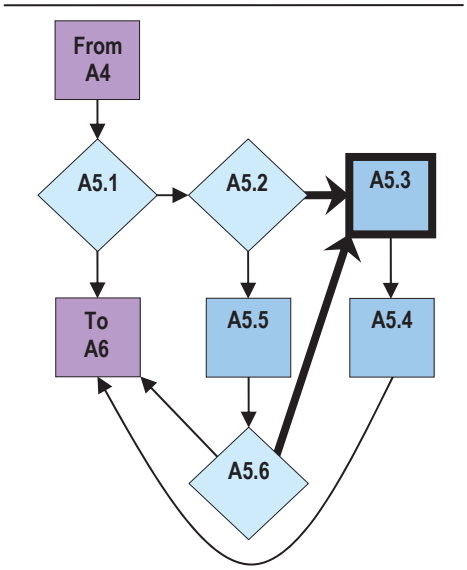
Description If there are digital components of records in an Ingest Project that are not compliant with the Archive's Format Standards Policy, the Archive must decide on the proper preservation strategy for these formats of digital components. The Archive may decide to transform the records into one of its preservation formats, manage the formats natively, or preserve both the native format and some preservation format. Although there are technical considerations to this determination, this is largely an appraisal decision. The Archive must determine how crucial the format is to the structure and ultimately the essential recordness of the record.

Uses Survey Report, Format Representation Information System, Preservation System Capabilities Report

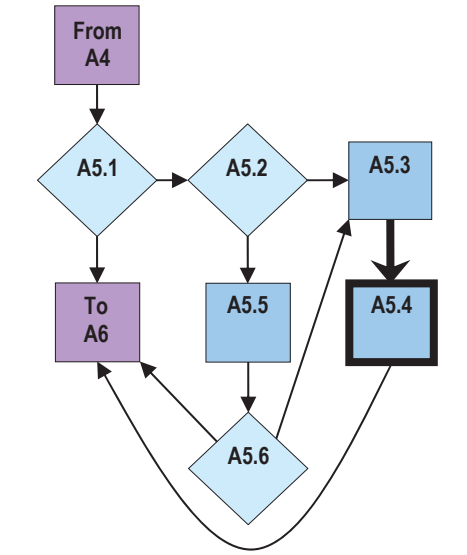
Produces/Modifies None



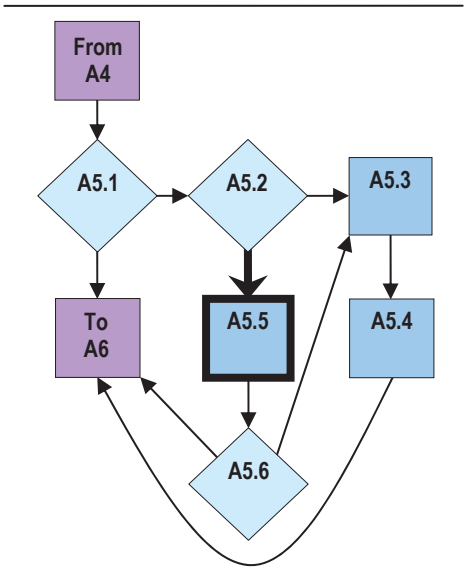
A5.3
Description If the Archive determines that it should preserve the records in the Preservation System in their existing native format(s), then it needs to establish this format in the Format Representation Information System and Formats Standards Policy.
Uses Designated Community Description
Produces/Modifies Format Representation Information System, Representation Information, Formats Standards Policy



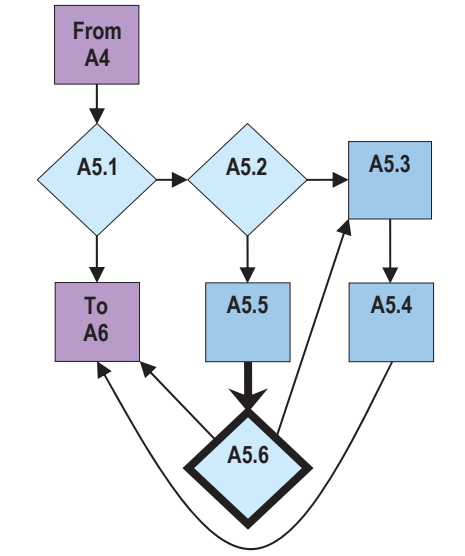
A5.4
Description In addition to the Representation Information, the Archive also needs to establish its rules and select appropriate standards for the new preservation format so it can manage and preserve records in that format successfully.
Uses None
Produces/Modifies Format Standards Policy, Transformation Plan



A5.5
Description If the Archive decides to transform the records, it must determine into which preservation format(s) to transform the records. See Step A5.2 for comments about the appraisal considerations involved in this step. If the Archive is dealing with a new combination of record type and format type, it will probably have to add to or modify its Transformation Policy.
Uses Survey Report, Transformation Policy
Produces/Modifies Transformation Policy, Transformation Plan

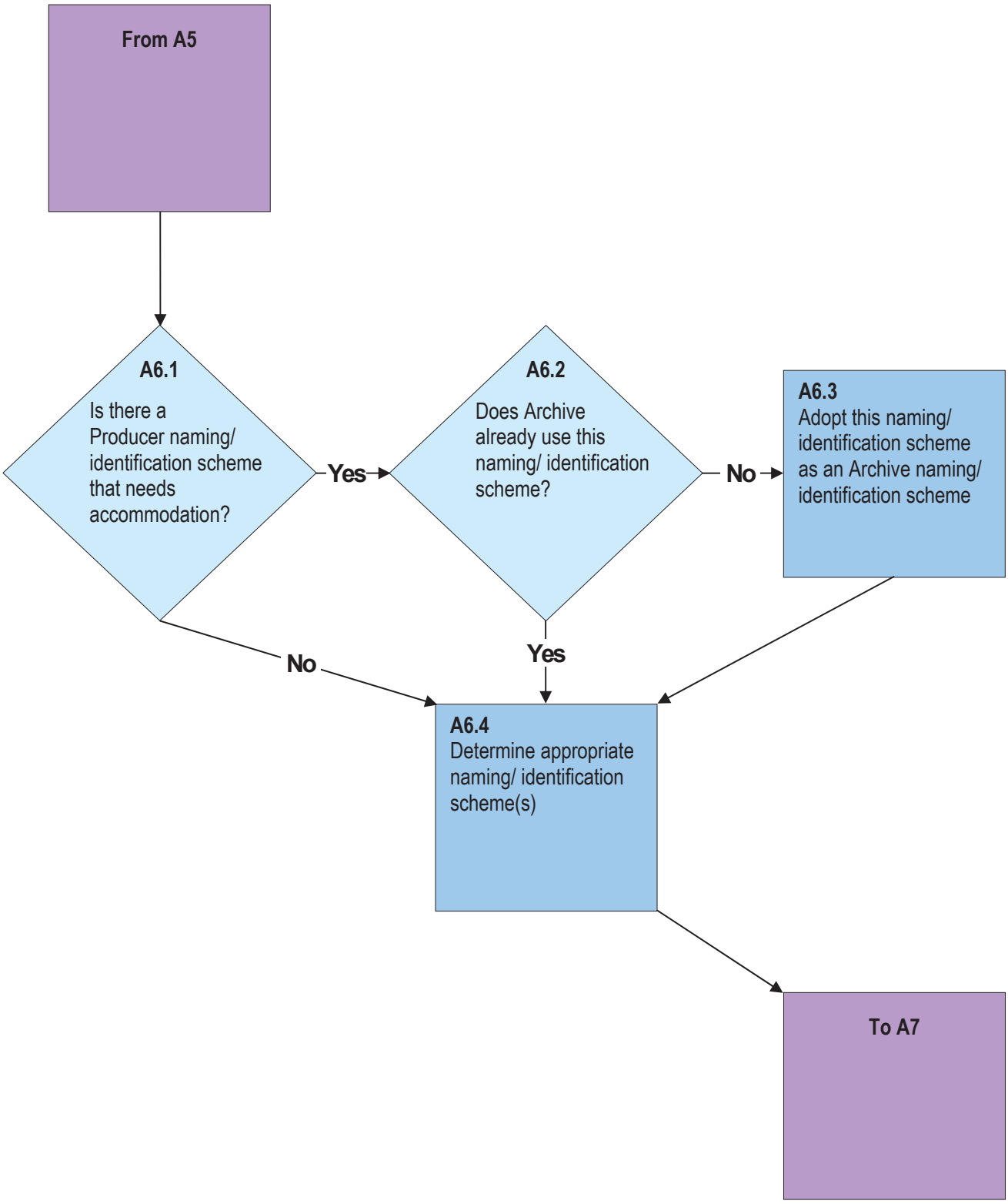


A5.6
Description The preservation format into which the Archive chooses to transform a record maybe either an existing preservation format or a new preservation format. If it is a new preservation format, the Archive goes to Step A5.3.
Uses None
Produces/Modifies None

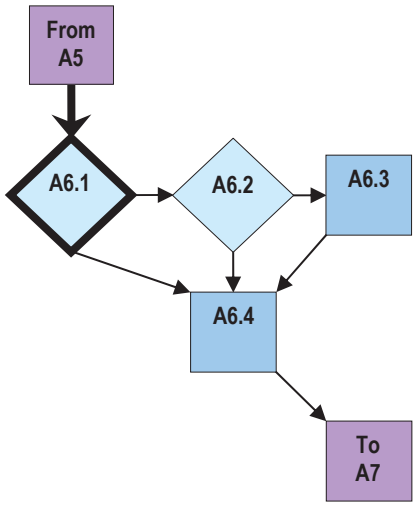


SECTION A: NEGOTIATE SUBMISSION AGREEMENT
Part A6: Assess Identifier Rules

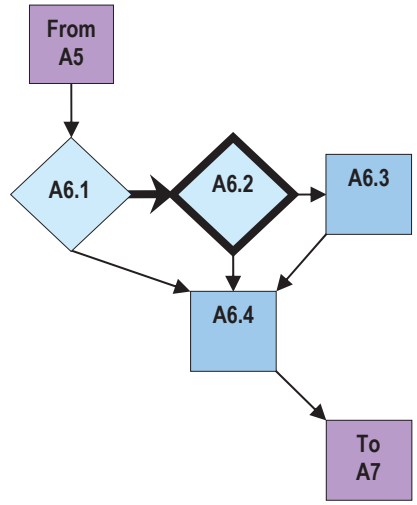
Overview
In this Part the Archive determines if it needs to preserve a Producer identification scheme with the records it should accession. For example, if a Producer's website is involved in an Ingest Project, the Archive may decide to preserve the file and path names of the HTML files to preserve the integrity of the website's internal links. In another instance the Archive may want to preserve a Producer's file and directory names of desktop applications stored on a network file system. If the Archive does need to preserve such a scheme, then the Archive must determine if it already manages the scheme in the Preservation System. If it does not, then it needs to create new rules to accommodate the new scheme.


From A5
A6.1
Is there a Producer naming/ identification scheme that needs accommodation?
Yes
A6.2
Does Archive already use this naming/ identification scheme?
No
A6.3
Adopt this naming/ identification scheme as an Archive naming/ identification scheme
No
Yes
A6.4
Determine appropriate naming/ identification scheme(s)
To A7

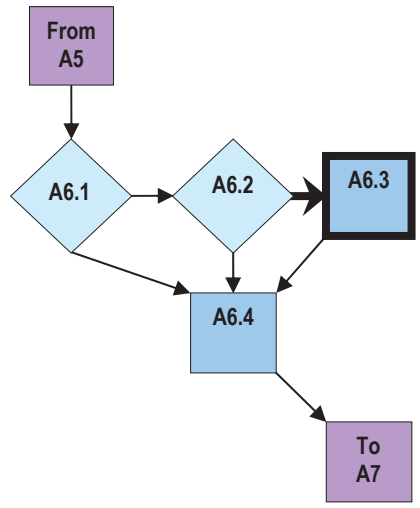
A6.1
Description Based on the information gathered in the Survey Report, the Archive determines if any of the records it should accession need to remain tied to a Producer's naming or identification scheme in the Preservation System.
Uses Survey Report, Producer Naming/Identification Scheme
Produces/Modifies None



A6.2
Description The Archive determines if it has already adapted that Producer Naming/Identification Scheme as an Archive Naming/Identification Scheme.
Uses Archive Naming/Identification Scheme, Producer Naming/Identification Scheme
Produces/Modifies None



A6.3
Description If the Archive has not adapted the Producer Naming/Identification Scheme as an Archive Naming/Identification Scheme, it will do so.
Uses Producer Naming/Identification Scheme
Produces/Modifies Archive Naming/Identification Scheme

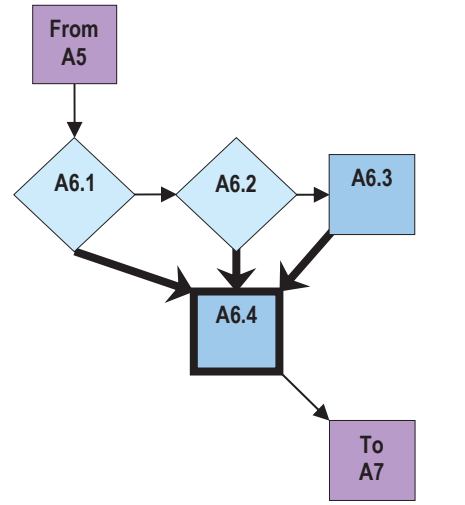


A6.4

Description The Archive determines which of its Naming/Identification Schemes it should use to name or identify the records in an Ingest Project. In many cases the Archive does not really have a choice to make because it only supports one scheme.

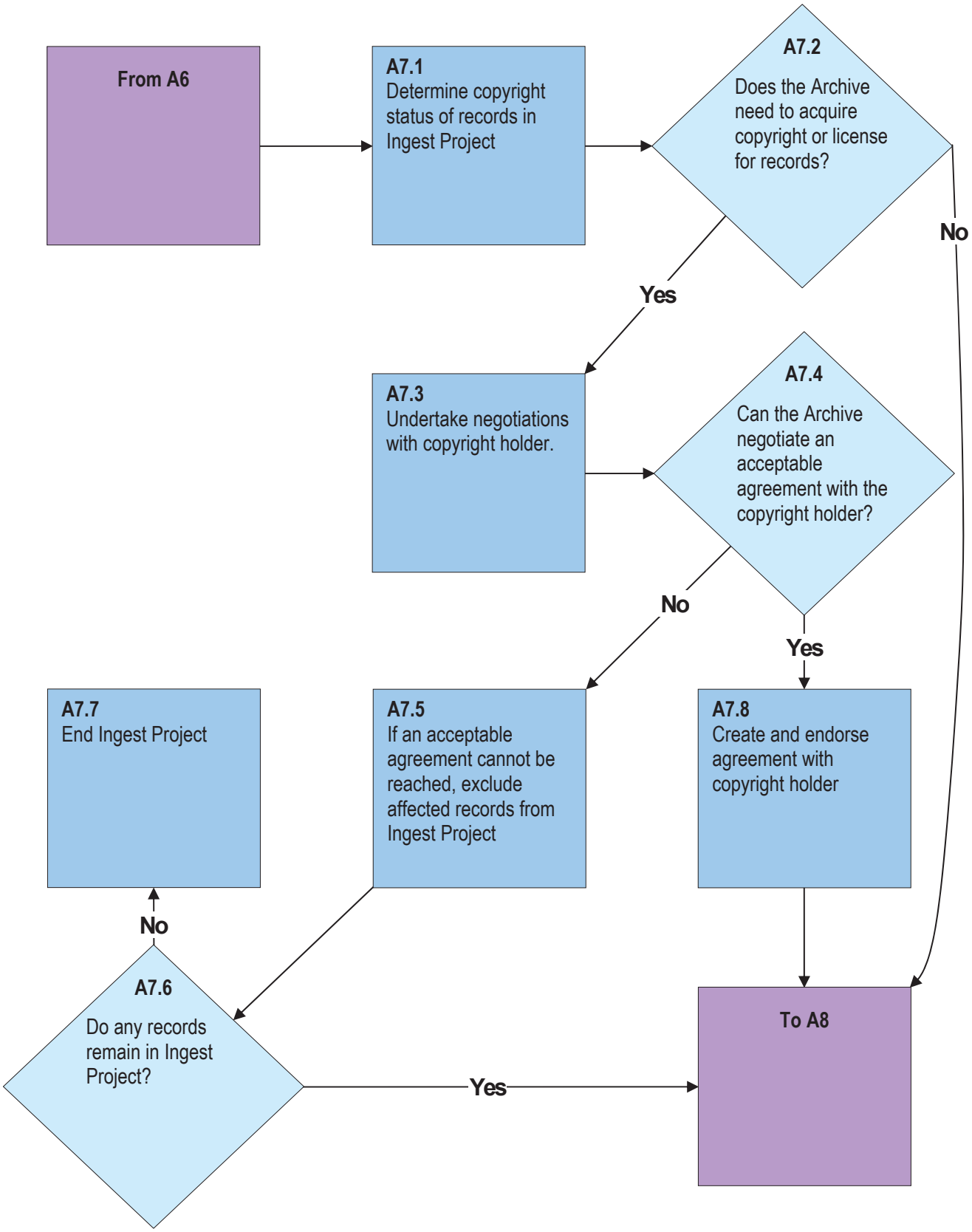
Uses Archive Naming/Identification Scheme, Survey Report

Produces/Modifies Archive Naming/Identification Scheme Decision



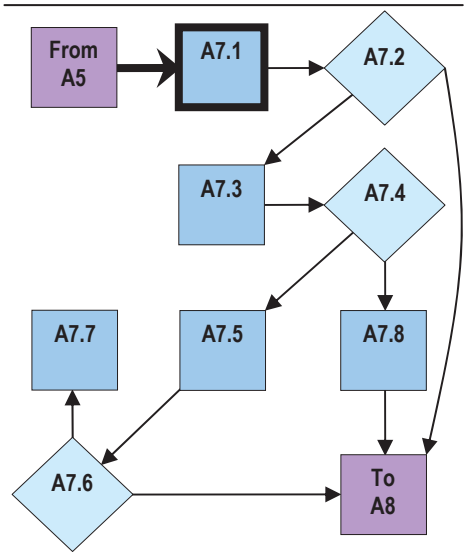
SECTION A: NEGOTIATE SUBMISSION AGREEMENT
Part A7: Assess Copyright

Overview
In this Part the Archive determines the copyright status of the records that it should accession and the copyright status of any associated software. In particular, the Archive determines if it already has the copyright of the records, or needs to obtain the copyrights or a licensing agreement for the records in an Ingest Project. This will allow the Archive to determine the impact copyright has on the feasibility of preserving the records it should accession.


From A6
A7.1
Determine copyright status of records in Ingest Project
A7.2
Does the Archive need to acquire copyright or license for records?
No
Yes
A7.3
Undertake negotiations with copyright holder.
A7.4
Can the Archive negotiate an acceptable agreement with the copyright holder?
No
Yes
A7.7
End Ingest Project
A7.5
If an acceptable agreement cannot be reached, exclude affected records from Ingest Project
A7.8
Create and endorse agreement with copyright holder
No
A7.6
Do any records remain in Ingest Project?
Yes
To A8

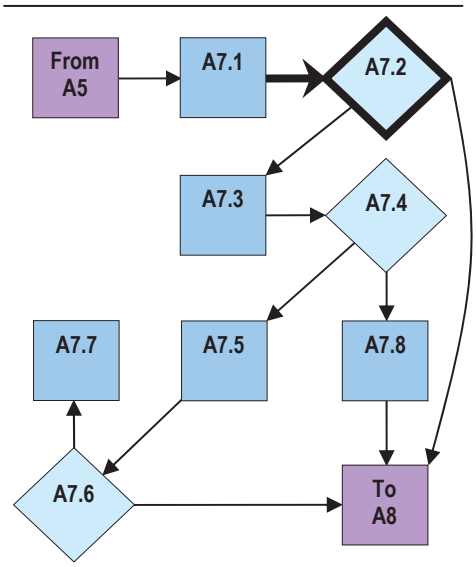
A7.1

Description Based on the information gathered in the Survey Report, the Archive determines the copyright status of the records it should accession. Consideration must be given not only to the copyright of the records' themselves, but also to the copyright of any associated software.
Uses Survey Report
Produces/Modifies Copyright Status

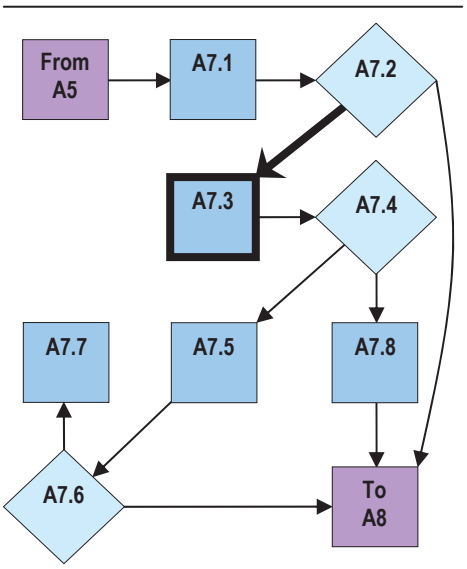


A7.2

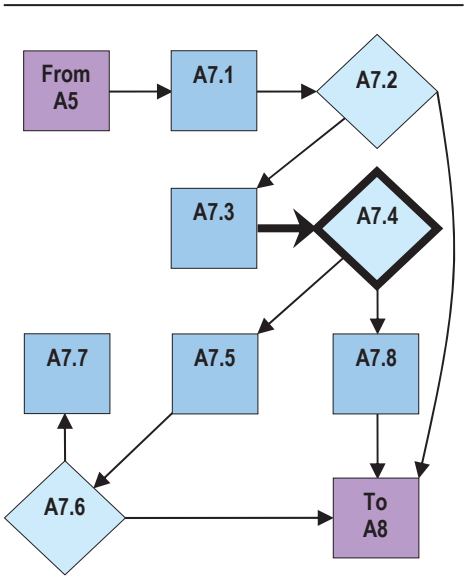
Description Based on its finding in Step A7.1, the Archive should determine if it needs to acquire the copyright of the records and/or associated software through a legal transfer or acquire a license for the use of the records and/or associated software from the copyright holder.
Uses Copyright Policy, Copyright Status
Produces/Modifies None



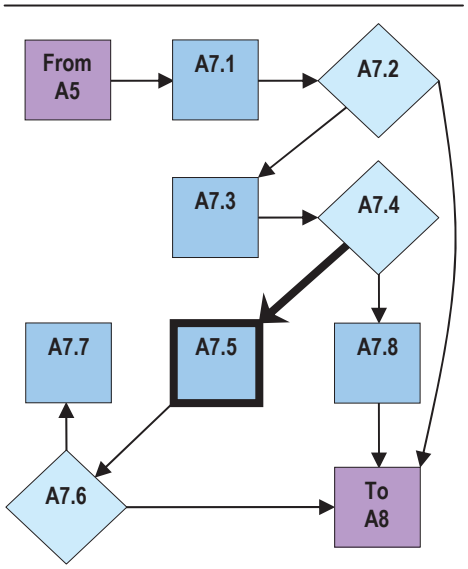
A7.3
Description If the Archive needs to acquire the copyright or licensing rights for at least some of the records and/or associated software it should accession, the Archive should initiate a negotiation process with the rights holder, which may or may not be the Producer. The Archive should use its Copyright Policy as its guide in these negotiations.
Uses Copyright Policy
Produces/Modifies None



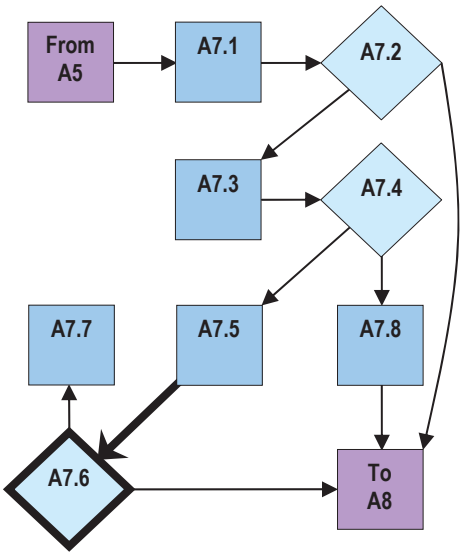
A7.4
Description The Archive determines if it can negotiate a rights transfer or licensing agreement with the rights holder, which it finds reasonable and with which it is capable of complying. Clearing and/or obtaining rights is likely to be resource intensive and difficult. The Archive should make this determination based on its Copyright Policy.
Uses Copyright Policy
Produces/Modifies None



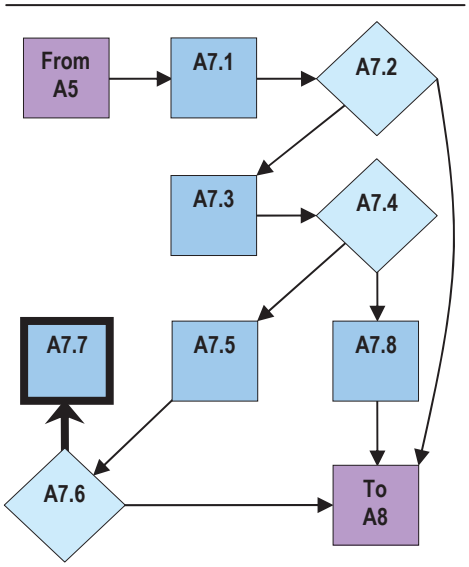
A7.5
Description If the Archive needs to acquire the copyright or licensing rights to at least some of the records and/or associated software it should accession, and the Archive determines it cannot negotiate a reasonable rights transfer or licensing agreement, it should exclude the affected records from the Ingest Project and modify the Survey Report accordingly.
Uses None
Produces/Modifies Survey Report



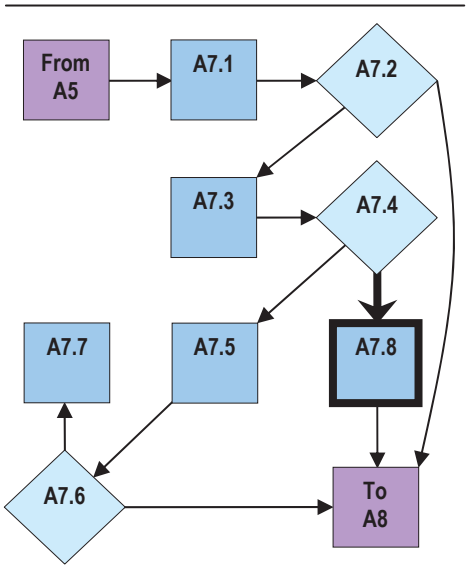
A7.6
Description Based on its actions in Step A7.5, the Archive determines if any records remain in the Ingest Project.
Uses Survey Report
Produces/Modifies None



A7.7
Description If no records remain in the Ingest Project, the Archive ends the Project.
Uses None
Produces/Modifies Ingest Project Termination Notice

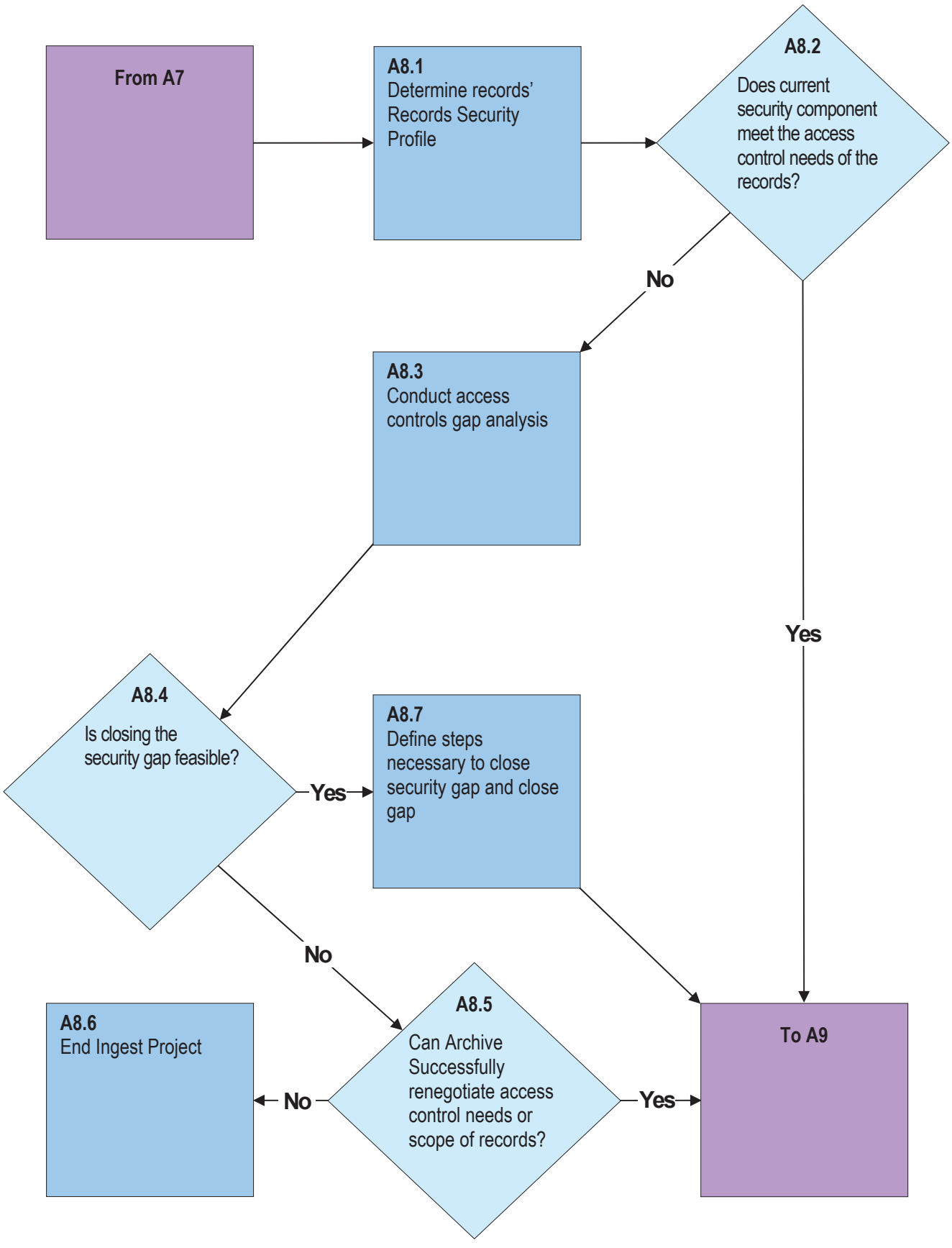


A7.8
Description If the Archive determines it can negotiate a reasonable copyright transfer or licensing agreement for at least some of the records in the Ingest Project and/or their associated software; it creates and endorses the agreement it has negotiated with the copyright holder.
Uses None
Produces/Modifies Copyright Transfer/License

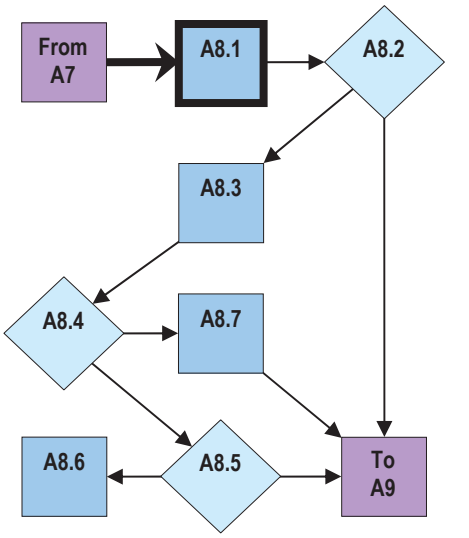


SECTION A: NEGOTIATE SUBMISSION AGREEMENT
Part A8: Assess Access Rights

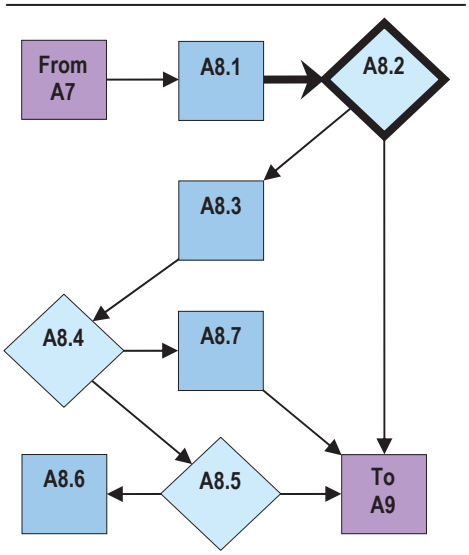
Overview
In this part, the Archive determines the appropriate Record Security Profile for the records in an Ingest Project. The Archive then determines if the security component of its Preservation System meets the access restriction requirements of the records earmarked for preservation. If the security system does not meet those requirements, the Archive determines if it is feasible to upgrade the security system to meet those requirements. If it is feasible, the Archive will upgrade the system; if it is not feasible, the Archive will not accept the records that require a level of security beyond what the Preservation System can provide.


From A7
A8.1
Determine records' Records Security Profile
A8.2
Does current security component meet the access control needs of the records?
No
Yes
A8.3
Conduct access controls gap analysis
A8.4
Is closing the security gap feasible?
Yes
A8.7
Define steps necessary to close security gap and close gap
No
A8.5
Can Archive Successfully renegotiate access control needs or scope of records?
No
Yes
A8.6
End Ingest Project
To A9

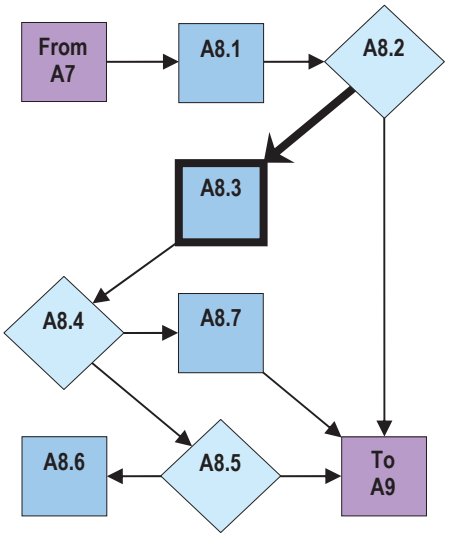
A8.1
Description Based on the information gathered in the records survey, along with its own Access Controls Policy, the Archive determines the appropriate Record Security Profile for the records in an Ingest Project. A Record Security Profile articulates the access control needs of a record. For some records, the Archive may have to create a new Record Security Profile.
Uses Survey Report, Access Controls Policy, Record Security Profile
Produces/Modifies Record Security Profile Decision, Record Security Profile



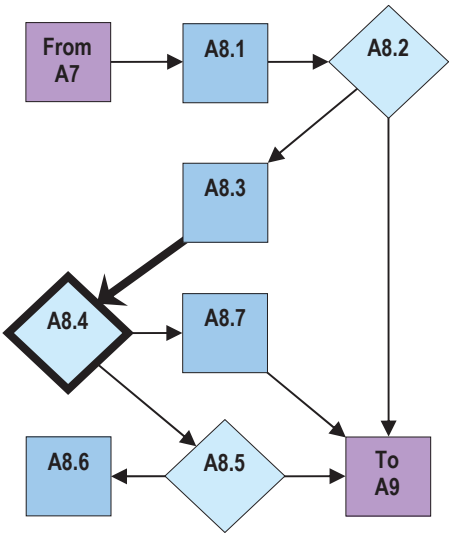
A8.2
Description The Archive determines if the current security component of its Preservation System meets the access control needs of the records in an Ingest Project.
Uses Access Controls Policy, Record Security Profile, Preservation System Capabilities Report
Produces/Modifies None



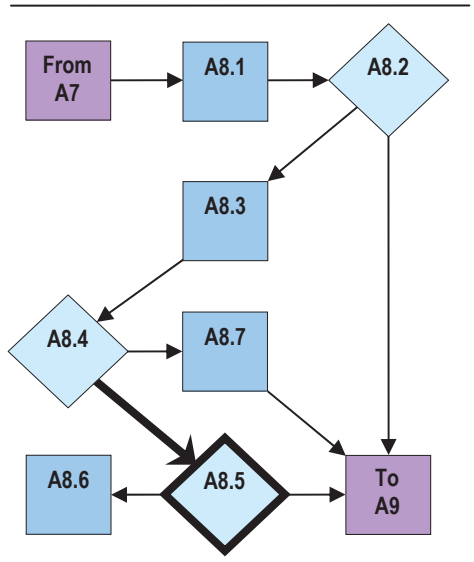
A8.3
Description If the Archive's security system does not meet the access control needs of the records in an Ingest Project, the Archive conducts a gap analysis of the security system and the records' access control needs to determine what improvements its security system would need to close the gap.
Uses Access Controls Policy, Record Security Profile, Preservation System Capabilities Report
Produces/Modifies Access Controls Gap Analysis



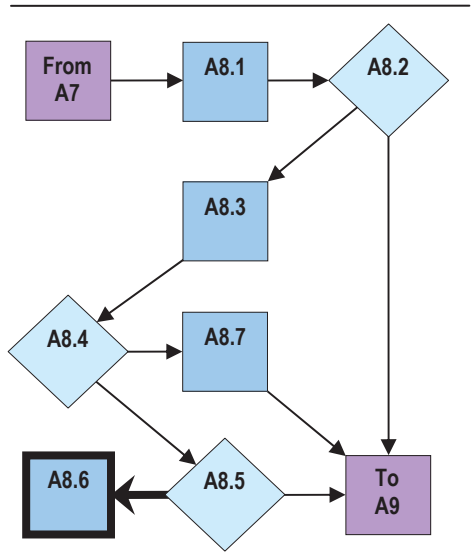
A8.4
Description The Archive determines if it is feasible to take the steps necessary to close the security gap as defined by the findings of A8.3
Uses Access Controls Gap Analysis
Produces/Modifies Access Controls Gap Analysis Feasibility Statement



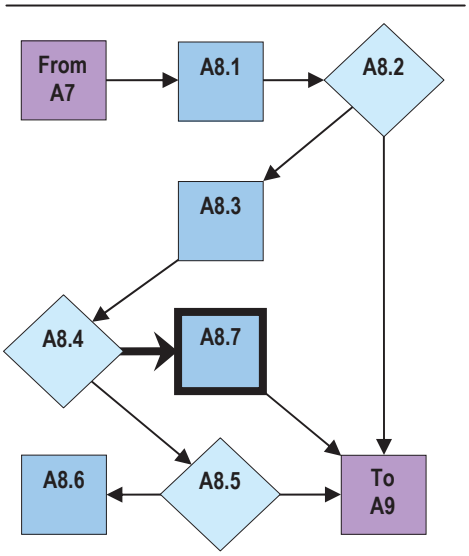
A8.5
Description: If the Archive determines that it is not feasible to close the security gap, then the Archive must renegotiate with the Producer to define a different set access control needs or a different scope for the Ingest Project, which excludes the affected records.
Uses Access Controls Gap Analysis, Survey Report
Produces/Modifies Survey Report, Record Security Profile



A8.6
Description If the Producer and the Archive are not able to renegotiate satisfactorily, either the Producer or the Archive ends the Project.
Uses None
Produces/Modifies Ingest Project Termination Notice

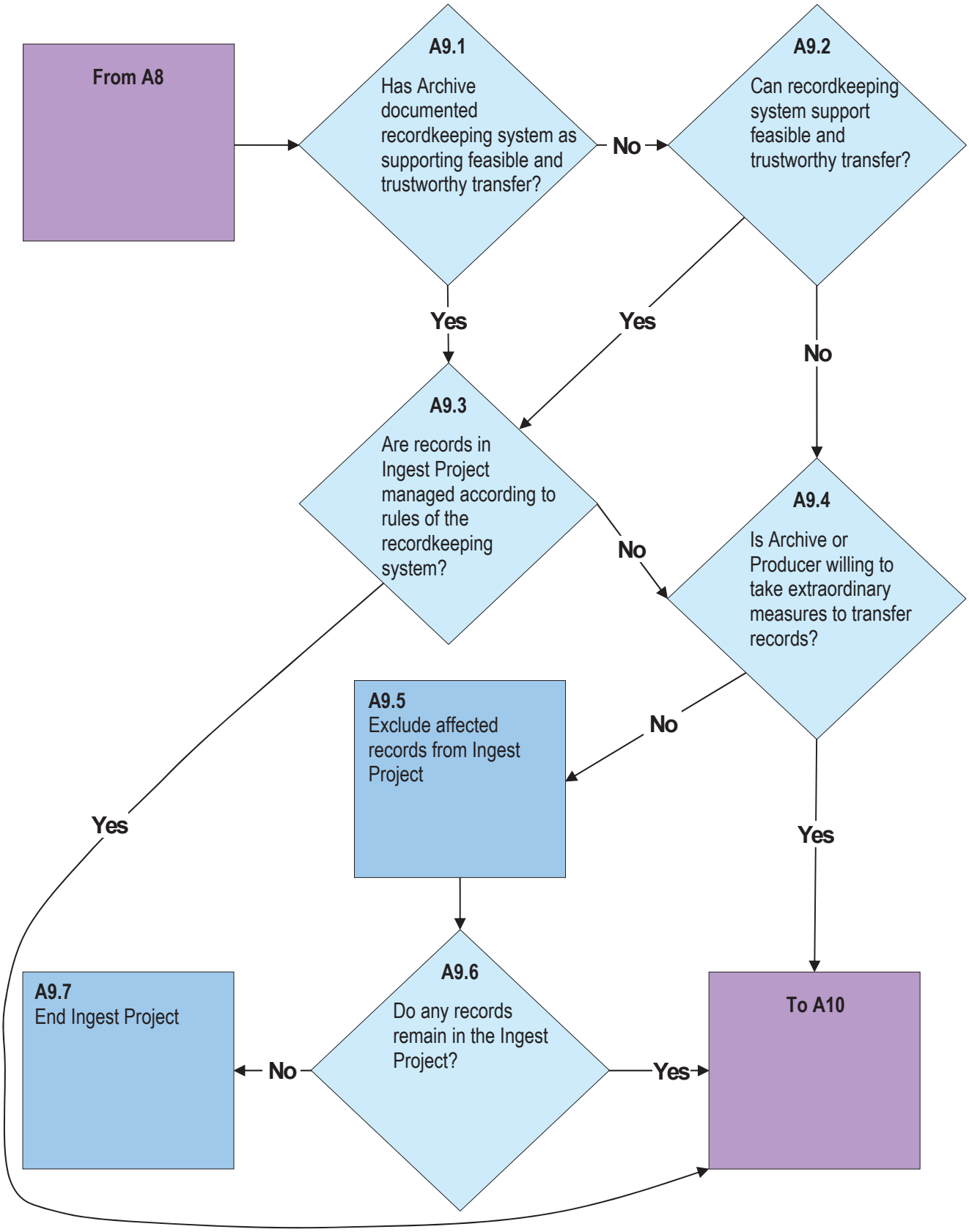


A8.7
Description If the Archive determines that it is feasible to take the steps necessary to close the security gap, then the Archive will define those steps and close the gap.
Uses Access Controls Gap Analysis, Access Controls Gap Analysis Feasibility Statement, Access Controls Policy
Produces/Modifies Access Controls Policy

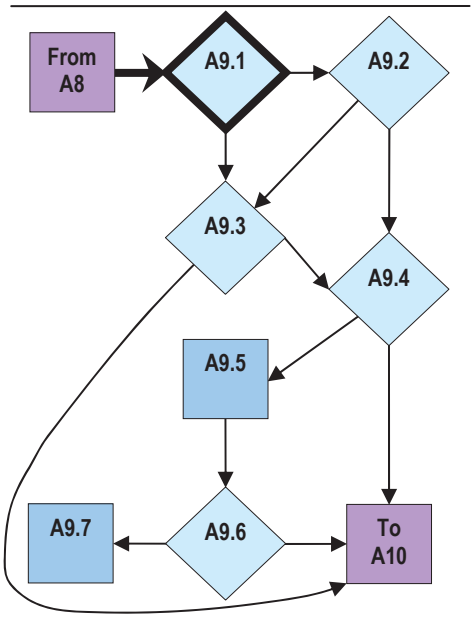


SECTION A: NEGOTIATE SUBMISSION AGREEMENT
Part A9: Assess Recordkeeping System

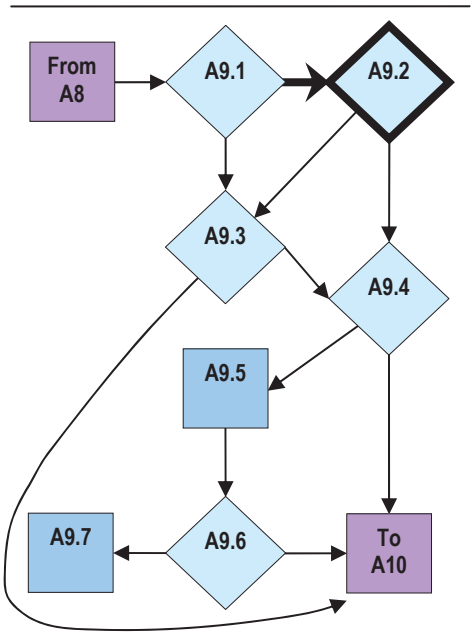
Overview
In this part the Archive determines if the active recordkeeping system managing the records in the Ingest Project allows, or is capable of enabling, a trustworthy transfer of records to the Archive in a feasible, scaleable manner. This assessment is entirely separate from any other assessment of the records themselves. Even if the records are authentic and are stored in acceptable record types and formats, and even if there are no problems with existing identifier rules, copyrights, or access rights, it may still simply be too difficult or expensive to facilitate a trustworthy transfer from the recordkeeping system to the preservation system. The Archive should base this determination on the "Requirements for Trustworthy University Electronic Records" or some other set of requirements for trustworthy recordkeeping systems (a Recordkeeping System Evaluation Tool) that evaluate the systems' ability to transfer records to preservation systems. The Archive may forgo this evaluation if it has already identified the capabilities of the recordkeeping system in a Recordkeeping System Report. If the Archive evaluates a new or modified system, it should document the system in a new or updated Recordkeeping System Report. Determining if records can be transferred in a feasible, scaleable, and trustworthy manner also includes checking that the records are managed according to the rules of the recordkeeping system.


From A8
A9.1
Has Archive documented recordkeeping system as supporting feasible and trustworthy transfer?
No
A9.2
Can recordkeeping system support feasible and trustworthy transfer?
Yes
Yes
No
A9.3
Are records in Ingest Project managed according to rules of the recordkeeping system?
No
A9.4
Is Archive or Producer willing to take extraordinary measures to transfer records?
A9.5
Exclude affected records from Ingest Project
No
Yes
Yes
A9.6
Do any records remain in the Ingest Project?
A9.7
End Ingest Project
No
Yes
To A10

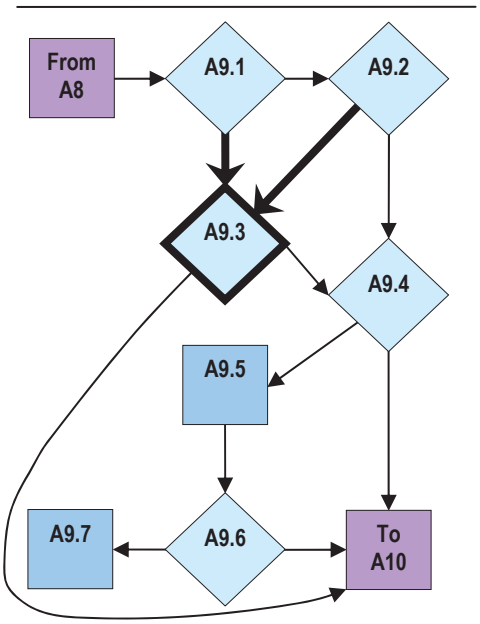
A9.1
Description The Archive determines if has already approved the recordkeeping system managing the records in the Ingest Project as a system from which it can transfer records to the Archive in a feasible, scaleable, and trustworthy manner.
Uses Recordkeeping System Report
Produces/Modifies None



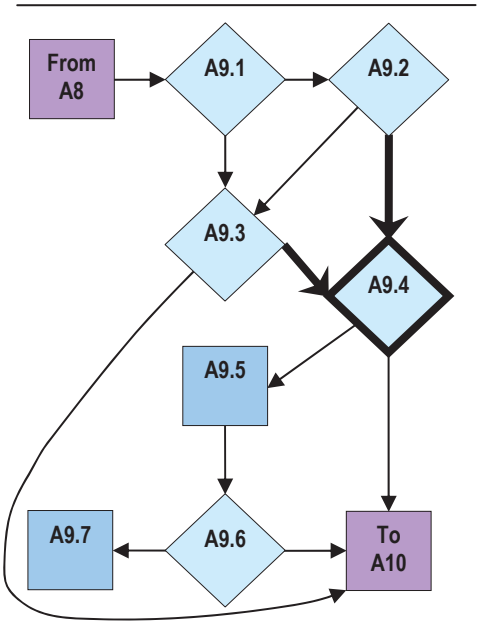
A9.2
Description The Archive evaluates and determines the recordkeeping system's ability to transfer records to the Archive in a feasible, scaleable, and trustworthy manner.
Uses Recordkeeping System Evaluation Tool
Produces/Modifies Recordkeeping System Report



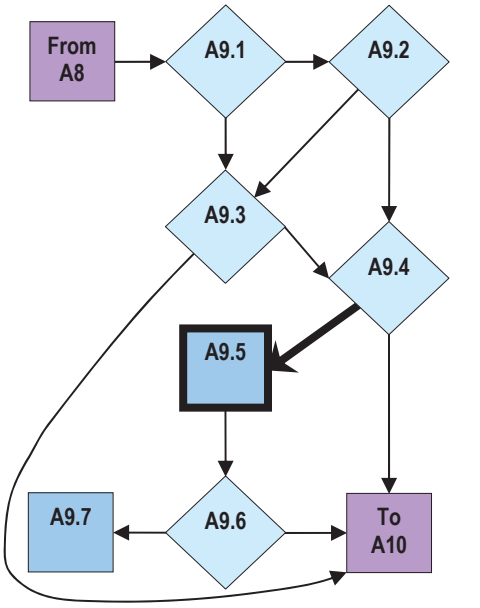
A9.3
Description The Archive determines if the records in the Ingest Project are managed in the recordkeeping system according to the rules of the recordkeeping system.
Uses Recordkeeping System Internal Rules
Produces/Modifies None



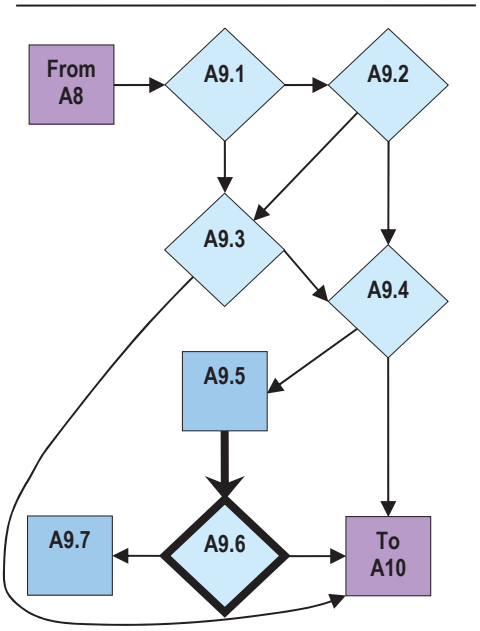
A9.4
Description The Archive and the Producer determine if either or both are willing to take extraordinary measures to transfer the records in the Ingest Project to the Archive in a trustworthy manner. These measures may range from building special tools to undertaking a software reengineering project. Such measures are rarely scaleable efforts that can be turned into regular procedures.
Uses None
Produces/Modifies SIP Creation Procedures



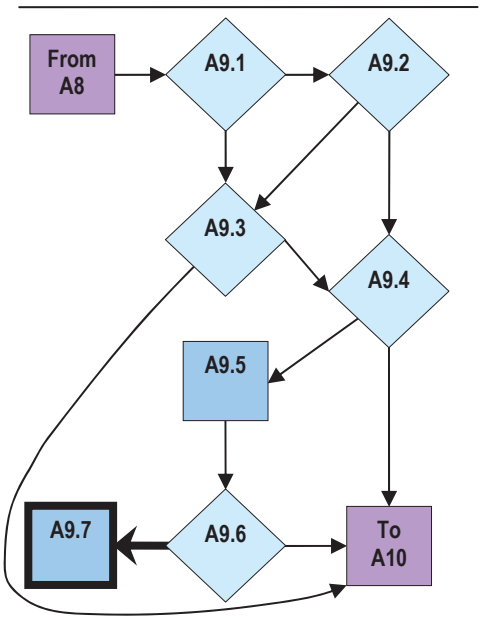
A9.5
Description The Archive excludes from the Ingest Project the records that it and the Producer cannot or are not willing to make the effort to transfer to the Archive in a trustworthy manner. The Archive should modify the Survey Report accordingly.
Uses None
Produces/Modifies Survey Report



A9.6
Description The Archive determines if any records remain in the Ingest Project after Step A9.6.
Uses Survey Report
Produces/Modifies None

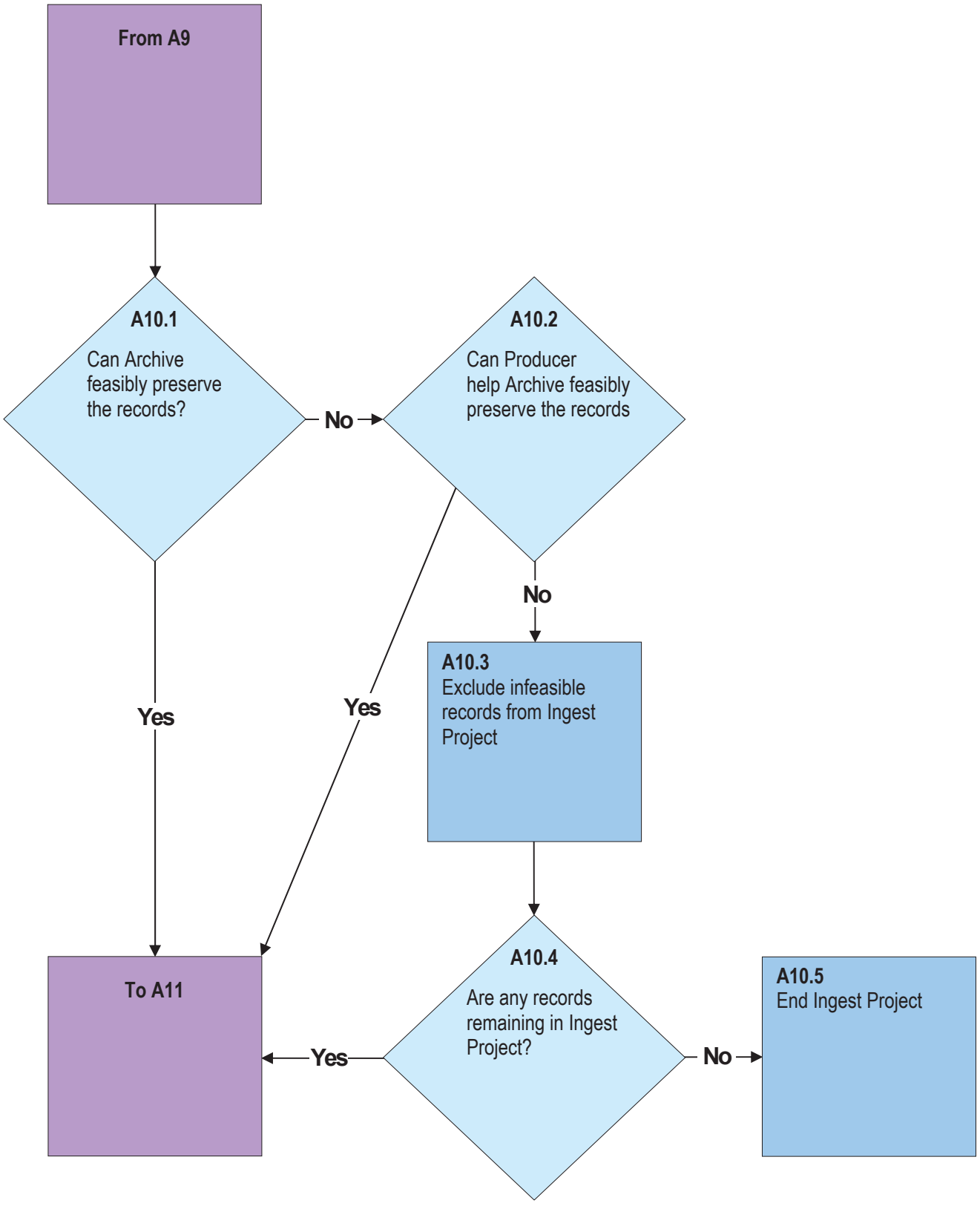


A9.7
Description If no records remain in its Ingest Project, the Archive ends the Project.
Uses None
Produces/Modifies Ingest Project Termination Notice

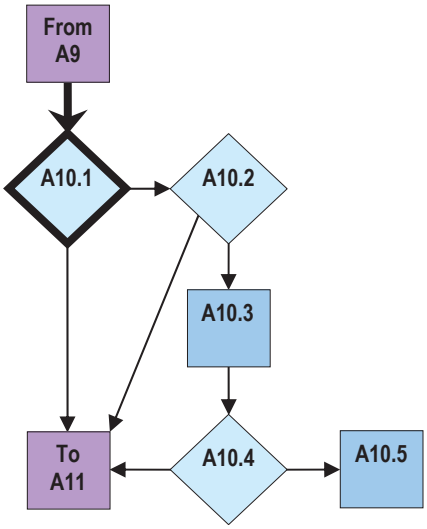


SECTION A: NEGOTIATE SUBMISSION AGREEMENT
Part A10: Assess Feasibility

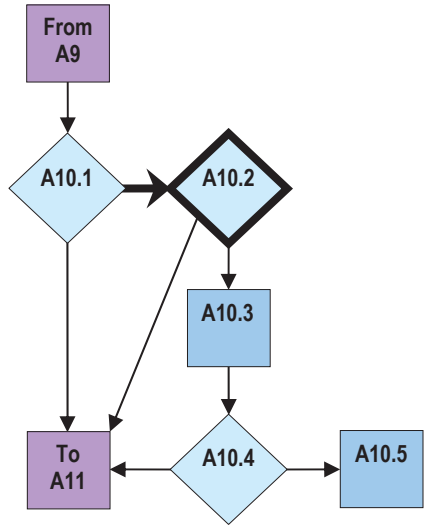
Overview
During this Part, the Archive assesses if it can accession, manage, and preserve the records it should accession in its Preservation System, either on its own, or with help from the Producer or a third party. This assessment is based on the information collected in the previous parts, including Assess Record Types, Assess Formats, Assess Identifier Rules, Assess Copyright, Assess Access Rights, and Assess Recordkeeping System. The Archive must determine if its existing resources for preservation formats, record types, identifier rules, creator records, security procedures, and system capabilities meets the needs of the records identified in Part A3. The resulting feasibility report should present a gap analysis if the Resources do not reflect the continuing value of any records assessed in A3. The Archive must then determine if it should modify or add to its Resources to meet those assessments or if it should reject or modify the scope of the records involved in the Ingest Project.


From A9
A10.1
Can Archive feasibly preserve the records?
No
A10.2
Can Producer help Archive feasibly preserve the records
Yes
No
Yes
A10.3
Exclude infeasible records from Ingest Project
To A11
A10.4
Are any records remaining in Ingest Project?
Yes
No
A10.5
End Ingest Project

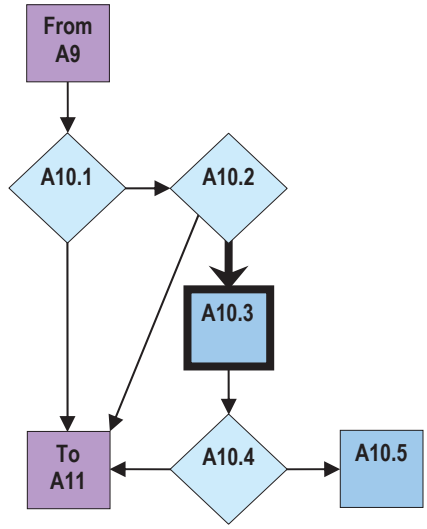
A10.1
Description The Archive assesses if it has the technical and staffing capacity or can draw on the resources of a third party to manage and preserve the records earmarked for preservation. The Archive documents its assessment in a Preservation System Availability Statement.
Uses Survey Report, Preservation System Capabilities Report
Produces/Modifies Preservation System Availability Statement



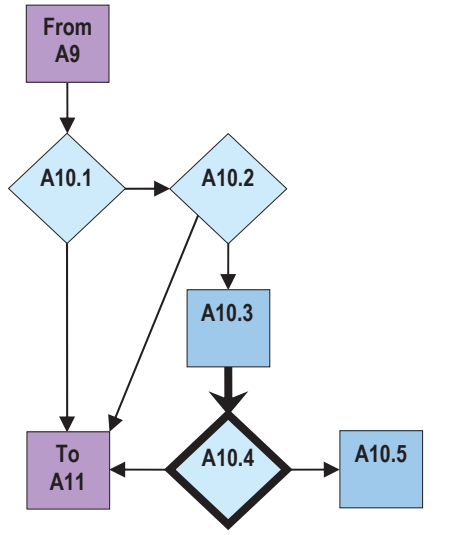
A10.2
Description If the Archive assesses that it cannot manage at least some of the records it should accession it asks the Producer if it can help the Archive make feasible the transfer of these records. If the Producer can help by producing resources, the Archive amends the System Availability Report to describe the resources that the Producer agrees to provide.
Uses Survey Report, Preservation System Capability Report
Produces/Modifies Preservation System Availability Statement



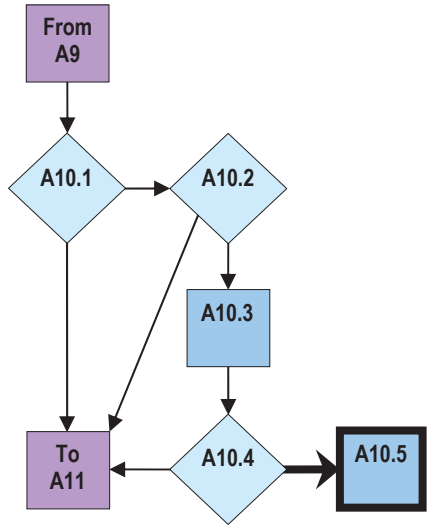
A10.3
Description If the Producer is unable or unwilling to help the Archive, then the Archive will exclude from the Ingest Project the records it cannot transfer, adjusting its Preservation System Availability Statement and Survey Report accordingly.
Uses Survey Report
Produces/Modifies Preservation System Availability Statement, Survey Report



A10.4
Description Based on its actions in Step A10.3, an Archive determines if any records remain in its Ingest Project.
Uses Survey Report
Produces/Modifies None

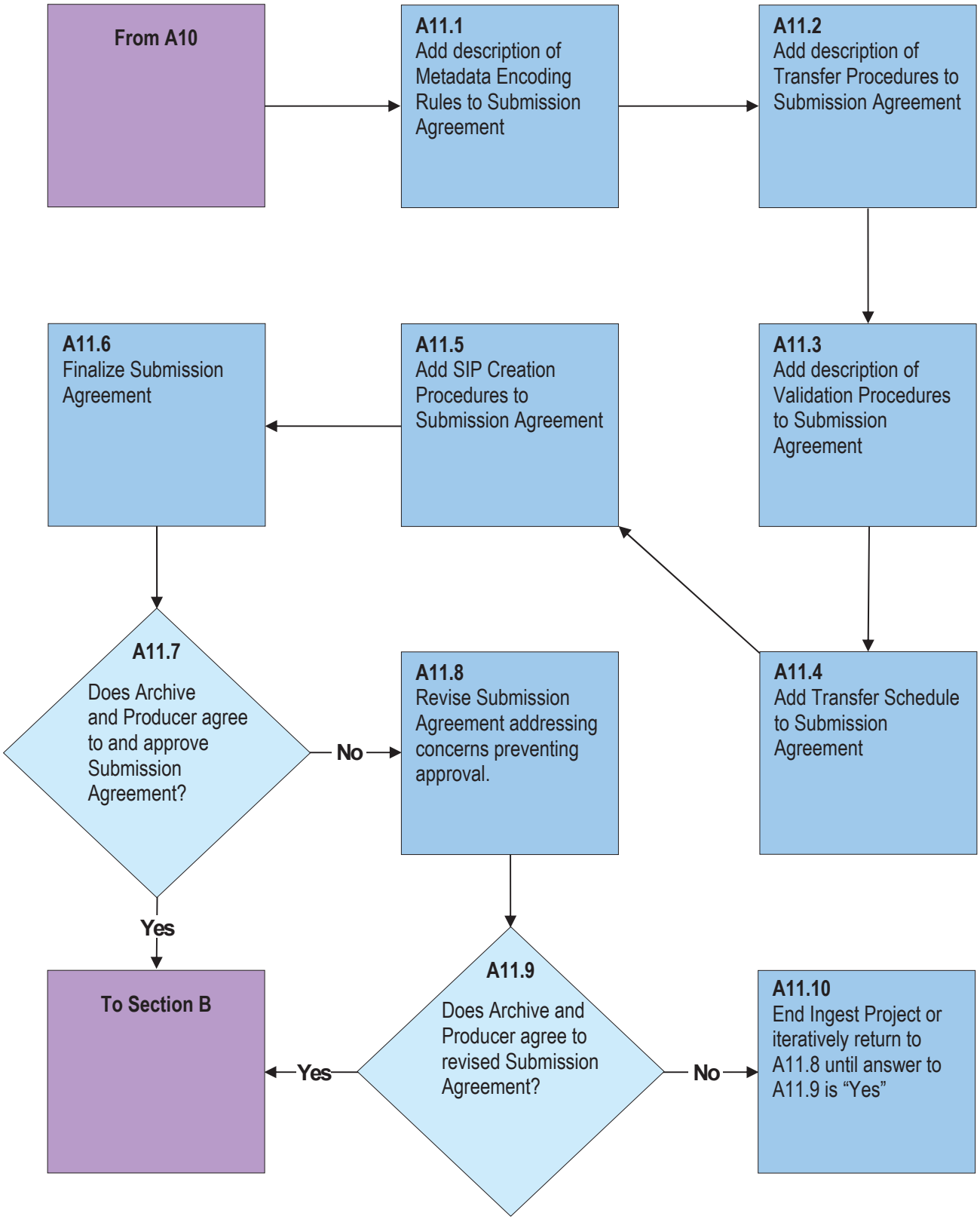


A10.5
Description If no records remain in its Ingest Project, the Archive ends the Project.
Uses None
Produces/Modifies Ingest Project Termination Notice

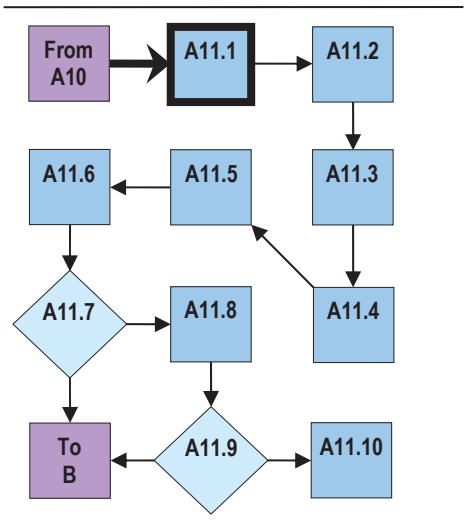


SECTION A: NEGOTIATE SUBMISSION AGREEMENT
Part A11: Finalize Submission Agreement

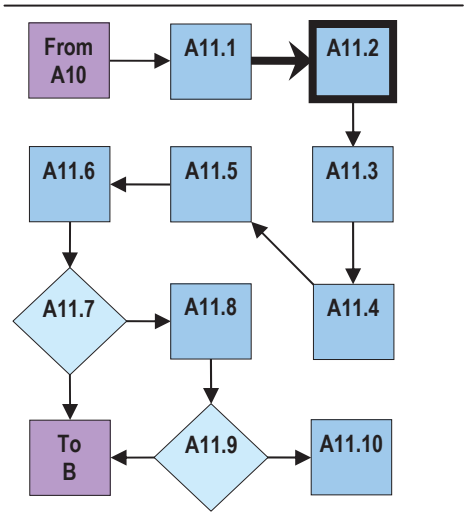
Overview
During this Part, the Archive adds description of Metadata Encoding Rules, Transfer Procedures and Schedules, and Validation Procedures to the Submission Agreement. Then the Archive and Producer work on finalizing the Submission Agreement until they both agree to endorse it.


From A10
A11.1
Add description of Metadata Encoding Rules to Submission Agreement
A11.2
Add description of Transfer Procedures to Submission Agreement
A11.3
Add description of Validation Procedures to Submission Agreement
A11.4
Add Transfer Schedule to Submission Agreement
A11.5
Add SIP Creation Procedures to Submission Agreement
A11.6
Finalize Submission Agreement
A11.7
Does Archive and Producer agree to and approve Submission Agreement?
No
Yes
A11.8
Revise Submission Agreement addressing concerns preventing approval.
A11.9
Does Archive and Producer agree to revised Submission Agreement?
Yes
No
To Section B
A11.10
End Ingest Project or iteratively return to A11.8 until answer to A11.9 is "Yes"

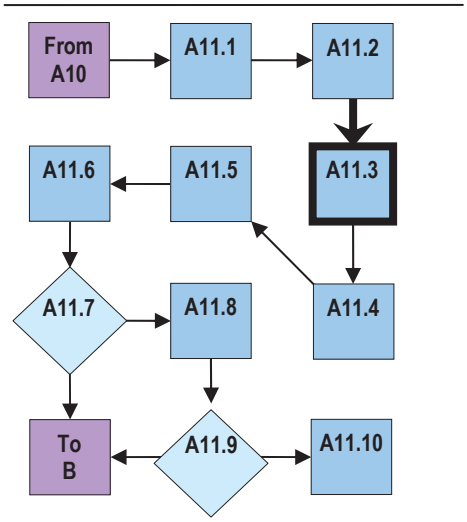
A11.1
Description The Archive attaches a description of its Metadata Encoding Rules to the Submission Agreement.
Uses Metadata Encoding Rules
Produces/Modifies Metadata Encoding Rules Decision



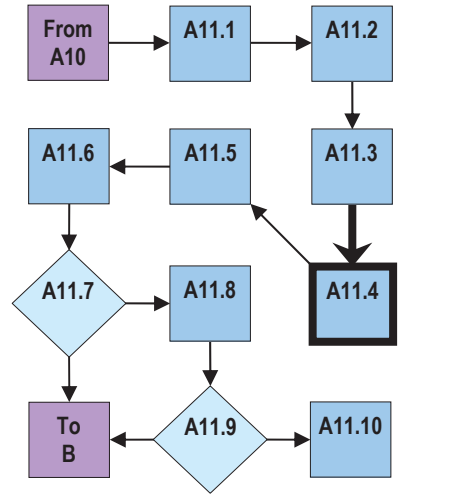
A11.2
Description The Archive attaches a description of its Transfer Procedures Rules to the Submission Agreement.
Uses Transfer Procedures
Produces/Modifies Transfer Procedures Decision



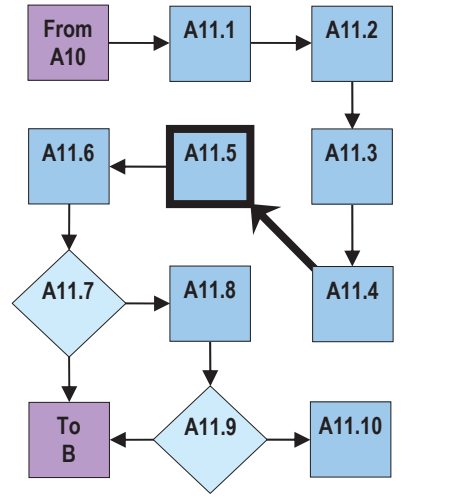
A11.3
Description The Archive attaches a description of its Validation Procedures to the Submission Agreement.
Uses Validation Procedures
Produces/Modifies Validation Procedures Decision



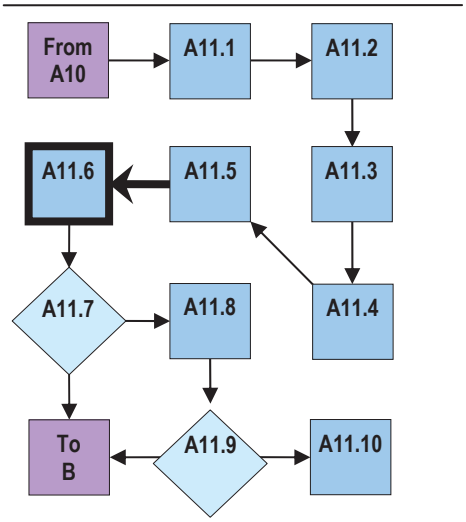
A11.4
Description The Archive determines a schedule for transferring the records in an Ingest Project to the Archive and attaches that schedule to the Submission Agreement.
Uses Transfer Procedures
Produces/Modifies Transfer Schedule



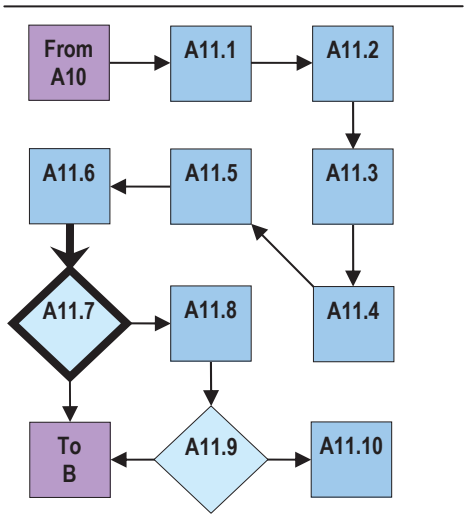
A11.5
Description The Archive attaches a description of its Submission Information Package (SIP) Creation Procedures to the Submission Agreement.
Uses SIP Creation Procedures
Produces/Modifies SIP Creation Procedures Decision



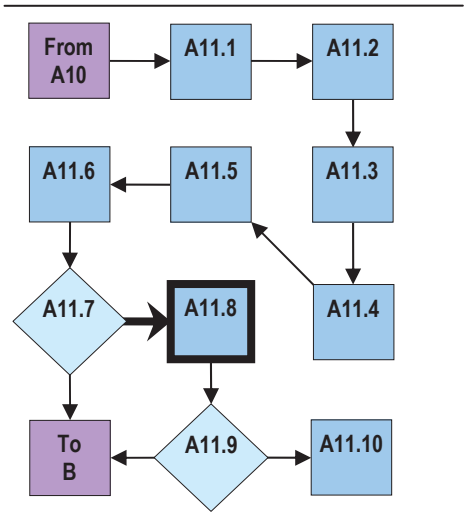
A11.6
Description The Archive takes actions needed to finalize the Submission Agreement that do not substantially change any decisions made earlier in Section A. The Archive constructs an actual draft Submission Agreement, typically using preset language.
Uses None
Produces/Modifies Draft Submission Agreement



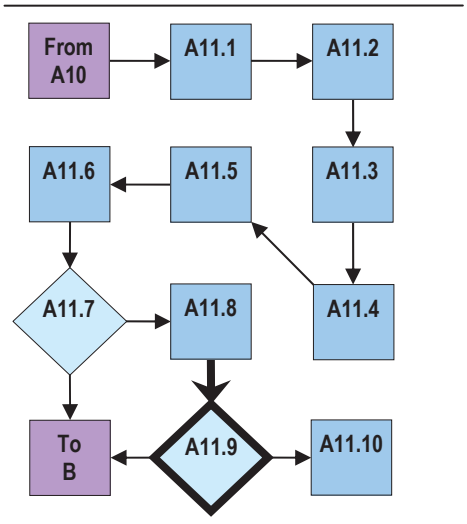
A11.7
Description The Archive and the Producer review the drafted Submission Agreement and determine if they are willing to endorse the Submission Agreement. If both are willing, the Archive and Producer endorse the Submission Agreement. This completes Section A: Negotiate Submission Agreement of the Ingest Guide. Proceed to Section B: Transfer and Validation to complete the Ingest Project.
Uses Draft Submission Agreement
Produces/Modifies Finalized and Endorsed Submission Agreement



A11.8
Description If either the Archive or the Producer is unwilling to endorse the Submission Agreement as it is drafted, the Archive, usually in conjunction with the Producer, will revise the Submission Agreement to address the Archive's or Producer's concerns.
Uses Draft Submission Agreement
Produces/Modifies Draft Submission Agreement



A11.9
Description The Producer and the Archive determine if they are willing to endorse the Submission Agreement with the revisions made in Step A11.8. If they are willing to endorse the Submission Agreement, they will do so. This completes Section A: Negotiate Submission Agreement of the Ingest Guide. Proceed to Section B: Transfer and Validation to complete the Ingest Project.
Uses Draft Submission Agreement
Produces/Modifies Finalized and Endorsed Submission Agreement

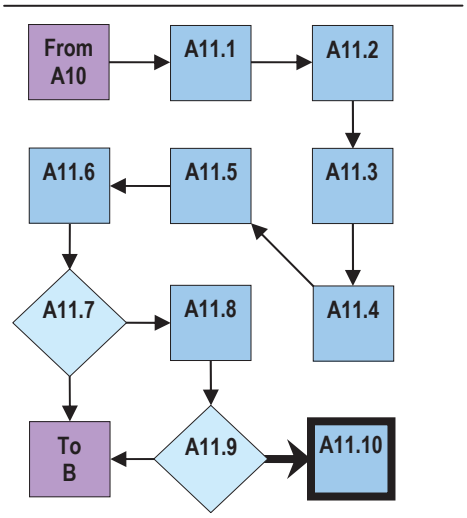


A11.10

Description If either the Producer or the Archive is still not willing to sign the revised submission agreement in Step A11.8, return to Step A11.8 to revise the Agreement again. Repeat this process until the Producer and the Archive agree to sign the Submission Agreement. If the Producer and the Archive cannot agree on a finalized version of a Submission Agreement, the Archive will stop the Ingest Project.

Uses Draft Submission Agreement

Produces/Modifies Ingest Project Termination Notice



SECTION B: TRANSFER AND VALIDATION

Overview

The Transfer and Validation section of the Ingest Guide describes the actions needed for the Archive and a Producer to deposit records into a preservation system. This portion of the Ingest Guide describes the actual transfer, validation, and transformation work of an Ingest Project. The steps for transfer, validation, and transformation are defined by Submission Agreements created in Step A.

This section is composed of six parts. During Part B1 the Producer packages the appropriate records in a Submission Information Package (SIP) with the proper metadata, as stipulated in the Submission Agreement. Once the Archive has received the SIP during Part B2, Automated Validation, it must verify the integrity, completeness, and correctness of the transfer and validate that the transferred records conform to the requirements of the Submission Agreement and to technical file format standards. This includes validating the SIP against the requirements of the Submission Agreement. Then in Part B3, Transform and Attach Metadata, the Archive transforms the records and attaches any needed metadata as prescribed by the Submission Agreement. In Part B4, AIP Formation, the Archive assembles the records involved in the Ingest Project into Archival Information Packages (AIPs). In Part B5, Final Appraisal, the Archive makes a final check of the records and the metadata in the AIP to ensure that they conform to the rules of the Archive and are indeed the records described in the Submission Agreement. Finally in Part B6, Formal Accession, the Archive formally accessions the records into the Preservation System and notifies the Producer of this formal accession.

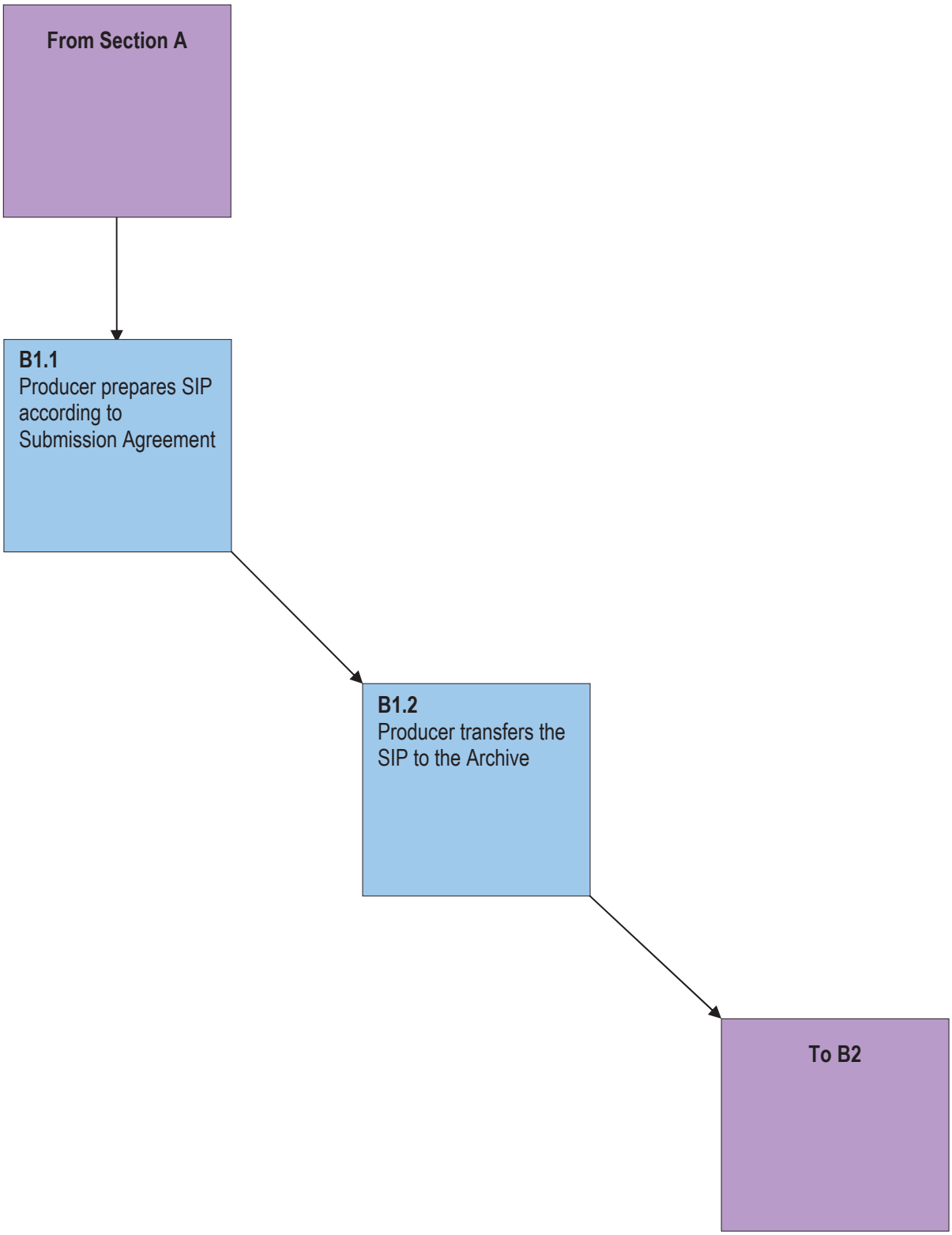
Part B6 is the final step of the Ingest Guide and is the last step of an Ingest Project.

The Ingest Guide uses the *OAIS* definition of SIP: "An Information Package that is delivered by the Producer to the OAIS for use in the construction of one or more AIPs."[7] It also uses the *OAIS* definition of AIP: "An Information Package, consisting of the Content Information and the associated Preservation Description Information (PDI), which is preserved within an OAIS."

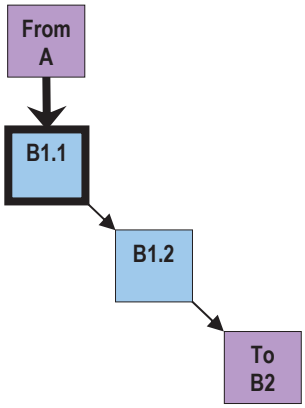
[7] ISO 14721:2003

SECTION B: TRANSFER AND VALIDATION
Part B1: Create and Transfer SIPs

Overview
During this Part a Producer creates a Submission Information Package (SIP) of the records the Archive will accession according to the terms of a Submission Agreement. A Producer will then transfer the SIP to the Archive.

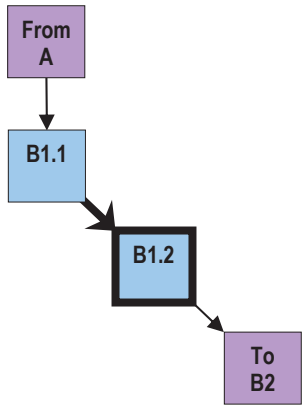

From Section A
B1.1
Producer prepares SIP according to Submission Agreement
B1.2
Producer transfers the SIP to the Archive
To B2

B1.1



Description The Producer prepares the records that it has agreed to transfer in the Submission Agreement by packaging them in a Submission Information Package (SIP) according to the SIP Creation Procedures articulated in the Submission Agreement. The SIP includes any necessary metadata, digital signatures, or Producer-side transformations that the Submission Agreement called for. Creation of the SIP may be technically complex, depending on the number of digital components and the number and complexity of file formats of those components. The Producer may or may not be able to accomplish this task on its own and may require technical assistance from the Archive. The Archive may wish to build a tool set to enable Producers to produce SIPs.
Uses SIP Creation Procedures Decision, SIP Creation Procedure
Produces/Modifies SIP

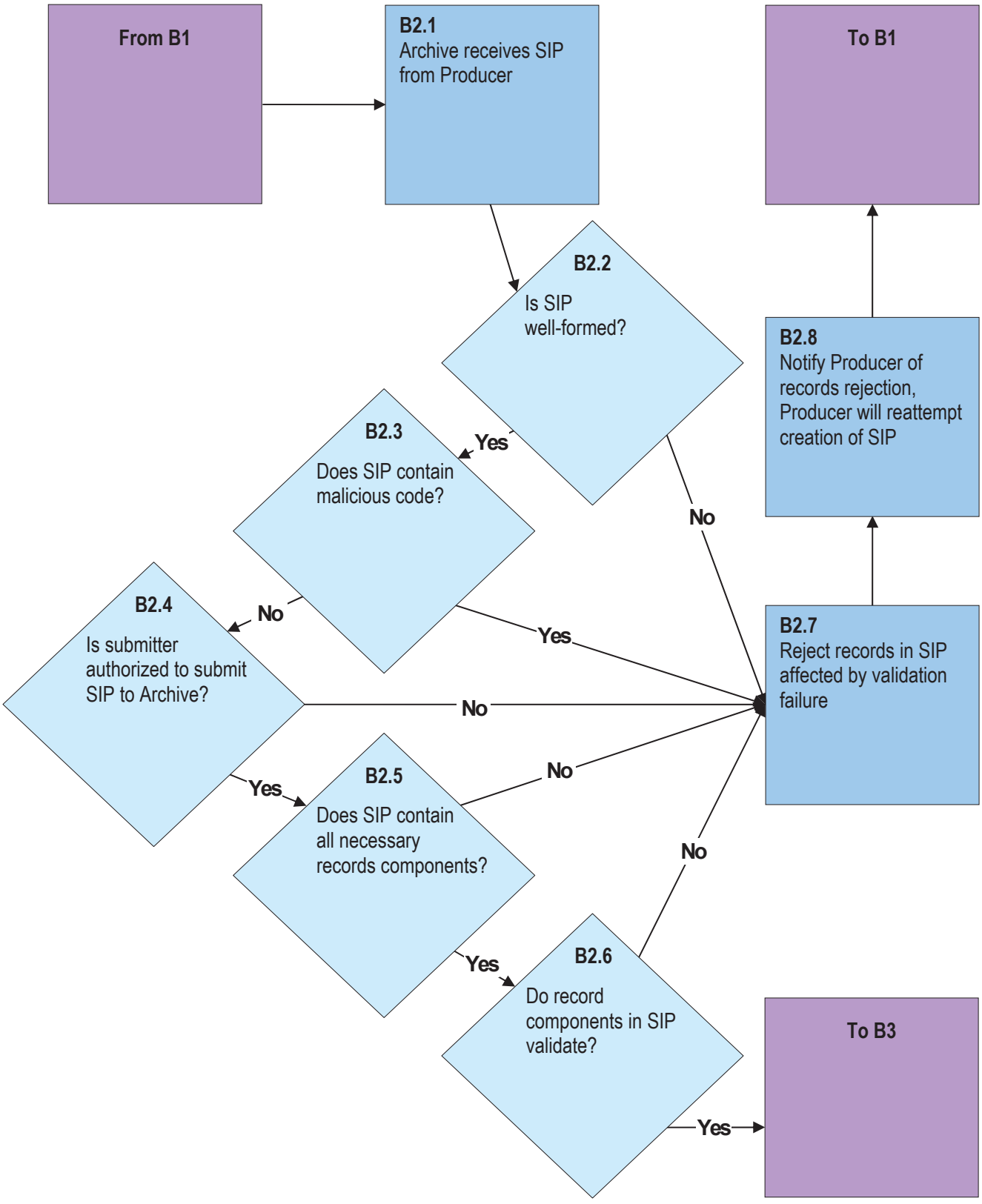
B1.2



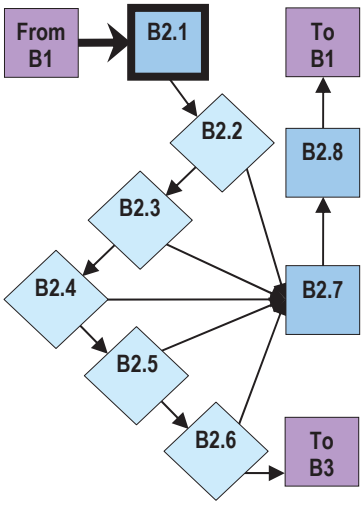
Description The Producer transfers the SIP to the Archive. This step represents the actual transfer moment in the ingest process. The transfer may be a physical exchange of storage media containing electronic records or a transfer undertaken over a computer network. None of the steps involved in completing either of these processes is described here.
Uses Transfer Schedule
Produces/Modifies None

SECTION B: TRANSFER AND VALIDATION
Part B2: Validate

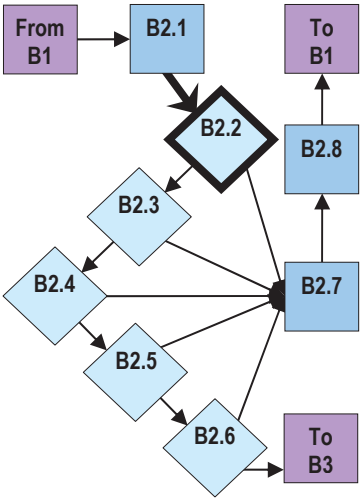
Overview
During this Part, the Archive validates the SIP and its included record components received from the Producer. The Archive checks that the SIP and its components are well-formed and whether they contain viruses. It also validates that the Producer was authorized to transfer the SIP and that the SIP conforms to the requirements of the Submission Agreement. If the SIP fails any of these validations, the Archive rejects the SIP and notifies the Producer to generate another SIP. An Archive can carry out all of the steps in this Part in an automated manner. While an Archive does not have to automate these Steps, automation greatly enhances productivity. In practice, an Archive may implement Steps B2.2 through B2.6 in a difference sequence.


From B1
B2.1
Archive receives SIP from Producer
To B1
B2.2
Is SIP well-formed?
B2.8
Notify Producer of records rejection, Producer will reattempt creation of SIP
B2.3
Does SIP contain malicious code?
Yes
No
B2.4
Is submitter authorized to submit SIP to Archive?
No
Yes
B2.7
Reject records in SIP affected by validation failure
No
B2.5
Does SIP contain all necessary records components?
Yes
No
No
B2.6
Do record components in SIP validate?
Yes
To B3

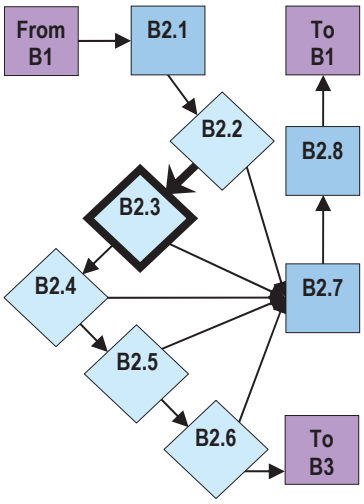
B2.1
Description The Archive receives the SIP from a Producer. The Archive may notify a Producer that it has received the SIP as long as the Archive makes it clear receiving the SIP does not mean accepting it.
Uses Institutional Identity Management System, Producer Record
Produces/Modifies Documentation of Receipt



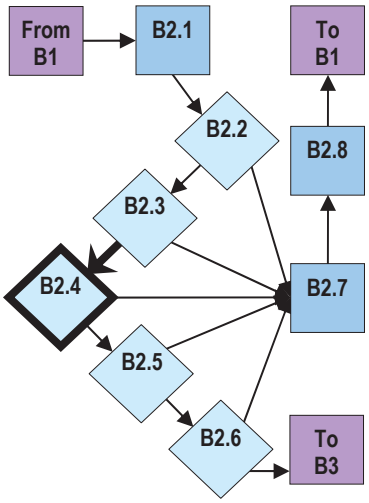
B2.2
Description This is the first of five validation steps that the Archive performs on a SIP. In this step, the Archive performs a format check on the SIP package to ensure that it meets the requirements of the Preservation System, as described in the Submission Agreement. This includes confirming the existence of all the records components listed in the SIP manifest.
Uses SIP Creation Procedures
Produces/Modifies SIP Validity Statement



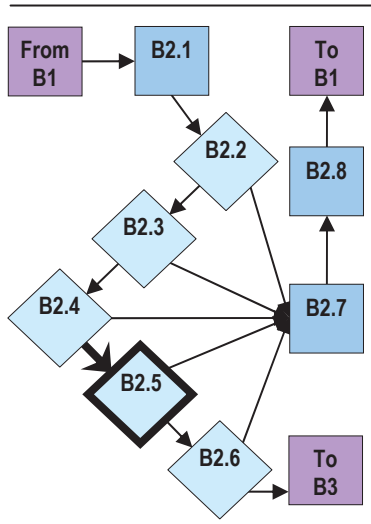
B2.3
Description The Archive checks if a SIP contains any malicious code (viruses). Because of the age of the digital components contained in the SIP, the Archive must ensure that malicious code check can recognize very old malicious code.
Uses SIP Creation Procedures, Virus Definition Files
Produces/Modifies SIP Validity Statement



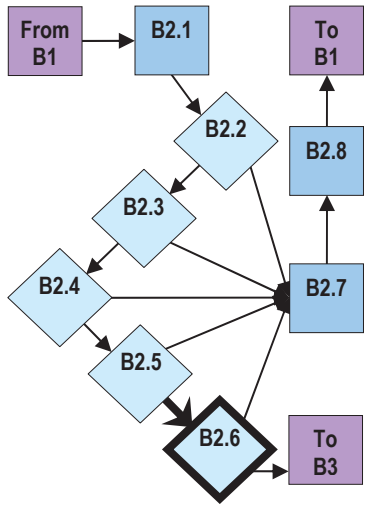
B2.4
Description The Archive checks if the submitter of a SIP is authorized to send the SIP to the Archive.
Uses Producer Entry
Produces/Modifies SIP Validity Statement



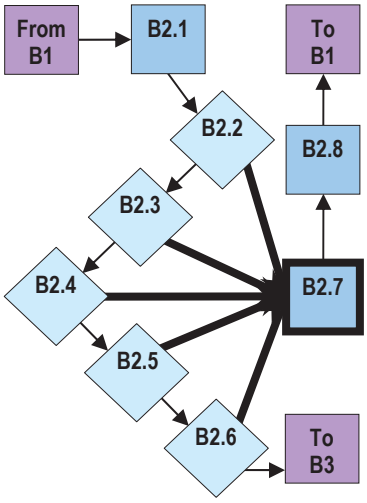
B2.5
Description The Archive checks if the records components in the SIP or SIPs of an Ingest Project properly form the records described in a Submission Agreement.
Uses Transformation Plan, Record Type Record
Produces/Modifies SIP Validity Statement



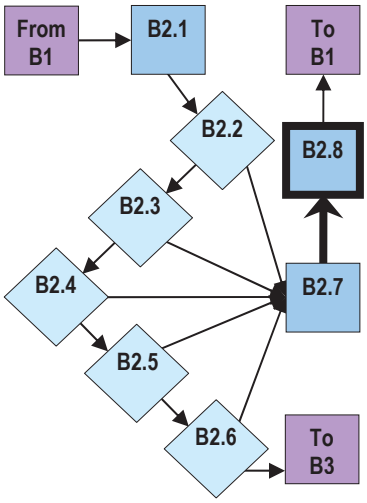
B2.6
Description The Archive validates that the file formats of the records components contained within a SIP conform to technical file format standards.
Uses Transformation Plan, Record Type Record, Format Representation Information System
Produces/Modifies SIP Validity Statement



B2.7
Description If the Archive finds any problems with a SIP or its components in Steps B2.2 through B2.6, the Archive rejects all of the records in the SIP affected by the validation failure.
Uses None
Produces/Modifies None

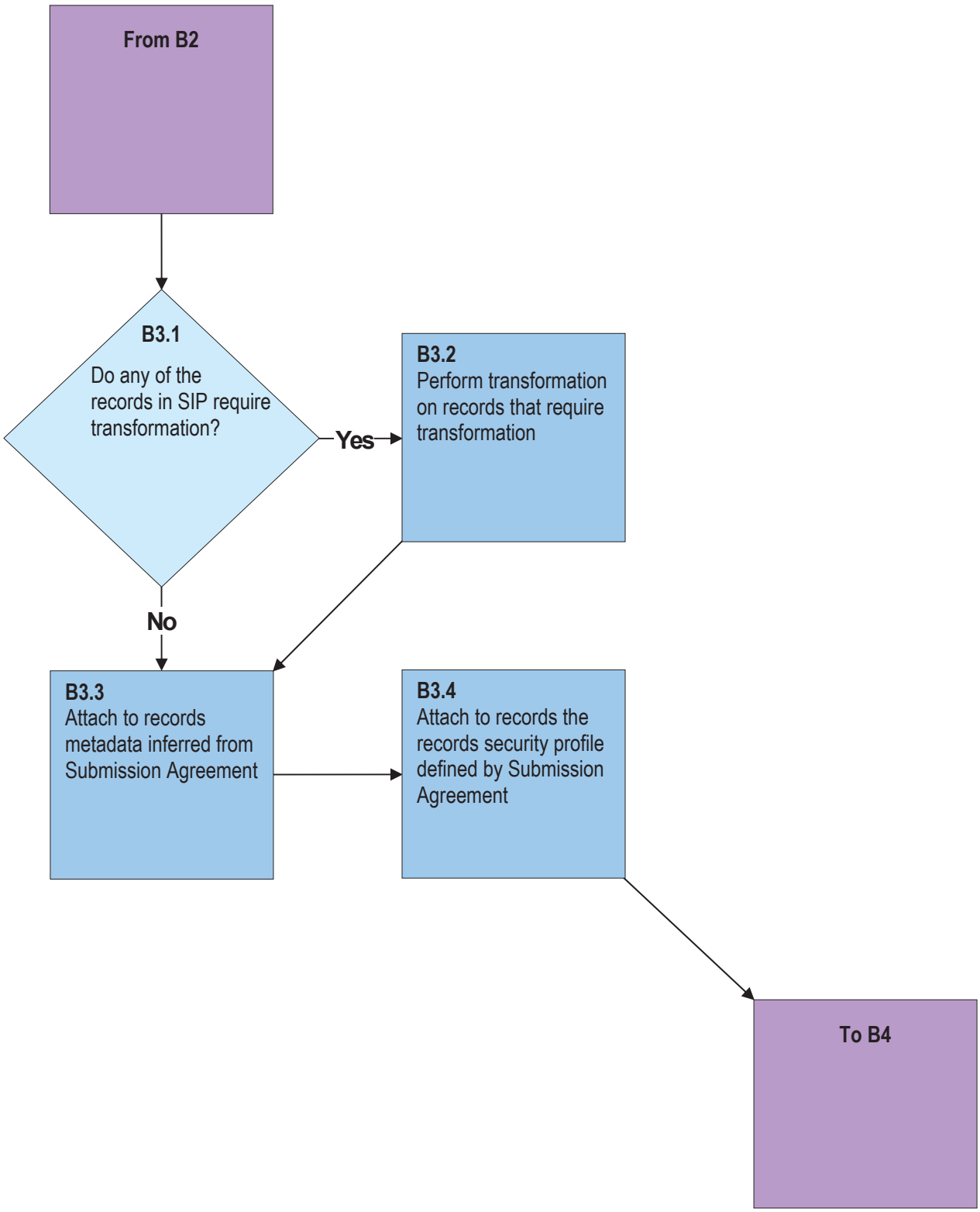


B2.8
Description The Archive notifies a Producer that it has rejected at least some records in a SIP. The Producer will correct the error and repeat Step B1.1, creating another SIP of those records for resubmission.
Uses None
Produces/Modifies SIP Rejection Notification

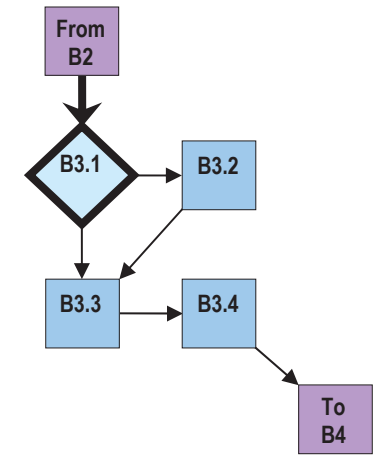


SECTION B: TRANSFER AND VALIDATION
Part B3: Transform and Attach Metadata

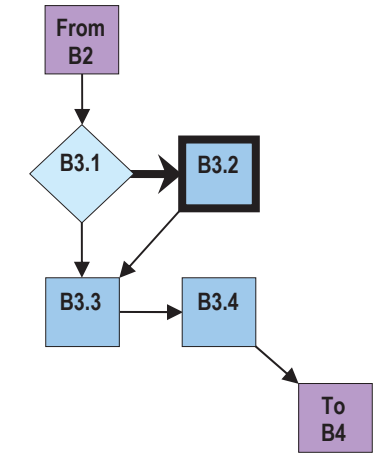
Overview
During this Part, the Archive transforms any records in an Ingest Project that require transformation. An Archive also attaches to the records metadata it can automatically infer from the Submission Agreement and its associated documentation. This metadata includes unique identifiers, Records Security Profiles, and information about the Producer, Record Types, formats, and time of transfer. This metadata should allow the Archive to administer the records. The Archive can add descriptive metadata to records after it has completed the Ingest Project.


From B2
B3.1
Do any of the records in SIP require transformation?
Yes
No
B3.2
Perform transformation on records that require transformation
B3.3
Attach to records metadata inferred from Submission Agreement
B3.4
Attach to records the records security profile defined by Submission Agreement
To B4

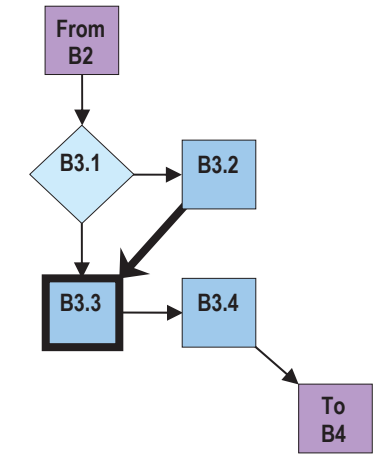
B3.1
Description The Archive determines if any of the records in an Ingest Project requires format transformation. The Archive determines this based on the findings it made in Part A5.
Uses Transformation Plan
Produces/Modifies None



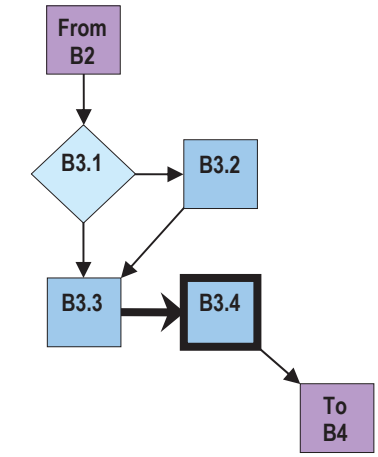
B3.2
Description If the Archive needs to transform any of the records in an Ingest Project, it will transform those records as determined in Part A5.
Uses Transformation Plan
Produces/Modifies Records Transformed



B3.3
Description The Archive attaches metadata it infers from the Submission Agreement to the records. This is an automated application of the metadata whose encoding rules are stipulated in Step A11.1.
Uses Metadata Encoding Rules
Produces/Modifies Records with Attached Metadata

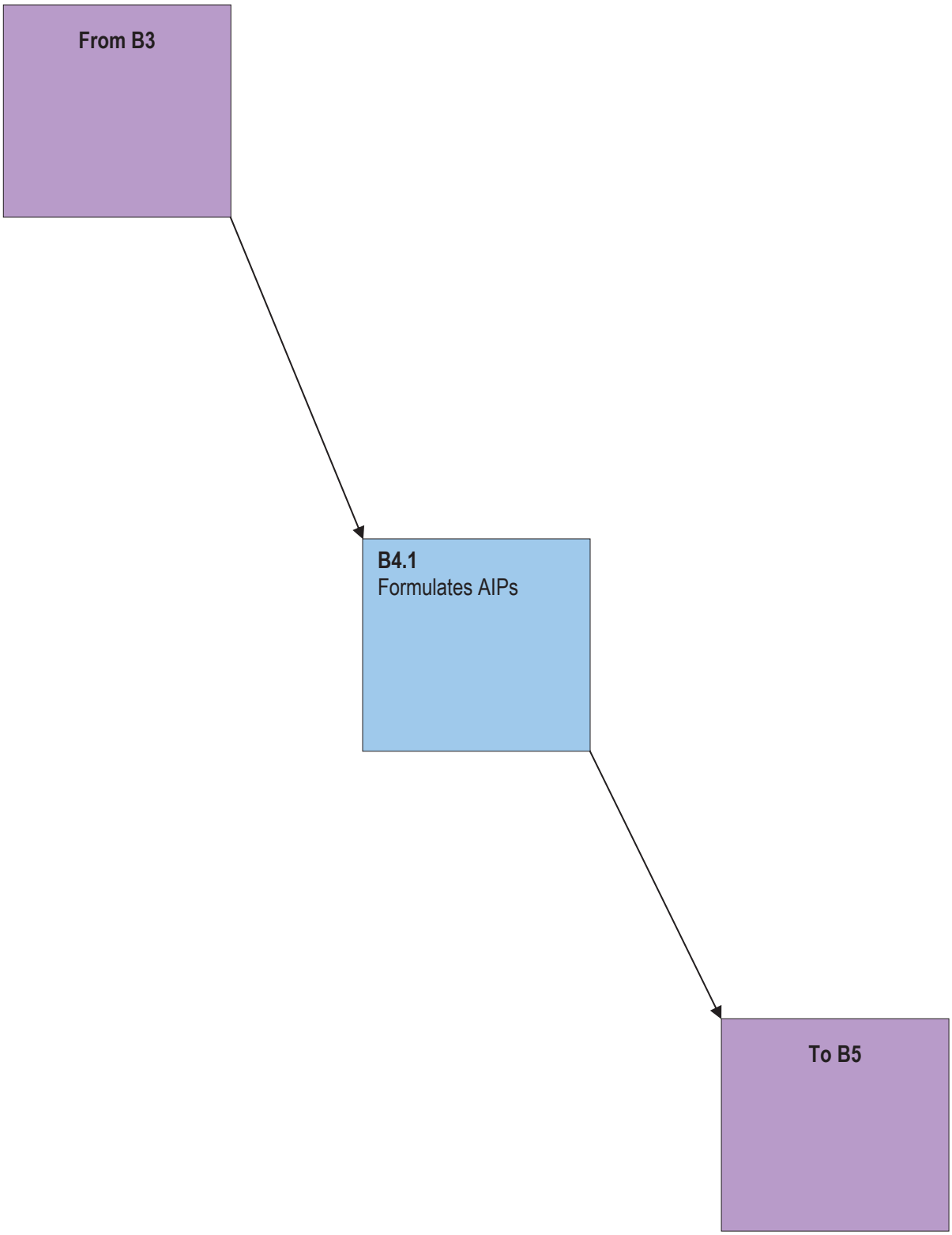


B3.4
Description The Archive attaches the proper records security profile to the records as defined in the Submission Agreement (Part A8).
Uses Records Security Profile
Produces/Modifies Records with Security Profile

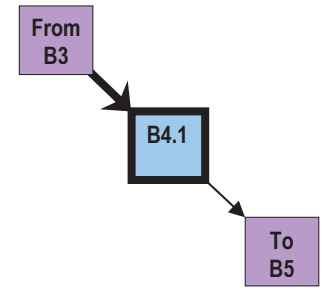


SECTION B: TRANSFER AND VALIDATION
Part B4: Formulate AIPs

Overview
During this Part, the Archive turns the records involved in an Ingest Project into Archival Information Packages (AIP) according to the rules and procedures of the Archive's Preservation System.

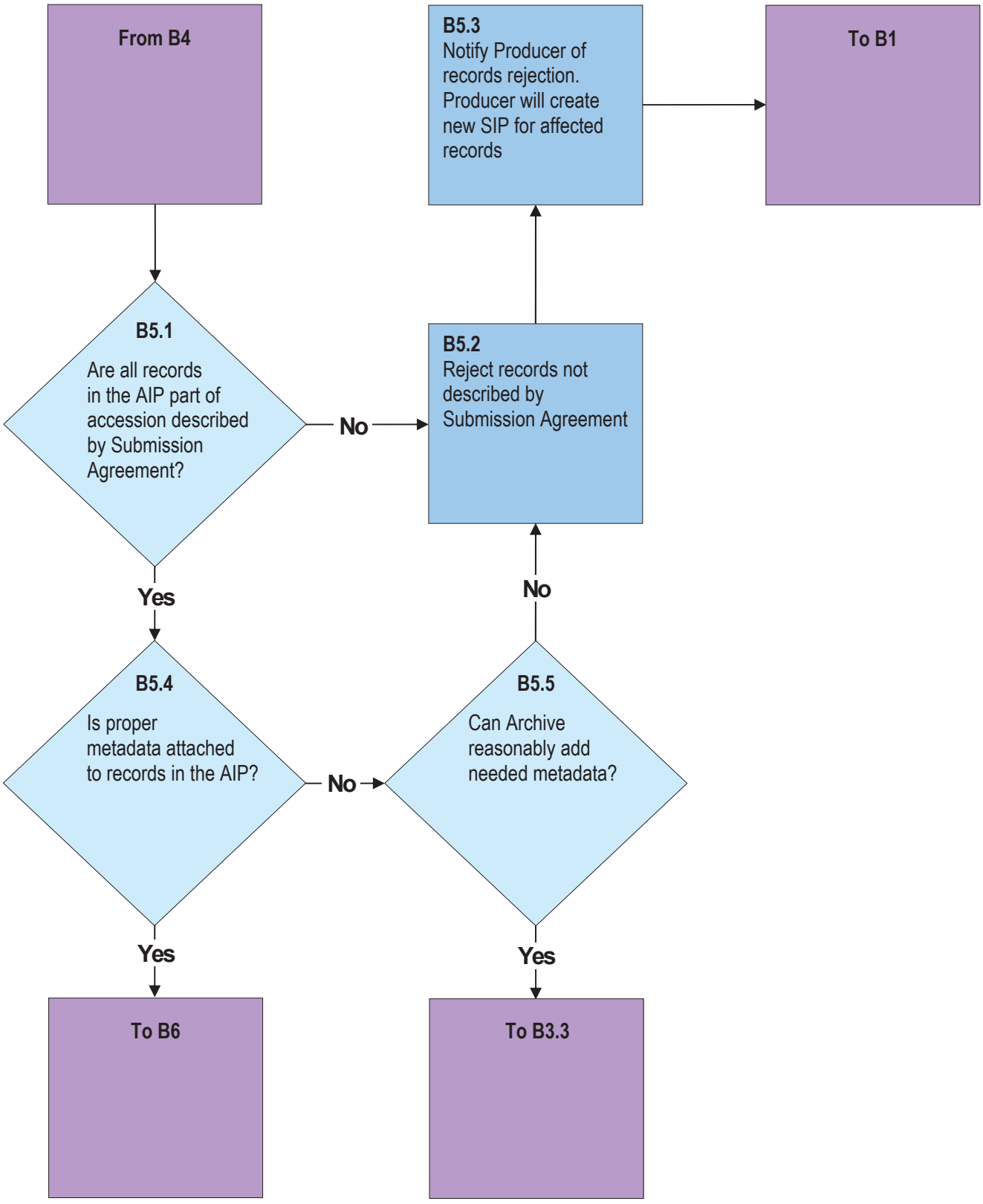

From B3
B4.1
Formulates AIPs
To B5

B4.1
Description The Archive turns the records involved in an Ingest Project into AIPs.
Uses AIP Configuration Rules
Produces/Modifies AIP

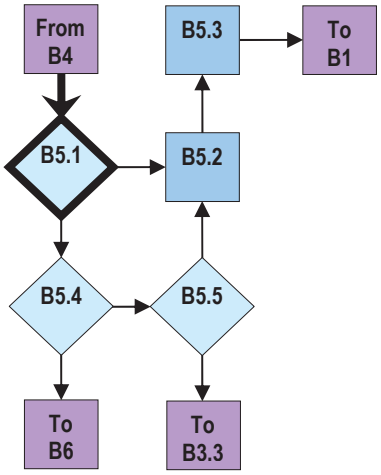


SECTION B: TRANSFER AND VALIDATION
Part B5: Assess AIPs

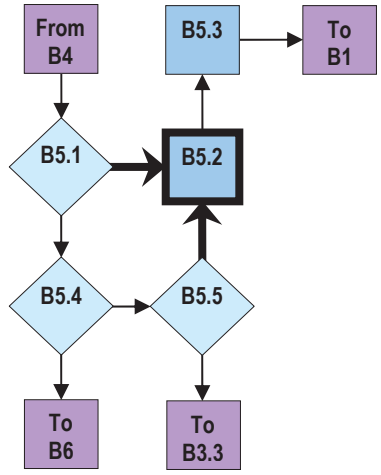
Overview
During this Part, the Archive conducts a final appraisal of the records involved in the Ingest Project. It ensures that the records in the newly formed AIPs are the records described in the Submission Agreement and that they have the proper metadata associated with them. If the AIP does not contain the correct records, the Archive rejects the records and notifies the Producer to generate a new SIP for the affected records. If the records in the AIP do not have the proper metadata, the Archive determines if they can be added with a reasonable amount of effort. If this effort is too great, the Archive rejects the affected records. Usually, the Archive checks a sample of records involved in the Ingest Project. The appropriate rate of sampling depends on the circumstances of the Archive and individual Ingest Projects.


From B4
B5.3
Notify Producer of records rejection. Producer will create new SIP for affected records
To B1
B5.1
Are all records in the AIP part of accession described by Submission Agreement?
No
B5.2
Reject records not described by Submission Agreement
Yes
No
B5.4
Is proper metadata attached to records in the AIP?
No
B5.5
Can Archive reasonably add needed metadata?
Yes
Yes
To B6
To B3.3

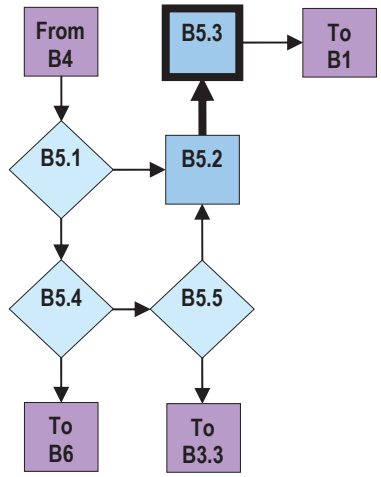
B5.1
Description The Archive manually checks the records in an AIP to determine if they are in fact the records described in the Submission Agreement.
Uses Survey Report
Produces/Modifies AIP Validity Statement



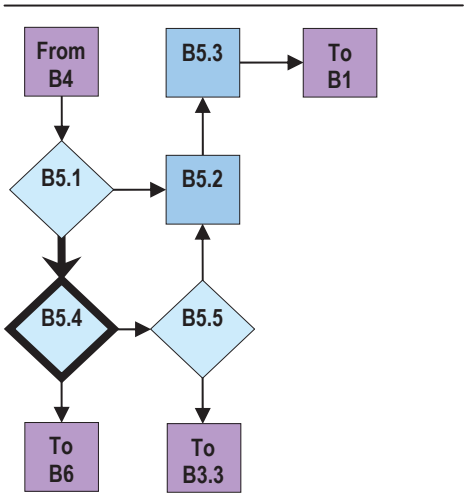
B5.2
Description If the records in an AIP do not match the records listed and described in a Submission Agreement, or lacks the necessary metadata the Archive cannot reasonably add on its own, the Archive rejects the affected records in the AIP.
Uses None
Produces/Modifies None



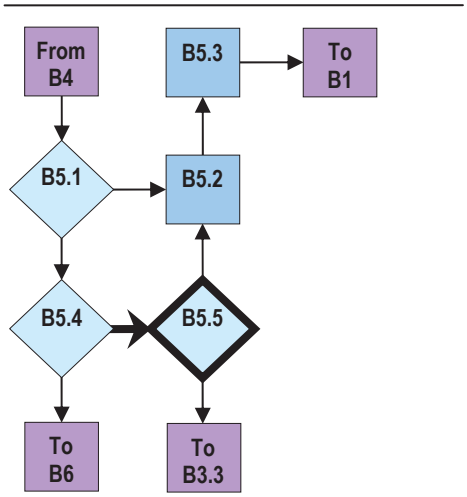
B5.3
Description The Archive notifies a Producer that it has rejected records in an AIP. The Producer corrects the error and repeats Step B1.1, creating a new SIP for the affected records.
Uses None
Produces/Modifies AIP Rejection Notice



B5.4
Description The Archive manually checks the records in an AIP to determine if they have the properly associated metadata.
Uses Metadata Encoding Rules
Produces/Modifies AIP Validity Statement

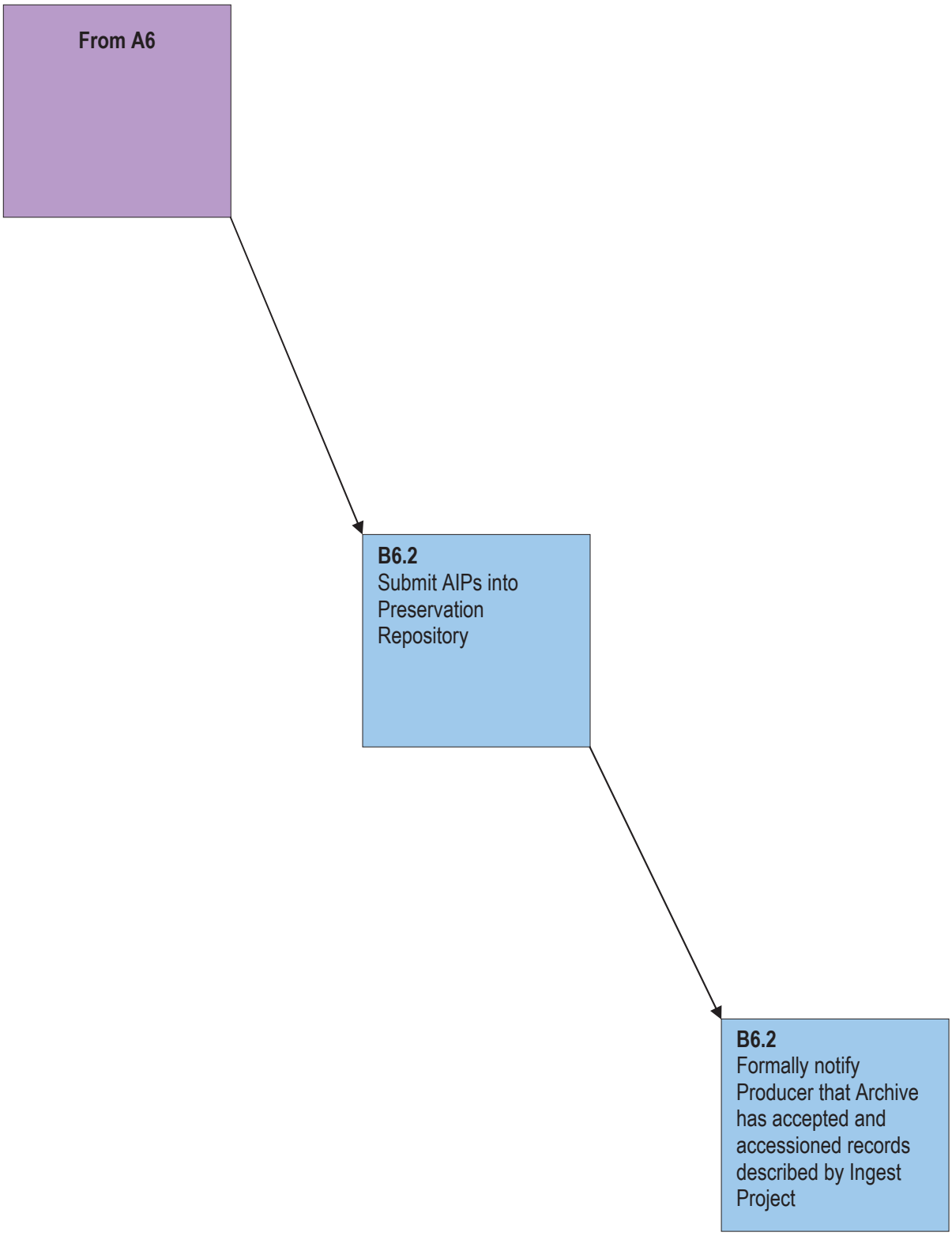


B5.5
Description If at least some of the records in an AIP do not have the proper metadata attached, the Archive determines if it can reasonably attach the proper metadata. If it can, it goes back to Step B3.3 to attach the proper metadata. If it cannot reasonably attach the proper metadata to the records in the AIP, the Archive moves to Step B5.2 and rejects the affected records.
Uses Metadata Encoding Rules
Produces/Modifies None

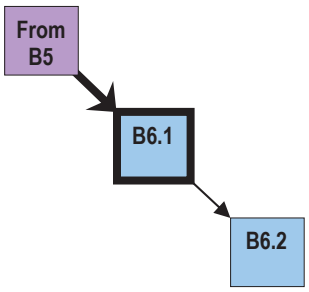


SECTION B: TRANSFER AND VALIDATION
Part B6: Formally Accession

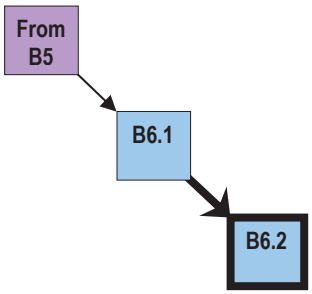
Overview
During this part, the Archive deposits the AIPs it has formulated during an Ingest Project into its Preservation System. Then it formally notifies the Producer that it has accepted and accessioned the records the Producer transferred to the Archive in a SIP or SIPs. This is the moment of formal transfer from the Producer to the Archive.


From A6
B6.2
Submit AIPs into Preservation Repository
B6.2
Formally notify Producer that Archive has accepted and accessioned records described by Ingest Project

B6.1
Description The Archive deposits the AIP(s) to its Preservation System according to its Preservation System rules.
Uses None
Produces/Modifies Preservation System-Managed AIP



B6.2
Description The Archive formally acknowledges that it has accepted and accessioned the records involved in the Ingest Project. This is the moment of formal transfer of the records in a SIP from the Producer to the Archive.
Uses None
Produces/Modifies Transfer Notice, Accession Log



SUBMISSION AGREEMENT

Overview

A Submission Agreement defines the nature and scope of the records involved in an Ingest Project and delineates the manner in which the Archive and the Producer execute the transfer, validation, and transformation of these records. In addition to guiding the work of the Producer and the Archive for transfer and transformation and serving as the benchmark for validation, it also provides both entities a document describing the terms of an Ingest Project that they can endorse and agree to. The Submission Agreement can cover a single Ingest Project or serial Projects.

The Submission Agreement documents the information needed to establish the terms of the scope, transfer, validation, and transformation of an Ingest Project. The Archive documents this as Components in the Submission Agreement. Most of these Components—and the decisions they represent—are tied to standing Resources. For example one of the elements of a Submission Agreement identifies the format types of the records in an Ingest Project. This Component in the Submission Agreement references a Formal Representation Information System on the format types the Archive employs as preservation formats in its Preservation System. Resources are usually policies, procedures, metadata records, or logs of action. While these Resources have an impact on nearly all of the Archive's Ingest Projects, they are not specific to any single Ingest Project.

Ideally, the Archive creates machine-readable and human-readable versions of its Submission Agreements. A human-readable version gives the Producer and the Archive a document both can agree to and endorse. A machine-readable version enables a degree of automated validation and transformation, usually coordinating calls to sets of other machine-readable code that dictate validation and transformation activities. This automation should help the Archive make the size and number of its Ingest Projects scalable. The degree of automation depends largely on the amount of detail an Archive's Resources has. For example, if the Archive's Format Representation Information System contains detailed, machine-readable, technical, and administrative metadata about each Format, it can use that metadata to automatically validate and transform the formats of records during Ingest. If the Archive's Format Representation Information System is a simple paper list and brief narrative description of formats, the Policy will not help the Archive automate the validation or normalization of records.

Some Ingest Projects—usually those with new types of records, formats, creators, or special circumstances—prompt the Archive to create a new version or add to one or more of its Resources. For example, if the Archive decides to preserve a record in a format that is not one of its existing preservation formats, it will have to add that format to its Format Representation Information System, producing technical and administrative metadata about the new format. Although this Resource development demands time and effort, it allows the Archive to automate a greater variety of Ingest Projects in the long run. As the Archive adds detail and breadth to its Resources, it will be able automate a greater portion of its Ingest Process for a broader range of records.

The Survey Report, created in Parts A2 and A3, is a critical part of the Ingest process: it inventories the records that the Archive should accession in a particular Ingest Project. A Survey Report which a Submission Agreement references, can exist as a separate entity, or it can be embedded into the Submission Agreement.

See Appendix B for an example of a Submission Agreement.

COMPONENTS, RESOURCES, PRODUCTS, AND DOCUMENTATION

Overview

Below are descriptions of the Components, Resources, Products, and Documentation that the Archive uses to undertake Ingest Projects and create Submission Agreements. Each Component, Resource, Product, and Documentation includes a description, an explanation of its role in the Ingest process, and a listing of all the steps that use, produce, and modify it. They are listed alphabetically.

How these Resources and Components manifest themselves varies from Archive to Archive. Therefore, this section of the Ingest Guides describes each Resource and Component generally but does not prescribe their composition in detail. This section essentially highlights the need for the Archive to have these Resources and Components in some form, but does not extensively describe the manner of their existence. This section describes twelve Components, thirty Resources, eight Products, and twelve instances of Documentation.

Components

Expressions or selections of Resources the Archive documents in a Submission Agreement. The Components document decisions the Archive makes for the scope, transfer, validation, and transformation of records in an Ingest Project.

Resources

The Archive's standing policies, procedures, metadata records, or logs of action. Resources dictate the actions the Archive undertakes during its Ingest Projects. Resources are not specific to any single Ingest Project. The Submission Agreement references Resources through Components.

Products

Objects created or modified as a result of the work the Archive and the Producer undertake during an Ingest Project.

Documentation

Expressions of decisions the Archive makes during an Ingest Project that the Archive does not need in its Submission Agreement to undertake transfer, validation, and transformation but does need to document all of its Ingest decisions.

Access Controls Gap Analysis
Description An analysis of the security capabilities of the Preservation System in comparison to the access control needs of the records in an Ingest Project. If the security capabilities fall short of the records' access control needs, the Archive measures this gap and describes it in the Gap Analysis.
Ingest Project Role Documentation
Steps that Use A8.4, A8.5, A8.7
Steps that Produce/Modify A8.3

Access Controls Gap Analysis Feasibility Statement
Description A statement that declares whether the Preservation System is or is not capable of meeting the access control needs of the records in the Ingest Project.
Ingest Project Role Documentation
Steps that Use A8.7
Steps that Produce/Modify A8.4

Access Controls Policy
Description An articulation of an Archive's policy on access to records stored in the Preservation System. This includes security measures an Archive takes to prevent unauthorized access to records. This also usually includes such access restriction profiles as "open access," "restricted, administrative records," and "restricted, personal records" which declare who can gain access to the records, under what circumstances, and when.
Ingest Project Role Resource
Steps that Use A8.1, A8.2, A8.3, A8.7
Steps that Produce/Modify A8.7

Accession Log
Description A record of the accessions an Archive has made. At a minimum, an accession entry should record a basic description or identification of the accessioned records, the date of transfer, and the Producer that transferred the records to the Archive.
Ingest Project Role Resource
Steps that Use A1.3
Steps that Produce/Modify B6.2

Activity Log
Description A record of the work an Archive has done with or for a Producer. Entries in an Activity Log may include accessions, surveys, consultations, or any other type of activities the Archive may engage in. Because Activity Logs vary so greatly among Archives, an Archive may create an entry at any point during an Ingest Project depending on what activities it wants to document. An Archive should only keep an Activity Log to the extent that it wants to document its activities in this form.
Ingest Project Role Resource
Steps that Use A1.1, A1.3, A2.3
Steps that Produce/Modify Varies

Archival Information Package (AIP)
Description An Archival Information Package, commonly referred to as an "AIP," is the form a record takes when it is managed in a preservation repository. *OAIS* defines an AIP as "An Information Package, consisting of the Content Information and the associated Preservation Description Information (PDI), which is preserved within an OAIS [Preservation System]."[8]
Ingest Project Role Product
Steps that Use None
Steps that Produce/Modify B4.1

Archival Information Package Configuration Rules
Description These rules articulate how an Archive needs to assemble its Archival Information Packages (AIPs) so it can successfully submit them to its preservation repository in order to manage and preserve them over time.
Ingest Project Role Resource
Steps that Use B4.1
Steps that Produce/Modify None

Archival Information Package Rejection Notice
Description The notification an Archive sends to a Producer that it has rejected that Producer's Archival Information Package (AIP). The Notification should describe why and when the Archive terminated the AIP. An Archive also retains a copy of the Notice for its collection management system.
Ingest Project Role Documentation
Steps that Use None
Steps that Produce/Modify B5.3

Archival Information Package Validity Statement
Description The statement an Archive makes declaring the validity of an AIP.
Ingest Project Role Documentation
Steps that Use None
Steps that Produce/Modify B5.1, B5.4

Archive Naming/Identification Scheme
Description This scheme defines how an Archive names and/or identifies the records it holds in its Preservation System. An Archive may support multiple Naming/Identification schemes, although many Archives only support one such scheme. If an Archive decides to use a Producer Naming/Identification Scheme to name or identify records in its Preservation System, the Archive will adapt that Producer scheme as an Archive Naming/Identification Scheme in Step A6.3.
Ingest Project Role Resource
Steps that Use A6.2, A6.4
Steps that Produce/Modify A6.3

[8] ISO 14721:2003, p. 1-7.

Archive Naming/Identification Scheme Decision
Description This documents the decision an Archive makes regarding the naming or identification scheme it applies to the records during an Ingest Project.
Ingest Project Role Component
Steps that Use None
Steps that Produce/Modify A6.4

Archives Directory
Description This directory lists archives, their contact information, and a brief summary of their collecting policies. An archive may use the directory to help direct a Producer to an appropriate Archive for its records.
Ingest Project Role Resource
Steps that Use A1.5
Steps that Produce/Modify None

Collection Policy
Description In addition to defining what type of records an Archive collects, a Collection Policy also identifies the Producers from whom an Archive will collect records. Usually, this covers Producers not addressed in a Records Authority Statement, although a Records Authority Statement and a Collection Policy may overlap.
Ingest Project Role Resource
Steps that Use A1.4, A3.3, A3.4
Steps that Produce/Modify None

Copyright Policy
Description The articulation of how an Archive manages the reproduction of records in light of their Copyright Status.
Ingest Project Role Resource
Steps that Use A7.2, A7.3, A7.4
Steps that Produce/Modify None

Copyright Status
Description An indication of the copyright status of the records and associated software in an Ingest Project, usually indicating the copyright holder and any applicable licensing agreement the Archive has with the copyright holder. This may include a description of the records' and associated software's moral rights.
Ingest Project Role Component
Steps that Use A7.2
Steps that Produce/Modify A7.1

Copyright Transfer/License
Description An agreement between an Archive and a copyright holder that is not the Producer in an Ingest Project. It either transfers copyright ownership of the records in an Ingest Project from the original copyright holder to the Archive, or it licenses to the Archive the right to reproduce the records under a set of conditions defined by the copyright holder.

Ingest Project Role Resource
Steps that Use None
Steps that Produce/Modify A7.8

Designated Community Description
Description A description of "an identified group of potential Consumers who should be able to understand a particular set of information. The Designated Community may be composed of multiple user communities."[9] Usually the description of a community's ability to understand a set of information focuses on the community's technical capability to functionally use the formats of a set of records.
Ingest Project Role Resource
Steps that Use A5.3
Steps that Produce/Modify None

Documentation of Receipt
Description A notification sent from an Archive to a Producer that it has received Submission Information Package(s) (SIP(s)) from the Producer. The notification is not a declaration that the Archive has formally accepted the records contained in the SIP(s) it has received.
Ingest Project Role Documentation
Steps that Use None
Steps that Produce/Modify B2.1

Draft Submission Agreement
Description A completed Submission Agreement that neither the Producer nor the Archive has endorsed.
Ingest Project Role Product
Steps that Use A11.7, A11.8, A11.9, A11.10
Steps that Produce/Modify A11.6, A11.8

Finalized and Endorsed Submission Agreement
Description A completed Submission Agreement that both the Producer and the Archive have endorsed.
Ingest Project Role Product
Steps that Use None
Steps that Produce/Modify A11.7, A11.9

Format Representation Information System
Description: A repository of information about all formats used by the Preserver along with documentation of verification, validation, and rendering tools for each format. The system should be able to associate digital components with corresponding file format specifications and the tools to work with them. These systems may be homegrown, created and maintained externally (PRONOM, Global Digital File Format Registry), or a hybrid of local and external representation information. Some of this information might be stored in paper documentation, but for the system to be useful it needs to be electronic and integrated into the Ingest process.

[9] ISO 14721:2003, p. 1-10.

Ingest Project Role Resource
Steps the Use A5.1, A5.2, A5.3, B2.6
Steps the Produce/Modify None

Format Standards Policy
Description A Policy that declares the formats an Archive is able to process at Ingest, and the formats the Archive uses to preserve records in its Preservation System. A preservation format is a format that an Archive is capable of functionally preserving over the long term. Each time an Archive decides to employ a new preservation format, it needs to add that format to its Format Standards Policy.
Ingest Project Role Resource
Steps that Use A5.1
Steps that Produce/Modify A5.3, A5.4

Ingest Project Termination Notice
Description A notification an Archive sends to a Producer that it has ended the Ingest Project. The Notification should describe why and when the Archive ended the Ingest Project. An Archive also retains a copy of the Notice for its collection management system.
Ingest Project Role Documentation
Steps that Use None
Steps that Produce/Modify A1.5, A2.5, A3.5, A7.7, A8.6, A9.7, A10.5, A11.10

Institutional Identity Management System
Description An institution-wide identity management system that the Archive may use to help confirm the identity of a Producer. The system may manifest itself in a variety of ways, from a simple paper directory to a Lightweight Directory Access Protocol (LDAP)-based system.
Ingest Project Role Resource
Steps that Use A1.2, A1.6, A2.3, B2.1
Steps that Produce/Modify None

Metadata Encoding Rules
Description An articulation of the schemas for the metadata associated to the records in an Archive's Preservation System.
Ingest Project Role Resource
Steps that Use A11.1, B3.3, B5.4, B5.5
Steps that Produce/Modify None

Metadata Encoding Rules Decision
Description This articulates the metadata schema(s) an Archive decides to use as the encoding standard for the metadata it associates to the records during an Ingest Project.
Ingest Project Role Component
Steps that Use None
Steps that Produce/Modify A11.1

Preservation System Availability Statement
Description A statement of the institutional, financial, staffing, and technical capabilities of the Preservation System to manage and preserve the records during an Ingest Project.

Ingest Project Role Documentation
Steps that Use None
Steps that Produce/Modify A10.1, A10.2, A10.3

Preservation System Capabilities Report
Description A serial report that states the institutional, financial, staffing, and technical capabilities of the Preservation System.
Ingest Project Role Resource
Steps that Use A5.2, A8.2, A8.3, A10.1, A10.2
Steps that Produce/Modify None

Preservation System-Managed Archival Information Package
Description An AIP that is stored, maintained, and preserved in a Preservation System.
Ingest Project Role Product
Steps that Use None
Steps that Produce/Modify B6.1

Producer Entry
Description A reference to a specific Producer Record about a particular Producer.
Ingest Project Role Component
Steps that Use B2.4
Steps that Produce/Modify A1.3, A1.6

Producer Naming/Identification Scheme
Description A scheme that defines how a Producer names and/or identifies its records.
Ingest Project Role Documentation
Steps that Use A6.1, A6.2, A6.3
Steps that Produce/Modify None

Producer Record
Description A record that authoritatively identifies and describes a Producer. It also describes the relationship of a Producer to an Archive and its relationship with other Producers. Records of producers may describe individuals or departments, offices, or units. An Archive should have a record of every Producer that has transferred records to the Archive. This way, the Archive can associate all of its records in its holdings with a Producer. A Producer Record is usually encoded as machine-readable metadata. For example, a Producer Record may be written to ISAAR(CPF) (International Standard Archival Authority Record for Corporate Bodies, Persons, and Families, 2nd ed.) and encoded in EAC (Encoded Archival Context). The more details a Producer Record has, the more an Archive can utilize that record to automate its management and description of the records in its holdings. An Archive might populate a Producer Record with data from its Institutional Identity Management System.
Ingest Project Role Resource
Steps that Use A1.3, A2.2, A2.3, B2.1
Steps that Produce/Modify A1.6

Record Security Profile
Description A description of a class of security characteristics that an Archive assigns to the records in its Preservation System. A Profile articulates the access control needs of a record. Profiles are usually records-specific articulations of an Archive's Access Controls Policy.
Ingest Project Role Resource
Steps that Use A8.1, A8.2, A8.3, B3.4
Steps that Produce/Modify A8.1, A8.5

Record Security Profile Decision
Description A reference to a particular Records Security Profile.
Ingest Project Role Component
Steps that Use None
Steps that Produce/Modify A8.1

Record Type List
Description A listing of the Record Types of the records in an Ingest Project.
Ingest Project Role Component
Steps that Use None
Steps that Produce/Modify A4.1

Record Type Record
Description Documentation that describes the type of records that an Archive may accession into its Preservation System. A Record Type Record may describe various properties of a record type, including its general composition, function, confidentiality status, its Producers, and their management by Producers and Archives. The more details a Record Type Record has, the more an Archive can utilize it to automate its management and description of the records in its holdings.
Ingest Project Role Resource
Steps that Use A4.1, B2.5, B2.6
Steps that Produce/Modify A4.2

Recordkeeping System Evaluation Tool
Description A set of requirement delineating the attributes a recordkeeping system needs to be considered a trustworthy system. These requirements usually include requirements to facilitate the trustworthy transfer of records to a Preservation System.
Ingest Project Role Resource
Steps that Use A3.2, A9.2
Steps that Produce/Modify None

Recordkeeping System Internal Rules
Description A recordkeeping system's set of rules delineating how it must manage records.
Ingest Project Role Resource
Steps that Use A9.3
Steps that Produce/Modify None

Recordkeeping System Report
Description Records that identify and describe recordkeeping systems. These records should document an Archive's determination of a system's trustworthiness—its characteristics that allows a person to presume the authenticity of the records it manages—and its ability to support the scaleable, feasible, and trustworthy transfer of records to the Archive. An Archive should update the Report of a recordkeeping system often enough to accurately reflect its current configuration.
Ingest Project Role Resource
Steps that Use 3.2, A9.1
Steps that Produce/Modify 3.2, A9.2

Records Authority Statement
Description A statement that gives an Archive the authority to serve as the Archive for Producers. If applicable, it provides evidence that an Archive has the right to serve as the Archive for the records in the Ingest Project.
Ingest Project Role Resource
Steps that Use A1.4
Steps that Produce/Modify None

Records Retention Policy
Description An articulation of the disposition and period of retention of an institution's records. A Records Retention Policy may declare some records to have a disposition of permanent retention in the institution's Archive.
Ingest Project Role Resource
Steps that Use A3.3, A3.4
Steps that Produce/Modify None

Records Transformed
Description Records that an Archive has normalized into one its Preservation Formats.
Ingest Project Role Product
Steps that Use None
Steps that Produce/Modify B3.2

Records with Attached Metadata
Description Records to which an Archive has associated metadata during Part B3.
Ingest Project Role Product
Steps that Use None
Steps that Produce/Modify B3.3

Records with Security Profile
Description Records to which an Archive has associated security profile(s) during Part B3.
Ingest Project Role Product
Steps that Use None
Steps that Produce/Modify B3.4

Representation Information
Description The details needed to make the content of a record (defined by *OAIS* as a Content Data Object) understandable to a Designated Community.[10] Representation Information allows for the full interpretation of the content into meaningful concepts. Generally, Representation Information is the technical information needed to fill in the gap between the configuration of a record's format and the knowledge base of a Designated Community. Each time an Archive decides to employ a new preservation format, it will need to generate Representation Information for that preservation format in Step A4.3. An Archive may also have to produce new Representation Information in response to a change in a Designated Community's knowledge base.
Ingest Project Role Resource
Steps that Use None
Steps that Produce/Modify A5.3

Submission Information Package (SIP)
Description A Submission Information Package, commonly referred to as a "SIP," is created by a Producer to prepare records for transfer to an Archive. *OAIS* defines a SIP as "an Information Package that is delivered by the Producer to the OAIS [Preservation System] for use in the construction of one or more AIPs."[11]
Ingest Project Role Product
Steps that Use None
Steps that Produce/Modify B1.1

Submission Information Package Creation Procedures
Description A set of procedures that describes how a Producer should create a SIP.
Ingest Project Role Resource
Steps that Use A11.5, B1.1, B2.2, B2.3
Steps that Produce/Modify A9.4

Submission Information Package Creation Procedures Decision
Description An articulation of an Archive's decision to use a set SIP Creation Procedures in an Ingest Project.
Ingest Project Role Component
Steps that Use B1.1
Steps that Produce/Modify A11.5

Submission Information Package Rejection Notification
Description The notification an Archive sends to a Producer that it has rejected the Producer's SIP. The Notification should describe why and when the Archive rejected the SIP.
Ingest Project Role Documentation
Steps that Use None
Steps that Produce/Modify B2.8

[10] ISO 14721:2003, p. 1-13.
[11] ISO 14721:2003, p. 1-13.

Submission Information Package Validity Statement
Description A statement that an Archive makes that it has validated a SIP.
Ingest Project Role Documentation
Steps that Use None
Steps that Produce/Modify B2.2, B2.3, B2.4, B2.5, B2.6

Survey Instrument
Description A tool that an Archive uses to gather information about the records it surveys. This tool should be designed to help the Archive execute its Survey Procedures and help ensure that the Archive gathers the information it needs to capture during a survey. The Survey Instrument usually manifests itself as a form but can be any type of tool or set of tools that the Archive uses to gather information about records.
Ingest Project Role Resource
Steps that Use A3.1
Steps that Produce/Modify None

Survey Procedures
Description A plan or set of methods established to guide an Archive in undertaking a Records Survey. Usually these procedures consist of the Archive interviewing a Producer, the Archive querying the Producer through some sort of questionnaire, the Archive querying the electronic records themselves, or some combination of these methods.
Ingest Project Role Resource
Steps that Use A3.1
Steps that Produce/Modify None

Survey Report
Description A Report that identifies the records an Archive should accession during the Ingest Project. Survey Reports can vary greatly in detail, from a general description of the records the Archive should accession to an item-level inventory of those records. An Archive may create an early working draft of the Report in Step A2.1, after it and the Producer agree on the scope of the records that will be surveyed. In Steps A3.1 and A3.2, the Archive describes in the Survey Report the records it surveyed to the level of detail it requires. In Steps A3.3 and A3.4, the Archive documents in the Report its decisions on which, if any, records in the survey it should accession and what essential elements of these records it needs to preserve. To guide the Archive's appraisal decisions in Steps A3.3 and A3.4 and to be useful in Parts A3 through A10, the Survey Report needs to identify the records' Producer, Record Types, format type, file size, confidentiality requirements, copyright status, and any Producer-created identifiers.
Ingest Project Role Documentation
Steps that Use A2.1, A3.1, 3.3, A3.4, A4.1, A4.2, A5.1, A5.2, A5.5, A6.1, A6.4, A7.1, A7.6, A8.1, A8.5, A9.6, A10.1, A10.2, A10.3, A10.4, B5.1
Steps that Produce/Modify A3.1, A3.2, A3.3, A3.4, A7.5, A8.5, A9.5, A10.3

Transfer Notice
Description A notice an Archive creates and sends to a Producer declaring that it has received and accessioned the records of an Ingest Project.
Ingest Project Role Documentation
Steps that Use None

Steps that Produce/Modify B6.1

Transfer Procedures
Description A set of procedures that articulates how a Producer transfer records to an Archive.
Ingest Project Role Resource
Steps that Use A11.2, A11.4
Steps that Produce/Modify None

Transfer Procedures Decision
Description An articulation of an Archive's decision to use a set Transfer Procedures in an Ingest Project.
Ingest Project Role Component
Steps that Use None
Steps that Produce/Modify A11.2

Transfer Schedule
Description A schedule that articulates when a Producer transfers a SIP or set of SIPs to an Archive during an Ingest Project.
Ingest Project Role Component
Steps that Use B1.2
Steps that Produce/Modify A11.4

Transformation Plan
Description An articulation of how an Archive plans to transform the records involved in an Ingest Project, detailing the process of accepting format types from the Producer and transforming them into appropriate Preservation Formats.
Ingest Project Role Component
Steps that Use B2.5, B2.6, B3.1, B3.2
Steps that Produce/Modify A5.1, A5.4, A5.5

Transformation Policy
Description A policy that articulates an Archive's approach to making appraisal decisions concerning its transformation of records. It guides an Archive's decision to transform a record from one format to another. This Policy can exist as very general policy stating broad appraisal principles, or it can exist as a very detailed policy mapping the transformation of a specific format to a preservation format for a specific record type. This Policy can exist as information embedded in the Records Retention Policy, Format Standards Policy, Format Representation Information System, and/or Collection Policy. If the Archive is dealing with a new combination of record type and format type in Step A4.5, it probably has to add to or modify its Transformation Policy.
Ingest Project Role Resource
Steps that Use A5.5
Steps that Produce/Modify A5.5

Validation Procedures
Description A set of procedures that articulates how an Archive validates records.

Ingest Project Role Resource
Steps that Use A11.3
Steps that Produce/Modify None

Validation Procedures Decisions
Description The articulation of an Archive's decision to use a set Validation Procedures during an Ingest Project.
Ingest Project Role Component
Steps that Use None
Steps that Produce/Modify A11.3

Virus Definition Files
Description These files define viruses.
Ingest Project Role Component
Steps that Use B2.3
Steps that Produce/Modify None

APPENDIX A: USING THE INGEST GUIDE

Overview

The Ingest Guide is a prescriptive guide. It describes the steps Archives need to undertake to have a trustworthy ingest process; the Guide is not a detailed procedure manual: it does not explain how an Archive should precisely execute these steps or construct its Resources. How the Guide is used will vary greatly from Archive to Archive depending on circumstances and needs.

At first reading, the Ingest Guide may appear to prescribe a daunting process, but much of what the Guide asks Archives to undertake they already do, often intuitively or informally. Archives already appraise records; determine their appropriate access restrictions; make transformation decisions (like photocopy newspaper clippings) and apply metadata (like finding aids) to records. The Ingest Guide calls for Archives to document their decisions carefully and to base them on well-documented procedures, polices, and standards. Because electronic records are less forgiving about preservation than paper records, this documentation is imperative. In addition, traditional archival practices have often produced less than ideal results. What archivist has not contended with poorly documented terms of use, transfer, or preservation decisions?

Although following the Ingest Guide entails more work than most traditional archival accessioning methods do, this more carefully documents the accession process. It also regularizes and streamlines many decision-making steps and offers the potential to automate a considerable portion of accessioning, preservation, and description. The Guide is geared towards enabling an Archive to ingest records in a semi-automated and scalable manner. The more an archive articulates its Resources as machine-readable objects, the more it will be able to automate its ingest process. Obviously, expressing Resources as machine-readable objects can take a considerable investment of effort. Each Archive will have to determine the degree of automation that is appropriate for its operations. Archives will get the biggest payoff from automating their Ingest process when they have serial Ingests Projects from the same Producer sending the same type of records. An Archive can rapidly repeat the Ingest decisions they previously made and the more it has automated its SIP creation, transfer, validation, and transformation steps, the more quickly it can turn its SIPs into AIPs and complete its Ingest Projects. This work slows down when an Archive undertakes an Ingest Project with a new producer or record type or format since it has to add to its stable of resources to accommodate this new type of accession.

However, reaching a highly automated, trustworthy Ingest process requires individual Archives or the records and digital preservation communities to undertake a number of tasks that this Guide points out but does not fully address. Three of the most significant of these tasks are:

1. Develop Resources
   In order for an Archive to have an automated Ingest process it needs to have fully developed rules for creating the Resources and a schema for articulating them in a machine-readable manner. Currently, most of the Resources do not have the needed rules or schemas.

2 Create SIPs (Part B1)
Creating the Resource "Submission Information Package Creation Procedures" alone represents a substantial amount of work for an Archive. Not only does a Producer have to configure records into a SIP properly, it has to extract those records from a recordkeeping environment in a trustworthy manner which is not a trivial task.

3 Appraise Records (Part A3)
An Archive has to use fully developed methodologies for surveying records and for determining a record's authenticity, disposition, and essential elements. Implementing these methodologies is a major undertaking for any Archive.

APPENDIX B: EXAMPLE OF A SUBMISSION AGREEMENT

Overview

Below are the machine-readable and the human-readable versions of a single Submission Agreement. Each version is a different expression of the same information. The data in machine-readable Submission Agreement are Components that reference Resources. The Components are expressed as underlined text in the human-readable version.

The Archive uses the human-readable version to document its own and the Producer's endorsements of the Submission Agreement. It uses the machine-readable version to automate its validation and transformation work.

In this example the Archive, the Digital Collections and Archives at Tufts University, accessions the website of the Task Force on the Undergraduate Experience. It has previously collected paper records from the Task Force.

This is a detailed illustration of a Submission Agreement; it is not a guide for constructing them. This is one example of the many ways an Archive can organize its Submission Agreements. A standard syntax for expressing Submission Agreements does not yet exist. In order for Archives to use these Agreements successfully, that syntax will need to be created.

```
<submission-agreement id="SA00023">

  <record-survey id="RS00023" />

  <archive>
    <identifier>US::TUFTSU::Central::0001</identifier>
    <name>Digital Collections and Archives</name>
  </archive>

  <producer>
    <identifier>US::TUFTSU::Taskforce::0001</identifier>
    <name>Task Force on the Undergraduate Experience</name>
  </producer>

 <circumstance>
  This Ingest Project occurs after a previous Ingest Project with the Task Force
  concerning its paper records.
 </circumstance>

  <description>
   The records in this Ingest Project are records on the website
   (http://ugtaskforce.tufts.edu) of the Task Force on the Undergraduate Experience
   (CID US::TUFTSU::Taskforce::0001). These include the President's charge, reports
   generated by the Task Force, and the additional records describing the Task Force's
   activities and findings.
  </description>

  <subscribed-users>
    <user>user1</user>
    <user>user2</user>
  </subscribed-users>

  <workflow>University Records</workflow>

  <sipcreation status="009" elements="ALL" />

  <transfer status="008" elements="ALL" />

  <descriptionstandard status="001" elements="ALL" />

  <copyright status="001" elements="ALL">
   Copyright held by Tufts University.
   </copyright>

  <confidentiality status="001" elements="ALL">
   Open access to the general public.
  </confidentiality>

  <metadata elements="ALL">
    <tag tagNS  ="http://purl.org/dc/elements/1.1/"
         tagName="publisher"
         value  ="Task Force on the Undergraduate Experience"/>

    <tag tagNS  ="http://purl.org/dc/elements/1.1/"
         tagName="rights"
         value  ="Copyright Tufts University 2001"/>
  </metadata>

  <relationships>
    <sa-relationship/>

    <relationship urn="tufts:central:dca:UA088:00001"
                  relNS="info:fedora/fedora-system:def/relations-external"
```

```
                relName="isMemberOf"/>

  <relationship urn="tufts:central:dca:PROD:00001"
                relNS="http://dca.tufts.edu/ns/relations/"
                relName="producedBy"/>
</relationships>

<element id="SA00023:001">
  <description>
    University president's charge to the Task Force.
  </description>

  <urnpool>
    <list>
      <urn>tufts:central:dca:nhprc-erec:UGT:00001</urn>
    </list>
  </urnpool>

  <date range begin="1999" end="1999" />

  <object-profiles>
    <profile name="0011"/>
  </object-profiles>

  <relationships>
    <relationship urn="tufts:central:dca:RTD:00019"
                  relNS="http://dca.tufts.edu/ns/relations/"
                  relName="hasRecordType"/>
  </relationships>
</element>

<element id="SA00023:002">
  <description>
    Various interim, status, and final reports created by the Task Force.
  </description>

  <urnpool>
    <range>
      <begin>tufts:central:dca:nhprc-erec:UGT:RP001</begin>
      <end>tufts:central:dca:nhprc-erec:UGT:RP999</end>
    </range>
  </urnpool>

  <date range begin="2000" end="2003" />

  <object-profiles>
    <profile name="0011"/>
  </object-profiles>

  <relationships>
    <relationship urn="tufts:central:dca:RTD:00011"
                  relNS="http://dca.tufts.edu/ns/relations/"
                  relName="hasRecordType"/>
  </relationships>
</element>

<element id="SA00023:003">
  <description>
    List of outreach activities undertaken by the Task Force
  </description>

  <urnpool>
    <list>
```

```
        <urn>tufts:central:dca:nhprc-erec:UGT:00002</urn>
      </list>
    </urnpool>

    <date range begin="2003" end="2003" />

    <object-profiles>
      <profile name="0031"/>
    </object-profiles>

    <relationships>
      <relationship urn="tufts:central:dca:RTD:00021"
                    relNS="http://dca.tufts.edu/ns/relations/"
                    relName="hasRecordType"/>
    </relationships>
  </element>

  <element id="SA00024:004">
    <description>
      List of links to online news stories concerning the Task Force
    </description>

    <urnpool>
      <list>
        <urn>tufts:central:dca:nhprc-erec:UGT:00003</urn>
      </list>
    </urnpool>

    <date range begin="2003" end="2003" />

    <object-profiles>
      <profile name="0043"/>
    </object-profiles>

    <relationships>
      <relationship urn="tufts:central:dca:RTD:00035"
                    relNS="http://dca.tufts.edu/ns/relations/"
                    relName="hasRecordType"/>
    </relationships>
  </element>

  <element id="SA00024:005">
    <description>
      Digitized print news stories concerning the Task Force
    </description>

    <urnpool>
      <range>
        <begin>tufts:central:dca:nhprc-erec:UGT:00100</begin>
        <end>tufts:central:dca:nhprc-erec:UGT:00199</end>
      </range>
    </urnpool>

    <date range begin="2003" end="2003" />

    <object-profiles>
      <profile name="0011"/>
    </object-profiles>

    <relationships>
      <relationship urn="tufts:central:dca:RTD:00035"
                    relNS="http://dca.tufts.edu/ns/relations/"
                    relName="hasRecordType"/>
```

```
</element>


<element id="SA00024:006">
  <description>
    List of Task Force members.
  </description>

  <urnpool>
    <list>
       <urn>tufts:central:dca:nhprc-erec:UGT:00004</urn>
    </list>
  </urnpool>

  <date range begin="2003" end="2003" />

  <object-profiles>
    <profile name="0031"/>
  </object-profiles>

  <relationships>
    <relationship urn="tufts:central:dca:RTD:00015"
                        relNS="http://dca.tufts.edu/ns/relations/"
                        relName="hasRecordType"/>
  </relationships>
</element>

<element id="SA00024:007">
  <description>
     List of links to studies concerning undergraduates at other institutions
     the Task Force used as benchmarks
  </description>

  <urnpool>
    <list>
       <urn>tufts:central:dca:nhprc-erec:UGT:00005</urn>
    </list>
  </urnpool>

  <date range begin="2003" end="2003" />

  <object-profiles>
    <profile name="0043"/>
  </object-profiles>

  <relationships>
    <relationship urn="tufts:central:dca:RTD:00015"
                        relNS="http://dca.tufts.edu/ns/relations/"
                        relName="hasRecordType"/>
  </relationships>
</element>

<element id="SA00024:008">
  <description>
    Content of the Task Force website as a whole.
  </description>

  <urnpool>
    <list>
       <urn>tufts:central:dca:nhprc-erec:UGT:00006</urn>
    </list>
  </urnpool>
```

```
    <date range begin="ca. 2000" end="2003" />

    <object-profiles>
      <profile name="0064"/>
    </object-profiles>

    <relationships>
      <relationship urn="tufts:central:dca:RTD:00017"
                    relNS="http://dca.tufts.edu/ns/relations/"
                    relName="hasRecordType"/>
    </relationships>
  </element>

</submission-agreement>
```

SUBMISSION AGREEMENT

Submission Agreement ID SA00023

This Submission Agreement defines the terms of the transfer of records described in this Agreement and in the Records Survey (RS00023) (hereafter known as "The Records") from the Task Force on the Undergraduate Experience (US::TUFTSU::Taskforce::0001) to the Digital Collections and Archives (DCA).

By agreeing to this Submission Agreement the Task Force on the Undergraduate Experience (US::TUFTSU::Taskforce::0001) declares that it has the proper authority to transfer the records to the DCA.

Special circumstances of Ingest Project
This Ingest Project occurs after a previous Ingest Project with the Task Force concerning its paper records.

General Description of the Records
The records in this Ingest Project are records on the website (http://ugtaskforce.tufts.edu) of the Task Force on the Undergraduate Experience (CID US::TUFTSU::Taskforce::0001). These records are the President's charge, reports generated by the Task Force, and the additional records describing the Task Force's activities and findings.

Detailed Description the Records

All of the records
Will be prepared for transfer according to the DCA SIP Creation rule for web-based records. (SIP Creation Rule 009)

All of the records
Will follow the DCA standard transfer procedures for university records on web server. (Transfer Rule 008)

All of the records
Will be part of Task Force on Undergraduate Experience Collection (UA088)

All of the records
Will be described according to the DCA standard descriptive rule for university records. (Metadata Descriptive Rule 01)

All of the records
Copyright held by Tufts University. (Copyright Status 001)

All of the records
Open access to the general public. (Records Security Profile 01)

President's Charge (SA00023:001)
These record(s) are the University President's charge to the Task Force. They are Charges (Record Type 00019) created from 1999 through 1999 in the form of PDF documents that the DCA will keep and preserve as PDF documents (Object-Profile 0011).

Various Reports (SA00023:002)
These record(s) are the Various interim, status, and final reports created by the Task Force. They are Reports (Record Type 00011) created from 2000 through 2003 in the form of PDF documents that the DCA will keep as PDF documents (Object-Profile 0011).

Outreach Activities List (SA00023:003)
These record(s) are the List of outreach activities undertaken by the Task Force. They are Event Records (Record Type 00021) created from 2003 through 2003 in the form of an HTML file that the DCA will normalize into plain text (Object-Profile 0031).

Links to News Stories (SA00023:004)
These record(s) are the List of links to online news stories concerning the Task Force. They are News Clippings (Record Type 00035) created from 2003 through 2003 in the form of an HTML file that the DCA will normalize into an XBEL file (Object-Profile 0043).

News Stories (SA00023:005)
These record(s) are the Digitized print news stories concerning the Task Force. They are News Clippings (Record Type 00035) created from 2003 through 2003 in the form of PDF documents that the DCA will keep as PDF documents (Object-Profile 0011).

Membership List (SA00023:006)
These record(s) are a List of Task Force members. They are Subject Files (Record Type 00015) created from 2003 through 2003 in the form of an HTML file that the DCA will normalize into plain text (Object-Profile 0031).

Benchmarking Studies (SA00023:007)
These record(s) are the List of links to studies concerning undergraduates at other institutions the Task Force used as benchmarks. They are Subject Files (Record Type 00015) created from 2003 through 2003 in the form of an HTML file that the DCA will normalize into an XBEL file (Object-Profile 0043).

Website (SA00023:008)
These record(s) are the content of the Task Force website as a whole. They are Publications (Record Type 00017) created in ca. 2000 through 2003 in the form of HTML files, PDF files, and JPEG files that the DCA will normalize together into an ZIP Tidy file (Object-Profile 0064).

Endorsement of Submission Agreement

The Digital Collections and Archives agrees to the conditions of this Submission Agreement, which commits the DCA to accession the records into its holdings according to the terms of the Submission Agreement.

_____
Eliot Wilczek
University Records Manager
Authorized Representative of the Digital Collections and Archives

The Task Force on the Undergraduate Experience, agrees to the conditions of this Submission Agreement, which commits the Task Force on the Undergraduate Experience to transfer the records to the DCA according to the terms of the Submission Agreement.

_____
Armand Greene
Director
Authorized Representative of Task Force on the Undergraduate Experience

APPENDIX C: PRODUCER-ARCHIVE INTERFACE METHODOLOGY ABSTRACT STANDARD CROSSWALK

Overview

The Ingest Guide builds on the work of the Consultative Committee for Space Data Systems, particularly its 2004 *Producer-Archive Interface Methodology Abstract Standard*,[12] which consists of four phases: Preliminary, Formal, Transfer, and Validation. In its Preliminary and Formal phases, the *Producer-Archive Interface* proposes a detailed description of the steps needed to produce a submission agreement between a Producer and an Archive. These steps form a foundation for the development of Section A of the Ingest Guide, Negotiate Submission Agreement. While the *PAI* separates the submission agreement process into separate Preliminary and Formal phases, the Ingest Guide does not make such a distinction. The *PAI's* Transfer, and Validation phases are much less detailed than its Preliminary and Formal phases and only guided in a general way the development of Section B of the Ingest Guide, Transfer and Validation.

The crosswalks from the Ingest Guide to the *PAI* should help uses understand how these two documents relate to each other.

[12] 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>.

Crosswalks

Ingest Guide To Producer-Archive Interface Methodology Abstract Standard Crosswalk	
Ingest Guide	Producer-Archive Interface Methodology Abstract Standard
A1.1	P1
A1.2	P1
A1.3	None
A1.4	None
A1.5	None
A1.6	None
A2.1	P2, P3
A2.2	None
A2.3	None
A2.4	None
A2.5	None
A2.6	None
A3.1	P2, P9, P12, P16, P23, P24, P29, P30
A3.2	None
A3.3	P3
A3.4	None
A3.5	None
A4.1	None
A4.2	None
A5.1	P10
A5.2	P9, P10
A5.3	P4, P5
A5.4	P10, P12, P14
A5.5	P10, P12, P14
A5.6	None
A6.1	P16
A6.2	P17
A6.3	P17
A6.4	P17
A7.1	P29
A7.2	P29
A7.3	P29
A7.4	P29
A7.5	P29
A7.6	None
A7.7	F13
A8.1	P23, P24, P25, P30
A8.2	P23, P24, P25, P30
A8.3	P23, P24, P25, P30
A8.4	P42
A8.5	P42

A8.6	None
A8.7	None
A8.8	TBD
A8.9	P6
A10.1	P8, P15, P18, P22, P27, P33, P36, P40, P42, P43
A10.2	P43
A10.3	None
A10.4	None
A10.5	None
A11.1	F4
A11.2	F14, F15, F16, F17, F18, F19
A11.3	F20, F21, F22, F23, F24, F25
A11.4	F16, F26
A11.5	F15
A11.6	F36
A11.7	None
A11.8	F36
A11.9	F36
A11.10	F36
B1.1	T2
B1.2	T2
B2.1	T2
B2.2	V2
B2.3	V2
B2.4	V2
B2.5	V2
B2.6	V2
B2.7	V3
B2.8	V3
B3.1	None
B3.2	None
B3.3	None
B3.4	None
B4.1	None
B5.1	V2
B5.2	V3
B5.3	T3
B5.4	V2
B5.5	None
B6.1	T2
B6.2	T2

Fedora and the Preservation of University Records Project

2.2 Ingest Projects

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/

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:UA069.004.001.00007
(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

Table of Contents

OVERVIEW

Although the Ingest Guide is a prescriptive guide for managing an ingest process, archives can implement the Guide in a variety of ways, from an entirely manual to an extensively automated process. Archives can use the Guide to manage a wide range of accessions: small or large acquisitions, complex or simple collections, single or recurring accessions. Both large and small archives in a variety of industries can use the Guide. In order to explore and illustrate how archives could implement the Ingest Guide, the project team undertook three Ingest projects.

The Ingest projects were

1. Website of an ad-hoc committee on undergraduate life
2. Working papers of the Board of Trustees saved as desktop applications on CDs
3. Library administration records saved as desktop applications stored in an instance of SharePoint Team Services, a web-based share-space environment.

These three projects are only examples of how archives *could* use the Ingest Guide; they are *not* instructions for how archives must use the Guide.

METHODOLOGY

Each Ingest project's description consists of an:

- Ingest Project Narrative
- Survey Report
- Ingest Guide Steps: Archive and Producer Actions

Ingest Project Narrative

The Ingest Project Narrative provides a descriptive summary of the actions the project team actually undertook during each Ingest project. Because the project team undertook the projects before the Guide was finalized, the actions they undertook do not precisely follow all the steps of the Guide. In addition, the Ingest Guide calls for archives to support their Ingest processes with a variety of resources that the project team did not have and their development was beyond the scope of this research project. Therefore the project team had to undertake these projects without the benefit of these resources.

Survey Report

While the survey reports in this document are based on the actual actions undertaken by Tufts and Yale during each of their six Ingest projects, project staff did not create these reports while undertaking their projects. The survey reports found below are theoretical explorations of how Tufts or Yale might construct their reports for a fully operational Ingest process described by the Ingest Guide.

The Ingest Guide calls for the Archive to start a survey report early in Section A of the Guide, Negotiate Submission Agreement, and to continue to add information to the survey throughout many of the Steps in Section A. The Ingest Guide defines a survey report as:

> A Report that identifies the records an Archive should accession during the Ingest Project. Survey Reports can vary greatly in detail, from a general description of the records the Archive should accession to an item-level inventory of those records. An Archive may create an early working draft of the Report in Step A2.1, after it and the Producer agree on the scope of the records that will be surveyed. In Steps A3.1 and A3.2, the Archive describes in the Survey Report the records it surveyed to the level of detail it requires. In Steps A3.3 and A3.4, the Archive documents in the Report its decisions on which, if any, records in the survey it should accession and what essential elements of these records it needs to preserve. To guide the Archive's appraisal decisions in Steps A3.3 and A3.4 and to be useful in Parts A3 through A10, the Survey Report needs to identify the records' Producer, Record Types, format type, file size, confidentiality requirements, copyright status, and any Producer-created identifiers.

The survey reports in these Ingest Project descriptions consist of five parts:

1 *Background*
This gives a brief overview of the Ingest project.

2 *Description of Records Surveyed*
This contains a general description the records and identifies the Producer, the functions of the records, the record types of the records, their format types, file size, any Producer identifiers, any confidentially and access restriction requirements, the copyright status, and listing (to varying degrees of detail) of the records within the purview of the Ingest project.

3 *Evaluation of Recordkeeping System*
This contains a description of how well the recordkeeping system storing and managing the records enables the feasible and scaleable transfer of those records to the Archive. This portion of the survey report also indicates if the records are managed according to the rules of the recordkeeping system.

4 *Evaluation of Authenticity*
This presents the Archive's judgment of the authenticity of the records involved in the Ingest Project. This section describes the reasoning behind the Archive's judgment.

5 *Appraisal Decision*
This documents the Archive's appraisal decision on the records in the Ingest Project. Each type of record listed the "Description of Records Surveyed" section has a disposition decision; a reason for disposition that usually references a retention schedule, collection policy, or other warrant for the decision; and a description of the essential elements of the records.[1]

Ingest Guide Steps: Archive and Producer Actions
These three-column spreadsheets describe the actions Tufts and Yale might take for each step of the Ingest Guide with a fully operational Ingest process described by the Ingest Guide. Like the survey reports, these spreadsheets are theoretical explorations of what Tufts and Yale might do, not what they actually did in the actual Ingest Projects.

The first column lists the steps of the Ingest Guide, the second column describes the actions Tufts or Yale would undertake for their Ingest projects, and the third column describes the actions of the Producer. A step from the Guide that requires no actions in a particular Ingest Project is not listed in that Project's spreadsheet. Terms in bold in the Archive Action or Producer Action columns refer to Resources, Products, or Documentation in the Ingest Guide.[2]

Many of the Archive's actions refer to an "ingest application." This application refers generically to an application that an archive would implement to handle many of the tasks described in Part

[1] The essential elements analysis, which considers documentary form, annotations, context, and medium, is largely based on the InterPARES, "Authenticity Task Force Report," *The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project,* <http://www.interpares.org/book/interpares_book_d_part1.pdf>
[2] See the "Components, Resources, Products, and Documentation" section of the Ingest Guide.

B, Transfer and Validate, of the Ingest Guide. These tasks include transfer, validation, format transformation, and turning SIPs into AIPs.[3]

[3] Tufts University developed the Tufts Ingest Prototype System (TIPS) as an initial attempt to develop such an application that would help an Archive semi-automate the steps in Part B of the Ingest Guide. See 2.3 Ingest Tools for a description of the application.

INGEST PROJECT ONE
Task Force on the Undergraduate Experience

Ingest Project Narrative

In 2001 Tufts University created the Task Force on the Undergraduate Experience to study undergraduate life at Tufts University. The Taskforce completed its mission in June 2003. In April 2003 the Task Force Project Coordinator contacted the Digital Collections and Archives (DCA) concerning the records of the Task Force. The DCA accessed paper records and several CDs from the Task Force in May 2003. However, the Task Force also created a website as part of its business. The DCA did not accession the website and only gave it a cursory appraisal at the time. The Task Force officially disbanded at the end of June 2003.

In October 2005 the project team decided to use the website of the Task Force as one of our Ingest Projects because it raises a number of interesting appraisal issues and represents a situation that most archives at colleges and universities are likely to face. There are important issues regarding how to view the website as a record, the essential elements of the website, and the way to run an ingest project in the absence of the original producer. Because the website is still live on a public webserver, the project team was able to access the website at any time.

To create the SIP, the project team used the wget utility[4] to fetch a copy of the website to a local filesystem. The team then manually extracted records from the website that the team judged to have enduring value. These records composed the website which included PDFs of reports and some of the text of the HTML files that made up the site. The project team then ran the HTML of the website through the W3C HTML tidy utility[5] to produce nearly-valid XHTML. Then the team packaged both the original and tidied versions of the entire site into separate ZIP files. The team packaged together all of the records into a single SIP.

The project team submitted this SIP to the Tufts Ingest Prototype System (TIPS).[6] This application performed all the checks of Part B.2 from the Ingest Guide (albeit with minimal validation) and then created one AIP for each record, with the tidied and untided versions of the whole website as one AIP. The ingest application was able to submit many of the AIPs to Tufts' Fedora 1.2.1 repository. However, the largest record (containing the contents of the website), caused the ingest system to crash. The problem is related to the way SOAP bindings to Fedora 1.2.1 are handled. Fixing this problem will require upgrading the Fedora SOAP interface to use SAAJ or MOTM.

The project team skipped the difficult task of building resources. We did not create record type records, format type information, or producer records and none already existed to utilize. How to represent, manage, and acquire this information remains an open question. This resulted in us

[4] Wget is a command-line tool for UNIX-like systems which retrieves files using the HTTP, HTTPS, and FTP protocols. See http://www.gnu.org/software/wget/wget.html for additional information on GNU wget.
[5] Tidy is a command-line tool which parses HTML and produces normalized HTML or XHTML. It is carefully written to maintain the visual appearance of the HTML as much as possible. See http://tidy.sourceforge.net/.
[6] See 2.3 Ingest Tools for a description of the application.

creating a number of records in the Fedora repository with dummy links to stub resources—essentially empty Fedora objects—that stood in for various resources.

Survey Report

Background
Eliot Wilczek, University Records Manager, Digital Collection and Archives (DCA) met with Armand Greene, Project Coordinator of the Task Force on the Undergraduate Experience on 04/10/03 to conduct a records survey. He sent a Records Survey Report to Armand Greene on 04/17/03. The report essentially deferred on the appraisal of the Task Force's website. The report said in part:

> "The website essentially provides access to a variety of documents. The DCA would focus on preserving these documents rather than saving the whole website as a website. Documents and records on the website include:
>
> President Bacow's charge
> Progress information which includes a meetings list
> Questions of the week
> Listing of people on the Task Force
> Reports and proposals
> Bibliography of the Task Force in the news with PDF news clippings
>
> "Transfer all to the Digital Collections and Archives except the PDF news clippings. The DCA has these publications so destroy the clippings when they are no longer needed. Transfer to the DCA the bibliography listing the articles concerning the Task Force.
>
> "Many of the documents on the website exist in electronic and/or paper format elsewhere and therefore the copies on the website may not need to be transferred to the DCA. This should be discussed in further detail."

The Task Force transferred records to the DCA on 05/22/03 in accordance with the records survey. These were paper records and electronic records in desktop application formats on CDs. The website or components of the website were not transferred to the DCA, although the reports on the website were transferred to the DCA in paper and desktop application format.

Armand Greene verbally informed Eliot Wilczek on 05/22/03 the Task Force website would remain at http://ugtaskforce.tufts.edu for sometime into the future, although Mr. Greene did not specify a time period. Eliot Wilczek verbally indicated to Armand Green that the DCA will look at accessioning the website or components of the website sometime in the future. Armand Green accepted this proposed effort.

On 06/30/03 the Task Force on the Undergraduate Experience concluded its activities and ceased as a unit of Tufts University in accordance with its mandate from the President of Tufts University.

On 10/12/05 Eliot Wilczek conducted a records survey of the Task Force website located at http://ugtaskforce.tufts.edu as part of a separate Ingest project.

Description of Records Surveyed

General Description

This is the website of the Task Force on the Undergraduate Experience. The website is located at http://ugtaskforce.tufts.edu. It is a relatively small and simple website. It has the following sections:

Home page
President's Charge
Progress
In the News
People
Contact Us

Producer

Task Force on the Undergraduate Experience

Producer Role(s) Creator

Function(s)

The website serves as the main vehicle for the Task Force to disseminate its findings and information about its activities to the Tufts community.

Record Type(s)

Charges
Reports
Event Records
News Clippings
Subject Files
Publications

Format Type(s)

Web pages composed in html.

For most pages in `<head></head>`

```
<meta name="GENERATOR" content="Microsoft FrontPage 5.0">
<meta name="ProgId" content="FrontPage.Editor.Document">
```

Not valid HTML, no `DOCTYPE` found, attempted validation HTML 4.01 Transitional.
Validation used: W3C Markup Validation Service v0.7.0
http://validator.w3.org accessed 10/13/05.

http://ugtaskforce.tufts.edu/contactus.html notes

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN">
```

Not valid HTML 4.0 Transitional.
Validation used: W3C Markup Validation Service v0.7.0
http://validator.w3.org accessed 10/13/05.

PDF files.

No validation performed.

JPG image files.
No validation performed.

File Size
Approximately 5 to 10 MB. A substantial portion of this is the pdfs of reports and news clippings.

Producer Identifier
All of the website and all of its components do not have producer identifiers that need preservation.

Confidentiality Requirements/Access Restrictions
All of the website and all of its components have no requirements for access restrictions. They should all have Category 2: Universal Distribution.

Copyright Status
Copyright of website as a whole and individual reports held by Tufts University.

Copyright of *Tufts Daily* articles in the news clippings held by the *Tufts Daily*.

Copyright of other student publications in the news clippings held by the author.

Copyright of Tufts University publications in the news clippings held by Tufts University.

List of Records Surveyed
President's Charge

Description	University President's charge to the Task Force.
Record Type	Charges
Dates	2001
Format	HTML and PDF

Various Reports

Description	Various interim, status, and final reports created by the Task Force.
Record Type	Reports
Dates	2001 through 2003
Format	HTML and PDF

Outreach Activities List

Description	List of outreach activities undertaken by the Task Force.
Record Type	Event Records
Dates	2003
Format	HTML

Links to News Stories

Description	List of links to online news stories concerning the Task Force.

Record Type	News Clippings
Dates	2003
Format	HTML

News Stories

Description	Digitized print news stories concerning the Task Force; provides bibliographic citations of the news stories.
Record Type	News Clippings
Date	2003
Format	PDF

Membership List

Description	List of Task Force members.
Record Type	Subject Files
Dates	2003
Format	HTML

Benchmarking Studies List

Description	List of links to studies concerning undergraduates at other institutions the Task Force used as benchmarks.
Record Type	Subject Files
Dates	2003
Format	HTML

Website

Description	The Task Force website as a whole.
Record Type	Publications
Dates	ca. 2001 through 2003
Format	HTML, PDF, JPG

Evaluation of Recordkeeping System

The PDF, image, and HTML files are stored on a public webserver accessible at http://ugtaskforce.tufts.edu. These files appear to be stored on the webserver in normal manner although the Digital Collections and Archives did not find any procedures or rules on the management of these files. Acting as both the Archive and the Producer, the DCA does not have direct access to the server and must access the files through a web browser. This is a trustworthy but not scaleable transfer process. However, because the website is small and the DCA will only have to capture files from the site once because the Task Force no longer exists and the website will not change, the DCA will make the effort to manually capture the files.

Evaluation of Authenticity

The Digital Collections and Archives judges the Task Force on the Undergraduate Experience website at http://ugtaskforce.tufts.edu and all of its component parts that it has evaluated in this survey to be authentic for the following reasons:

- At the 05/22/03 survey interview, Armand Green declared that the Task Force created the website in the normal course of its business.
- At the 05/22/03 survey interview, Armand Green identified the website at http://ugtaskforce.tufts.edu as the Task Force website.
- During the 10/12/05 survey, Eliot Wilczek determined the website at http://ugtaskforce.tufts.edu was the same website that Armand Green identified as the website on 05/22/03 based on a brief visual review of the website's appearance and content.
- The probability someone would maliciously alter the website—particularly its content—and try to hide that content alteration is extremely low because no one has a reasonable motivation to undertake such an action.

Appraisal Decision

President's Charge
- Disposition: Transfer to DCA
- Reason for Disposition: Retention Schedule gen071
- Essential Elements
  - Documentary Form
    - Content: *Verbatim*
    - Presentation: Textual Structure: *Verbatim*; Webpage Structure: *General Appearance*; PDF Structure: *General Appearance*
    - Signs: None
  - Annotations: None
  - Context: *Information* that delivered as webpage and PDF document via HTTP protocol with no restrictions
  - Medium: *Information* that HTML and PDF

Various Reports
- Disposition: Transfer to DCA
- Reason for Disposition: Retention Schedule gen065
- Essential Elements
  - Documentary Form
    - Content: *Verbatim*
    - Presentation: Textual Structure: *Verbatim*; Webpage Structure: *General Appearance*; PDF Structure: *General Appearance*
    - Signs: None
  - Annotations: None
  - Context: *Information* that delivered as webpage and PDF document via HTTP protocol with no restrictions
  - Medium: *Information* that HTML and PDF

Outreach Activities List
- Disposition: Transfer to DCA
- Reason for Disposition: Retention Schedule gen034

Essential Elements
Documentary Form
Content *Verbatim*
Presentation Textual Structure: *General Appearance*
Webpage Structure: *General Appearance*
Signs None
Annotations None
Context *Information* that delivered as webpage document via HTTP protocol with no restrictions
Medium *Information* that HTML

Links to News Stories
Disposition Transfer to DCA
Reason for Disposition Retention Schedule gen024
Essential Elements
Documentary Form
Content *Verbatim*
Presentation Textual Structure: *General Appearance*
Webpage Structure: *General Appearance*; *Information* of hyperlink URL, hyperlink functionality not needed
Signs None
Annotations None
Context *Information* that delivered as webpage document via HTTP protocol with no restrictions
Medium *Information* that HTML

News Stories
Disposition Destroy when no longer needed. [Because it is difficult to remove these records from the website when the DCA captures the Website as a whole, the DCA will probably transfer the News Stories to the DCA. However, the DCA would make no effort to preserve the News Stories.]
Reason for Disposition Although Retention Schedule gen024 applies, the DCA already holds the publications that contain these News Stories.

Membership List
Disposition Transfer to DCA
Reason for Disposition Retention Schedule gen010
Essential Elements
Documentary Form
Content *Verbatim*
Presentation Textual Structure: *General Appearance*
Webpage Structure: *General Appearance*
Signs None

- Annotations: None
- Context: *Information* that delivered as webpage document via HTTP protocol with no restrictions
- Medium: *Information* that originally in HTML

Benchmarking Studies List
- Disposition: Transfer to DCA
- Reason for Disposition: Retention Schedule gen010
- Essential Elements
  - Documentary Form
    - Content: *Verbatim*
    - Presentation: Textual Structure: *General Appearance*
      Webpage Structure: *General Appearance*; Information of hyperlink URL, hyperlink functionality not needed
    - Signs: None
  - Annotations: None
  - Context: *Information* that delivered as webpage document via HTTP protocol with no restrictions
  - Medium: *Information* that originally in HTML

Website (as a whole)
- Disposition: Transfer to DCA
- Reason for Disposition: Retention Schedule gen015
- Essential Elements
  - Documentary Form
    - Content: *Verbatim*
    - Presentation: Textual Structure: *General Appearance*
      Webpage Structure: *General Appearance*; *Information* of internal and external hyperlink URL, hyperlink *functionality* not needed. [Preserving *functionality* of internal hyperlink because it is an easier preservation strategy than preserving *information* of internal hyperlink is ok]
    - Signs: None
  - Annotations: None
  - Context: *Information* that delivered as website via HTTP protocol with no restrictions
  - Medium: *Information* that originally in HTML

Ingest Guide Steps: Archive and Producer Actions		
Steps	**Archive Actions**	**Producer Actions**
A Negotiate Submission Agreement		
A1 Establish Relationship		
A1.1 Initiate Contact/Ingest Project	The Archive had previously worked with the Producer on another Ingest Project. At that time, the Archive and Producer informally agreed that after the Producer ceases operations, the Archive would act as both the Archive and the Producer in this Ingest Project.	The Producer had previously worked with the Archive on another Ingest Project. At that time, the Archive and Producer informally agreed that after the Producer ceases operations, the Archive would act as both the Archive and the Producer in this Ingest Project.
A1.2 Identify Producer	The Archive identifies the Task Force on the Undergraduate Experience as the Producer.	
A1.3 Has the Archive already defined its relationship with the Producer?	Yes, the Archive already defined its relationship with the Producer when it ingested the paper records of this Task Force. *The Archive skips Steps A1.4 through A1.6.*	
A2 Define Project		
A2.1 Identify records at issue, agreeing upon scope of survey	The Archive and Producer (during previous Ingest Project) identify the contents of the Task Force website as the records of this Ingest Project.	The Producer and Archive (during previous Ingest Project) identify the contents of the Task Force website as the records of this Ingest Project.
A2.2 Does the Producer have custody/authority over the records identified in A2.1?	Yes, the Archive determines that the Task Force has the appropriate authority over the records identified in Step A2.1. *The Archive skips Steps A2.3 through A2.6.*	
A3 Collect Information and Assess Value of Records		
A3.1 Conduct Records Survey, note attributes of records	The Archive conducts a survey and produces a **Survey Report**.	The Archive as the Producer allows access to the records for the survey.
A3.2 Judge authenticity of records	The Archive examines content of records and speaks to the original Task Force Project Coordinator (during previous Ingest Project) and determines that they are very probably authentic, and updates the **Survey Report**.	
A3.3 Should the Archive accession at least some of the records?	Yes, the Archive determines that the content of the website as a whole and the reports contained therein have archival value, and judges them to be authentic. The Archive updates the **Survey Report** to reflect this decision.	

A3.4 Determine the essential elements of the records that should be accessioned	The Archive examines records and determines that the essential elements of the reports include the content of the text and the formatting. It determines the essential elements of the website include the content of the pages and evidence of the way in which the Task Force presented itself via the site. The Archive adds its determinations to the **Survey Report**.	
A4 Assess Record Type		
A4.1 Are all records identified as a Record Type?	Yes, the Archive determines that the reports are of record type *Report*, and the website is of record type *Publication*. These record types are already known. The Archive creates a **Record Type List** which references **Record Type Records.** *The Archive skips Step A4.2*.	
A5 Assess Formats		
A5.1 Are any records in file formats that are not a preservation format?	Yes, the Archive determines the following: The website is non-valid HTML, which is not a preservation format; the reports are in PDF format, which is a preservation format; the images are in GIF and JPEG formats, which are preservation formats. For the non-valid HTML files the Archive produces a **Transformation Plan**.	
A5.2 Should Archive transform or natively handle these formats?	The Archive will transform the non-valid HTML. *Archive skips Steps A5.3 through A5.4.*	
A5.5 Choose appropriate format	The Archive will transform the non-valid HTML into valid XHTML. The Archive updates its **Format Transformation Plan** to reflect this decision.	
A5.6 Is format chosen in A5.5 a Preservation Format the Archive already uses?	Yes, the Archive determines that valid XHTML is a preservation format based on its **Formats Standards Policy**.	
A6 Assess Identifier Rules		
A6.1 Is there a Producer naming/ identification scheme that needs accommodation?	No, the Archive determines there is no **Producer Naming/ Identification Scheme** that it needs to preserve or accommodate. The *Archive skips Steps A6.2 through A6.3.*	
A6.4 Determine appropriate naming/ identification scheme(s).	The Archive decides to use the standard **Archive Naming/ Identification Scheme**. The Archive records its **Naming/Identification Scheme Decision**.	

A7 Assess Copyright		
A7.1 Determine copyright status of records in Ingest Project	The Archive determines that the records are under copyright of Tufts University. The Archive records the **Copyright status**.	
A7.2 Does the Archive need to acquire copyright or license for records?	No, the Archive determines that the University holds copyright. *Archive skips Steps A7.3 through A7.8.*	
A8 Assess Access Rights		
A8.1 Determine records' Records Security Profile	The Archive determines that Task Force previously made the records available to the general public and has should have a **Records Security Profile** of open access. The Archive documents this decision as a **Records Security Profile Decision** which references the **Records Security Profile**.	
A8.2 Does current security component meet the access control needs of the records?	Yes, the Archive determines the security component of the Preservation System meets the needs of the **Records Security Profile**. *Archive skips Steps A8.3 through A8.7.*	
A9 Assess Recordkeeping System		
A9.1 Has the Archive documented recordkeeping system as supporting feasible and trustworthy transfer?	No, the Archive has not previously examined the webserver that stores and serves the website.	
A9.2 Can recordkeeping system support feasible and trustworthy transfer?	The Archive determines that the recordkeeping system can support the trustworthy transfer of records to the Archive, but not in a scalable manner. It can support a feasible transfer if the volume of records is low. The Archive produces a **Recordkeeping System Report.** *The Archive skips Step A9.3.*	
A9.4 Is Archive or Producer willing to take extraordinary measures to transfer records?	The Archive determines that it is willing to manually transfer the records from the recordkeeping system to the Archive because of the small volume of records involved in the Ingest Project. The Archive produces **SIP Creation Procedures** for moving the records from the recordkeeping system to the Archive. *The Archive Steps A9.5 through A9.7.*	
A10 Assess Feasibility		

A10.1 Can the Archive feasibly accession the records?	Yes. The Archive determines that it is feasible to manage and preserve the records and its requirements without extraordinary effort or special accommodation. The Archive produces a **Preservation System Availability Statement**. *The Archive skips Steps A10.2 through A10.5.*	
A11 Finalize Submission Agreement		
A11.1 Add description of Metadata Encoding Rules to Submission Agreement.	The Archive chooses standard Dublin Core metadata encoding and documents its **Metadata Encoding Rules Decision**.	
A11.2 Add description of Transfer Procedures to Submission Agreement	The Archive determines that it will retrieve the contents of the site via the HTTP protocol from the public webserver that currently hosts the website. The Archive documents its **Transfer Procedures Decision** in the Submission Agreement.	
A11.3 Add description of Validation Procedures to Submission Agreement	The Archive documents the validation procedures for HTML, JPEG, GIF and PDF files in the **Validation Procedures Decision**.	
A11.4 Add Transfer Schedule to Submission Agreement	The Archive selects an indefinite schedule, and documents this decision in the **Transfer Schedule Decision**.	
A11.5 Add SIP Creation Procedures to Submission Agreement	The Archive chooses its standard SIP format and documents its **SIP Creation Procedures Decision**.	
A11.6 Finalize Submission Agreement	The Archive draws up the **Draft Submission Agreement** based on its previous decisions.	
A11.7 Does Archive and Producer agree to and approve the Submission Agreement?	Yes, the Archive and Archive-as-Producer agree and produce the **Finalized Submission Agreement**. The Archive submits finalized submission agreement to the ingest system. The Archive's ingest application accepts and validates a machine-readable version of the **Finalized Submission Agreement.** *Archive skips Steps A11.8 through A11.10.*	Yes, Archive-as-Producer and Archive agree and produce the **Finalized Submission Agreement**.
B **Transfer and Validation**		
B1 **Create and Transfer SIPs**		
B1.1 Producer prepares SIP according to Submission Agreement		The Archive-as-Producer retrieves the content of the website to a workspace under Archive control. There the Archive manually extracts the subparts of the website and constructs the **SIP**. The Archive-as-Producer signs the SIP with its own digital signature.

B1.2 Producer transfers the SIP to the Archive		The Archive-as-Producer places the SIP in the ingest application drop-box.
B2 **Validate**		
B2.1 Archive receives SIP from Producer	The Ingest application accepts the SIP from the drop-box and produces **Documentation of Receipt** and delivers it to the Archive-as-Producer.	The Archive-as-Producer receives **Documentation of Receipt**.
B2.2 Is SIP well formed?	The ingest application checks the SIP format. The SIP is well formed. The Application updates the **SIP validity statement.**	
B2.3 Does SIP contain malicious code?	The ingest application scans the SIP components for viruses and other malicious code. All SIP components are clean. The application updates the **SIP validity statement**.	
B2.4 Is the submitter authorized to submit SIP to the Archive?	The ingest application validates the SIP signatures and validates identities against its database of certificates. The SIP is signed by an authorized person. The application updates the **SIP validity statement**.	
B2.5 Does SIP contain all necessary records components?	The ingest application checks all included records for completeness. All records in the SIP are complete. The application updates the **SIP validity statement**.	
B2.6 Do the record components in SIP validate?	The ingest application tests the record components for validity, where necessary. All record components validate. Application updates the **SIP validity statement**.	
B3 **Transform and Attach Metadata**		
B3.1 Do any of the records in SIP require transformation?	Yes, the Archive determines that the HTML files require tidying and transformation to valid XHTML according to the **Format Transformation Plan** in the Submission Agreement.	
B3.2 Perform transformation on records that require transformation	The Archive tidies and transforms the HTML files into valid XHTML, producing **Transformed Records**.	
B3.3 Attach to records metadata inferred from Submission Agreement	The ingest application attaches stock metadata from Submission Agreement, **Producer Record**, and **Record Type Records**, creating **Records with Attached Metadata**.	
B3.4 Attach to records the Records Security Profile defined by Submission Agreement	The ingest application attaches the **Record Security Profile** identified in the Submission Agreement, creating a **Records with Security Profile**.	
B4 **Formulate AIPs**		

B4.1 Formulate AIPs	The Ingest application creates an **AIP** for each record.	
B5 **Assess AIPs**		
B5.1 Are all of the records in the AIP part of accession described by Submission Agreement?	The Archive verifies that all of the records to be accessioned come from the website and produces an **AIP Validity Statement**. The ingest application accepts the **AIP Validity Statement**. *The Archive skips Steps B5.2 through B5.3.*	
B5.4 Is proper metadata attached to records in the AIP?	The Archive verifies that all of the records have sufficient and correct metadata and updates the **AIP Validity Statement**. The ingest application accepts the **AIP Validity Statement**. *The Archive skips Step B5.5.*	
B6 Formally Accession		
B6.1 Submit AIPs into Preservation Repository.	The ingest application submits the AIPs to the Preservation Repository.	
B6.2 Formally notify Producer that Archive has accepted and accessioned records described by Ingest Project.	The ingest application generates a **Transfer Notice** for the Producer and an entry in the **Accession Log**.	The Producer receives the **Transfer Notice**.

INGEST PROJECT TWO
Board of Trustees

Ingest Project Narrative

The Board of Trustees of Tufts University meets three times a year. Before each of these meetings, the Office of the Board of Trustees compiles a set of materials relevant to the upcoming meeting, including agendas, reports, minutes of past meetings, and other meeting records. In the past, staff from the Office of the Trustees assembled a ring binder containing these materials for each board member, mailing it to him or her several weeks before the meeting. However, in 2003, the Board began distributing these materials electronically, by burning CDs containing electronic versions of the working papers rather than filling binders.

The project team decided to make these meeting packet CDs one of ingest projects because it represents an interesting case of dealing with physical media, because the meeting records of the Board of Trustees clearly have enduring value, and because the project team had easy access to copies of the media. The organization of files on the CD represents an interesting appraisal situation by itself. Records are organized into directories by committee, and they sometimes have additional subdirectories. Most text documents have the original MS Word file along with a PDF version of the same file. Finally, there is an additional video on each CD, in which the University President describes events occurring around the Board meeting and points out items of particular note on the meeting agenda.

We decided to preserve each packet of documents that compose a record of a Board meeting—really a series of meetings of the full Board and its various committees held over the course of two days—as a single complex object containing all the data on the CD. The project team made this decision because:

1. The filesystem on the CD represented a significant element of the original order of the materials
2. The filesystem standard itself (ISO9660) is widely used, standardized, and quite preservable
3. There is little need to format-shift the bulk of the materials on the CD, because they are represented in both their original formats (MS Word) and in the more preservable PDF format.

As the project team had little experience with video formats, it felt it could not take any meaningful preservation action for the video at the time of the ingest project.

The project team took bit-for-bit copies of the CDs, using standard facilities found in Mac OS X. The team loaded these images from disk to ensure that they were not corrupted during the copy. The team packaged each individual filesystem image into a SIP and submitted to the Tufts Ingest Prototype System (TIPS),[7] which them performed the checks from Part B.2 of the Ingest Guide (however, with very minimal verification) and then created AIPs. However, the project team encountered the problem that Fedora would not accept records components over a certain size.

[7] See 2.3 Ingest Tools for a description of the application.

The meeting packets are all about 300 MB in size, and all caused ingest problems. Thus, we were unable to complete the ingest of any of the packets.

Survey Report

Background

Eliot Wilczek, University Records Manager, Digital Collection and Archives (DCA) met with Lydia Evans, Secretary of the Faculty, on 04/22/05 to discuss a variety of records management issues concerning Board of Trustees records. Lydia Evans indicated that the Office of the Board of Trustees has stored meeting records from 2000 through 2004 on CDs.

Lydia Evans agreed to give Eliot Wilczek two CDs with two sets of meeting records covering Trustees meetings from May 2003 and February 2004 so the Digital Collections and Archives (DCA) could produce this survey report on the Trustees' meeting records stored on CDs. She gave Eliot Wilczek the CDs on the day of their meeting.

Description of Records Surveyed

General Description

These are the meeting records of Board of Trustees of Tufts University. Each CD contains a website that helps a Trustee navigate through the meeting records, a video of the University President welcoming the Board members and giving an overview of the upcoming meetings, an agenda and schedule of the Board meetings and events (a Board "meeting" is really composed of several meetings and events that usually occur over two days), contact information for the Trustees and background information about Tufts.

Producer

Office of the Board of Trustees

Producer Role(s) Creator

Function(s)

The Office of the Board of Trustees sends Board members a set of meeting records. From 2000 to 2004 the Office mailed each Board member a CD of the meeting records for an upcoming meeting. The meeting records give a Board member the information he or she needs to make votes at Board meeting in an informed manner.

Record Type(s)

Meeting Records

Format Type(s)

Web pages composed in html.

For most pages in `<head></head>`

```
<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-
1">
<meta name="GENERATOR" content="Mozilla/4.79 [en]C-CCK-MCD {C-UDP;
Tufts University granite 2/11/2002}  (Windows NT 5.0; U) [Netscape]">
<title>index</title>
```

Valid HTML 4.0 Transitional.

Validation used: W3C Markup Validation Service v0.7.0

http://validator.w3.org accessed 04/29/05.

PDF files.
No validation performed.

JPG image files.
No validation performed.

Microsoft Word Documents.
No validation performed.

MOV video file.
No validation performed.

Microsoft Excel Worksheet file.
No validation performed.

Executable programs—installers.
No validation performed.

File Size
February 2004 meeting: 294 MB.
May 2003 meeting: 257 MB.

Producer Identifier
All of meeting records and all of their components do not have producer identifiers that need preservation.

Confidentiality Requirements/Access Restrictions
All of the records have requirements for access restrictions. They should all have Category 1: Confidential University Records.

Copyright Status
Copyright of all records held by Tufts University.

List of Records Surveyed
Board of Trustees Meeting Records

Description	Documents Trustees receive before Board meetings to help they make informed decisions. Meeting records are composed of a variety of other record types, usually reports, meeting agendas, and meeting minutes. Records of a Board meeting—really a series of meetings of the full Board and its various committees held over the span of two days—are bundled together on a CD. In this context, these records compose meeting records. The meeting records have two copies of all text documents, one copy in Word format and one copy in PDF.
Record Type	Meeting Records
Dates	2003-2004

Format HTML, PDF, WORD, JPEG, XSL, MOV

Evaluation of Recordkeeping System

All files are stored on CDs. The Office of the Board of Trustees distributes use copies to members of the Board and senior administrators. The Office maintains record copies. The Office carefully creates and manages these records but does not have written procedures for its recordkeeping of these records.

Evaluation of Authenticity

The Digital Collections and Archives judges the meeting records and all of the components of the May 2003 and February 2004 Board meetings on the CDs the Office of the Board of Trustees gave to the DCA to be authentic for the following reasons:

- At the 04/22/05 meeting, Lydia Evans declared that the Office of the Board of Trustees created the May 2003 and February 2004 meeting records in the normal course of its business.
- At the 04/22/05 meeting, Lydia Evans declared that from 2000 to 2004 the Office of the Board of Trustees stored meeting records on CDs in the normal course of its recordkeeping activities.
- Shortly after the 04/22/05 meeting, Eliot Wilczek briefly reviewed the files on the CDs that Lydia Evans gave to him and determined that they contained the May 2003 and February 2004 meeting records as she claimed.
- The probability someone would maliciously alter the records—particularly its content—and try hide that content alteration is extremely low because the opportunity and motivation to undertake such an action are both low.
- The DCA should be able presume the authenticity of meeting records the Office of the Board of Trustees directly gives to the DCA on CDs.

Appraisal Decision

Meeting Records

Disposition	Transfer to DCA
Reason for Disposition	Retention Schedule gen091
Essential Elements	
Documentary Form	
Content	*Verbatim*
Presentation	Textual Structure: *Verbatim*
	Video Structure: *General Appearance*
	PDF Structure: *General Appearance*
	Word Structure: *None*
	Excel Structure: *General Appearance—no calculations*
	HTML Structure: *General Appearance*
Signs	None
Annotations	None
Context	*Information* that delivered on CDs to Board members before board meetings

Medium	*Information* that HTML, PDF, MOV, Word, and Excel stored on CDs

Ingest Guide Steps: Archive and Producer Actions		
Steps	**Archive Actions**	**Producer Actions**
A Negotiate Submission Agreement		
A1 Establish Relationship		
A1.1 Initiate Contact/Ingest Project	The Archive had previously worked with the Producer on a variety of Ingest Projects and other work. The Archive contacts the Producer about Trustee meeting records stored on CDs.	The Producer had previously worked the Archive on a variety of Ingest Projects and other work. The Archive contacts the Producer about Trustee meeting records stored on CDs.
A1.2 Identify Producer	The Archive identifies the Office of the Board of Trustees as the Producer.	
A1.3 Has the Archive already defined its relationship with the Producer?	Yes, the Archive already defined its relationship with the Producer during previous Ingest Projects. *The Archive skips Steps A1.4 through A1.6.*	
A2 Define Project		
A2.1 Identify records at issue, agreeing upon scope of survey	The Archive and Producer identify the meeting records of the Full Board of Trustees and its various committees for its May 2003 and February 2004 meetings as the records of this Ingest Project.	The Archive and Producer identify the meeting records of the Full Board of Trustees and its various committees for its May 2003 and February 2004 meetings as the records of this Ingest Project.
A2.2 Does the Producer have custody/authority over the records identified in A2.1?	Yes, the Archive determines that the Office of the Board of Trustees has the appropriate authority over the records identified in Step A2.1. *The Archive skips Steps A2.3 through A2.6.*	
A3 Collect Information and Assess Value of Records		
A3.1 Conduct Records Survey, note attributes of records	The Archive conducts a survey and produces a **Survey Report**.	The Producer provides the Archive with copies of the CDs.
A3.2 Judge authenticity of records	The Archive examines content of records and received the CDs directly from the Producer and determines that they are very probably authentic, and updates the **Survey Report**.	
A3.3 Should the Archive accession at least some of the records?	Yes, the Archive determines that the meeting records have archival value, and judges them to be authentic. The Archive updates the **Survey Report** to reflect this decision.	

A3.4 Determine the essential elements of the records that should be accessioned	The Archive examines records and determines that the essential elements of the meeting records are the content and formatting of the PDF, Word, and Excel files; the content, formatting, and functionality of the HTML wrapper; and the content of the video. The Archive adds its determinations to the **Survey Report**.	
A4 Assess Record Type		
A4.1 Are all records identified as a Record Type?	Yes, the Archive determines that records surveyed include meeting records. This record type is already known to the Archive. The Archive creates a **Record Type List** in the Submission Agreement which references **Record Type Records.** *The Archive skips Step A4.2.*	
A5 Assess Formats		
A5.1 Are any records in file formats that are not a preservation format?	Yes, the Archive determines the following: The meeting record consists of PDF, Word, Excel, MOV formats and the HTML wrapper consists of valid HTML and JPEG images. According to the Formats Standards Policy, PDF, JPEG, and HTML are preservation formats, while Word, Excel, and MOV are not. For the Word, Excel, and MOV files the Archive produces a **Format Transformation Plan**.	
A5.2 Should Archive transform or natively handle these formats?	The Archive will transform the Excel files, and handle the Word, MOV, PDF, JPEG, and HTML files natively. *Archive skips Steps A5.3 through A5.4.*	
A5.3 Identify needed Representation Information for new preservation format	The Archive will add the MOV image format to its **Formats Standards Policy** and create **Representation Information** for the format, adding it to its **Format Representation Information System**. Although it will handle the Word files natively, it will no make no effort to preserve these files because the records that exist in Word format also exist in PDF format. Therefore, the Archive will not add Word to its **Formats Standards Policy** or create **Representation Information** for the format.	
A5.5 Choose appropriate format	The Archive will transform the Excel files into comma-separated values format (CSV) files. The Archive updates its **Format Transformation Plan** to reflect this decision.	
A5.6 Is format chosen in A5.5 a Preservation Format the Archive already uses?	Yes, the Archive determines that the CSV format is a preservation formats based on its **Formats Standards Policy**.	
A6 Assess Identifier Rules		

A6.1 Is there a Producer naming/ identification scheme that needs accommodation?	No, the Archive determines there is no **Producer Naming/ Identification Scheme** that it needs to preserve or accommodate. The *Archive skips Steps A6.2 through A6.3.*	
A6.4 Determine appropriate naming/ identification scheme(s).	The Archive decides to use the standard **Archive Naming/ Identification Scheme**. The Archive records its **Naming/Identification Scheme Decision**.	
A7 Assess Copyright		
A7.1 Determine copyright status of records in Ingest Project	The Archive determines that the records are under copyright of Tufts University. The Archive documents the **Copyright status**.	
A7.2 Does the Archive need to acquire copyright or license for records?	No, the Archive determines that the University holds copyright. *Archive skips Steps A7.3 through A7.8.*	
A8 Assess Access Rights		
A8.1 Determine records' Records Security Profile	The Archive determines that the records are administrative records and are closed for 75 years from the date of creation and should have a **Records Security Profile** of closed access. The Archive documents this decision as a **Records Security Profile Decision** which references the **Records Security Profile**.	
A8.2 Does current security component meet the access control needs of the records?	Yes, the Archive determines the security component of the Preservation System meets the needs of the **Records Security Profile**. *Archive skips Steps A8.3 through A8.7.*	
A9 Assess Recordkeeping System		
A9.1 Has the Archive documented recordkeeping system as supporting feasible and trustworthy transfer?	No, the Archive has not previously examined the CDs and the Producer's method of managing the records on the CDs.	
A9.2 Can recordkeeping system support feasible and trustworthy transfer?	The Archive determines that the recordkeeping system can support the trustworthy transfer of records to the Archive, but not in a scalable manner. It can support a feasible transfer if the volume of records is low. The Archive produces a **Recordkeeping System Report.** *The Archive skips Step A9.3.*	
A9.4 Is Archive or Producer willing to take extraordinary measures to transfer records?	The Archive determines that it is willing to manually transfer the records from the recordkeeping system to the Archive because of the small volume of records involved in the Ingest Project. The Archive produces **SIP Creation**	

	Procedures for moving the records from the recordkeeping system to the Archive. *The Archive Steps A9.5 through A9.7.*	
A10 Assess Feasibility		
A10.1 Can the Archive feasibly accession the records?	Yes. The Archive determines that it is feasible to manage and preserve the records and its requirements without extraordinary effort or special accommodation. The Archive produces a **Preservation System Availability Statement**. *The Archive skips Steps A10.2 through A10.5.*	
A11 Finalize Submission Agreement		
A11.1 Add description of Metadata Encoding Rules to Submission Agreement.	The Archive chooses standard Dublin Core metadata encoding and documents its **Metadata Encoding Rules Decision**.	
A11.2 Add description of Transfer Procedures to Submission Agreement	The Producer has provided copies of the CDs to the Archive. The Archive will create the SIPs from the CDs at a convenient time. The Archive documents its **Transfer Procedures Decision** in the Submission Agreement.	The Office of the Board of Trustees has provided CDs to the Archive.
A11.3 Add description of Validation Procedures to Submission Agreement	The Archive documents the validation procedures for HTML, JPEG, PDF, Word, Excel, and MOV files in the **Validation Procedures Decision**.	
A11.4 Add Transfer Schedule to Submission Agreement	The Archive selects an indefinite schedule, and documents this decision in the **Transfer Schedule Decision**.	
A11.5 Add SIP Creation Procedures to Submission Agreement	The Archive chooses its standard SIP format and documents its **SIP Creation Procedures Decision**.	
A11.6 Finalize Submission Agreement	The Archive draws up the **Draft Submission Agreement** based on its previous decisions.	
A11.7 Does Archive and Producer agree to and approve the Submission Agreement?	Yes, the Archive and Producer agree and produce the **Finalized Submission Agreement**. The Archive submits finalized submission agreement to the ingest system. The Archive's ingest application accepts and validates a machine-readable version of the **Finalized Submission Agreement.** *Archive skips Steps A11.8 through A11.10.*	Yes, the Producer and Archive agree and produce the **Finalized Submission Agreement**.
B **Transfer and Validation**		
B1 **Create and Transfer SIPs**		

B1.1 Producer prepares SIP according to Submission Agreement		The Producer gathers the CDs containing meeting records and ensures that the CDs contain the appropriate records according to the **SIP Creation Procedures Decision.**
B1.2 Producer transfers the SIP to the Archive		The Producer manually delivers CDs containing meeting records to the Archive according to the **Transfer Procedures Decision**.
B2 **Validate**		
B2.1 Archive receives SIP from Producer	The Archive accepts the SIP from the Producer and extracts the data from the CDs as TAR files and places the files into the ingest application. The Archive then produces a **Documentation of Receipt** and delivers that to the Producer.	The Producer receives the **Documentation of Receipt**.
B2.2 Is SIP well formed?	The ingest application checks the SIP format. The SIP is well-formed. The Application updates the **SIP validity statement.**	
B2.3 Does SIP contain malicious code?	The ingest application scans the SIP components for viruses and other malicious code. All SIP components are clean. The application updates the **SIP validity statement**.	
B2.4 Is the submitter authorized to submit SIP to the Archive?	The ingest application validates the SIP signatures and validates identities against its database of certificates. The SIP is signed by an authorized person. The application updates the **SIP validity statement**.	
B2.5 Does SIP contain all necessary records components?	The ingest application checks all included records for completeness. All records in the SIP are complete. The application updates the **SIP validity statement**.	
B2.6 Do the record components in SIP validate?	The ingest application tests the record components for validity, where necessary. All record components validate. Application updates the **SIP validity statement**.	
B3 **Transform and Attach Metadata**		
B3.1 Do any of the records in SIP require transformation?	Yes, the Archive determines that the Excel files require transformation to CSV files according to the **Format Transformation Plan** in the Submission Agreement.	
B3.2 Perform transformation on records that require transformation	The Archive transforms the Excel files to CSV producing **Transformed Records**.	
B3.3 Attach to records metadata inferred from Submission Agreement	The ingest application attaches stock metadata from Submission Agreement, **Producer Record**, and **Record Type Records**, creating **Records with Attached Metadata**.	

B3.4 Attach to records the Records Security Profile defined by Submission Agreement	The ingest application attaches the **Record Security Profile** identified in the Submission Agreement, creating a **Records with Security Profile**.	
B4 **Formulate AIPs**		
B4.1 Formulate AIPs	The Ingest application creates an **AIP** for each record.	
B5 **Assess AIPs**		
B5.1 Are all of the records in the AIP part of accession described by Submission Agreement?	The Archive verifies that all of the records to be accessioned come from the CDs and produces an **AIP Validity Statement**. The ingest application accepts the **AIP Validity Statement**. *The Archive skips Steps B5.2 through B5.3.*	
B5.4 Is proper metadata attached to records in the AIP?	The Archive verifies that all of the records have sufficient and correct metadata and updates the **AIP Validity Statement**. The ingest application accepts the **AIP Validity Statement**. *The Archive skips Step B5.5.*	
B6 Formally Accession		
B6.1 Submit AIPs into Preservation Repository.	The ingest application submits the AIPs to the Preservation Repository.	
B6.2 Formally notify Producer that Archive has accepted and accessioned records described by Ingest Project.	The ingest application generates a **Transfer Notice** for the Producer and an entry in the **Accession Log**.	The Producer receives the **Transfer Notice**.

INGEST PROJECT THREE
University Library Council and University Library Council Teams

Ingest Project Narrative

The University Libraries Council (ULC) at Tufts University is composed of the directors of the six libraries at the Tufts. The ULC establishes and manages University-wide library policies. The ULC also oversees a number of inter-library teams that attend to a variety of library functions and issues. These teams include Staff Development, Fair Use, and Copier Contract Compliance, among others. The teams store their working documents, meeting minutes, bylaws, policies, and other documents in a web collaboration environment called SharePoint® Team Services (http://lib.tufts.edu), which acts as a recordkeeping system. The documents stored in lib are mostly MS Word documents with occasional Excel spreadsheets, HTML files, and plain text files. ULC and team members have read/write access to the records in SharePoint.

The project team chose this accession because it presents issues with troublesome format types and with records stored in a troublesome recordkeeping system. This is a situation common to recordkeeping systems at many universities.

The project team contacted Carol Johnson—director of University Library Technology Services, which manages the SharePoint instance for the ULC—about participating in the grant project. She was amenable to the pilot project. In order to accession the records of the lib system, the project team set up a comparable hardware and OS environment to the system which runs lib. Then the team developed tools which would allow us to extract records from the lib system. The team aimed to replicate lib onto its own hardware before running the extraction utilities to avoid running relatively untested code on their production server. It turned out that replicating the SharePoint instance was not easy. Furthermore, SharePoint stores almost no metadata, so preparing records for ingest required significant metadata reconstruction.

The file formats in this accession present a difficult but all-too-common problem; the data is largely stored in a proprietary format with no easy transformation route. Because the project team wished to model scalable processes, manual transformation (via loading into MS Office and saving in a different format) was not an option. Furthermore, such transformations usually lose information. The OpenOffice.org (OOo) project is able to correctly open and render a large class of MS Office documents and save them in a well-documented format, but again the project team wanted to avoid manual transformations. However, OpenOffice.org has a programming interface that allows interaction-free access to loading and saving capabilities. Unfortunately, this interface is clumsy and OOo was never designed to be used in an interaction-free way. All this make the process of using OOo difficult, but it appears to be the best option.

The project team acquired the data from the SharePoint instance and loaded it onto a server under our control. From there the team extracted the records of interest using custom-built tools to create a number of SIPs. Because the SharePoint system does not maintain metadata that is necessary to handle the records in an archival setting, the SIP extraction process involved applying rules to create metadata.

The project team fed these SIPs into the Tufts Ingest Prototype System (TIPS)[8] which is able to accession the records into a Fedora system in a limited way. The team used the OpenOffice suite to perform office document transformation. However, the version OpenOffice which the team configured to use with TIPS is an older version which does not support the target format (OpenDocument). Instead the team transformed the Word and Excel documents into PDF documents. Furthermore, because of limitations in Fedora 2.0, the project team had to encode all complex objects in Base64 and wrap them in an xml tag. This makes retrieving the files from Fedora cumbersome.

[8] See 2.3 Ingest Tools for a description of the application.

Survey Report

Background

Eliot Wilczek, University Records Manager, Digital Collection and Archives (DCA) called Carol Johnson, Director, University Library Technology Services (ULTS) on 06/08/05 to discuss the records on the SharePoint Team Services site at http://lib.tufts.edu. Lib serves as the recordkeeping system for the records of the University Library Council (ULC) and its teams. Lib also serves as the recordkeeping system for some ULTS records.

During this conversation Carol Johnson verbally agreed to allow the DCA to survey the records on Lib and transfer any records that it appraises to have archival value to the DCA.

On 10/28/05 Eliot Wilczek conducted a survey of the records on Lib. Mr. Wilczek excluded the ULTS records from the survey because they appeared to be a small and incomplete set of ULTS' records.

Description of Records Surveyed

General Description

SharePoint Team Services is the recordkeeping system for the records of the ULC and its teams, which include:

Collections and Licensing
Copier Contract and Compliance
Electronic Dissertations & Theses
Fair Use
Integrated Library System Implementation
Publicity and Marketing
Staff Development
Services Steering

The ULC and the teams go through a self-selection process when they post records to Lib. A high percentage of the records in Lib have archival value, but the records in Lib do not represent all of the records of the ULC and its teams.

Producer

University Library Technology Services

Producer Role(s) Custodian

Notes The *Creator* of these records is the University Library Council and its teams listed above. The ULC create and direct the teams—the teams are not independent entities. ULTS manages the records on behalf of the ULC and its teams and has the authority to execute their disposition. *Also note* as Director of the ULTS, Carol Johnson is a member of the ULC.

Function(s)

The records on Lib document the important decisions and actions of the ULC and its teams. Some records also serve as working files for the teams and ULC. Posting the records onto Lib allows all members of the teams and the ULC to share these records with each other.

Record Type(s)
Charges
Reports
Meeting Minutes
Subject Files

Format Type(s)
SharePoint Team Services presented as web pages via http protocol.

For all pages—based on small sample of pages
`<html dir="ltr" xmlns:o="urn:schemas-microsoft-com:office:office">`
and
`<HEAD>`
`<META Name="GENERATOR" Content="Microsoft FrontPage 5.0">`

No validation performed.

MS Word files.

No validation performed.

MS Excel files.

No validation performed.

Web pages composed in html. This does not include external web pages not on Lib.

No validation performed.

Executable programs—installers.

No validation performed.

File Size
Most of the records are smaller than 1mb. Many of them are significantly smaller than 1 MB. However, Lib has a few executable files (installers) that are 20 to 40 MB in size.

Producer Identifier
The records do not have producer identifiers that need preservation.

Confidentiality Requirements/Access Restrictions
All of the records have requirements for access restrictions. They should all have Category 3: General University Records. Creators and the Producer can have immediate access.

Copyright Status
Copyright of all the records are held by Tufts University.

List of Records Surveyed
Charges

Description The ULC's charge to the various teams. All charges should be in "Charges" directories.

Record Type	Charges
Dates	1999 through 2005
Format	MS Word

Reports

Description	Mostly periodic but some subject-oriented reports created by the ULC and its teams. Most reports should be in "Annual Reports" directories, some are in "Shared Documents" or other directories.
Record Type	Reports
Dates	1999 through 2005
Format	MS Word

Meeting Minutes

Description	Written records of decisions made and actions taken by the ULC and its teams at their respective meetings. All meeting minutes should be in "Minutes" directories.
Record Type	Meeting Minutes
Dates	1999 through 2005
Format	MS Word

Subject Files

Description	A wide variety of records that often serve as the working files of the ULC and its teams. Most subject files should be in "Shared Documents" directories.
Record Type	Subject Files
Dates	1999 through 2005
Format	MS Word, MS Excel, HTML, Executable programs—installers

SharePoint Team Services Instance

Description	A web-based file-sharing system used as a recordkeeping system for records created by the ULC and its teams.
Record Type	Recordkeeping System
Dates	2002 through 2005
Format	SharePoint Team Services

Evaluation of Authenticity

The Digital Collections and Archives judges the records on the SharePoint Team Services instance at http://lib.tufts.edu, which it has evaluated in this survey, to be authentic for the following reasons:

- At the 06/08/05 survey interview, Carol Johnson indicated that ULTS installed and maintains the SharePoint instance on behalf of the ULC and its teams in the normal course of their business.
- At the 06/08/05 survey interview, Carol Johnson identified the SharePoint instance at http://lib.tufts.edu as the recordkeeping system for the ULC and its teams.

- During the 10/28/05 survey, Eliot Wilczek determined the SharePoint instance at http://lib.tufts.edu was the same that Carol Johnson identified on 06/08/05 based on a brief visual review of the SharePoint instance, its directory of files and the content of a sample of the records it contained.
- The probability someone would maliciously alter the SharePoint instance or any of the individual records managed by that instance—particularly their content—and try to hide that content alteration is extremely low because no one has a reasonable motivation to undertake such an action.

Appraisal Decision

For all records that have a disposition of "Transfer to DCA," it is critical to preserve the information of the file path. The file structure—which serves as a de facto taxonomy—will provide essential information for post-processing descriptive metadata work, particularly assigning records to the correct collections and series.

This instance of SharePoint is an active system; appropriate staff members add records to Lib continuously and can update records already in Lib at anytime. The records are arranged with no chronological cut-offs. Appraisal of these records will have to occur at systematic time intervals so the DCA can track which records it has previously appraised and accessioned.

Charges

Disposition	Transfer to DCA
Reason for Disposition	Retention Schedule gen042
Essential Elements	
Documentary Form	
Content	*Verbatim*
Presentation	Textual Structure: *Verbatim*
	MS Word Structure: *General Appearance*
Signs	None
Annotations	None
Context	*Information* that managed and shared in SharePoint Team Services instance via http delivered as MS Word documents with restrictions limiting access to ULC and Team members.
Medium	*Information* that MS Word

Reports

Disposition	Transfer to DCA
Reason for Disposition	Retention Schedule gen065
Essential Elements	
Documentary Form	
Content	*Verbatim*
Presentation	Textual Structure: *Verbatim*
	MS Word Structure: *General Appearance*
Signs	None

- Annotations: None
- Context: *Information* that managed and shared in SharePoint Team Services instance via http delivered as MS Word documents with restrictions limiting access to ULC and Team members.
- Medium: *Information* that MS Word

Meeting Minutes
- Disposition: Transfer to DCA
- Reason for Disposition: Retention Schedule gen025
- Essential Elements
  - Documentary Form
    - Content: *Verbatim*
    - Presentation: Textual Structure: *Verbatim*; MS Word Structure: *General Appearance*
    - Signs: None
  - Annotations: None
  - Context: *Information* that managed and shared in SharePoint Team Services instance via http delivered as MS Word documents with restrictions limiting access to ULC and Team members.
  - Medium: *Information* that MS Word

Subject Files
- Disposition: Transfer to DCA. Many Subject Files do not have permanent value but are so thoroughly mixed in with subject files of archival value that all subject files should be transferred to the DCA. The DCA will make no effort to preserve the Installers because none of them have archival value.
- Reason for Disposition: Retention Schedule gen024
- Essential Elements
  - Documentary Form
    - Content: *Verbatim,* includes *Data Integrity* for records in MS Excel
    - Presentation: Textual Structure: *General Appearance*; MS Word Structure: *General Appearance*; MS Excel Structure: *General Appearance*; HTML Structure: *General Appearance* *Information* of hyperlink URL, hyperlink functionality not needed; Installers: *None*
    - Signs: None
  - Annotations: None
  - Context: *Information* that managed and shared in SharePoint

	Team Services instance via http delivered as MS Word documents with restrictions limiting access to ULC and Team members.
Medium	*Information* that either MS Word, MS Excel, HTML, or executable applications—installers.
Post-Processing Notes	Determine if DCA needs to remove and destroy non-archival records from subject files.

SharePoint Team Services instance

Disposition	Retire when no longer needed and all records it contains are transferred to the DCA or other disposition has been executed. Destroy all data when instance retired. However, retain *information* that the Producer stored and managed the records described above in a SharePoint Team Services instance at http://lib.tufts.edu with read/write access for members of the ULC and its teams.
Reason for Disposition	The SharePoint instance in and of itself is not a record and its functionality is not an essential element of any of the record it contains.
Essential Elements	
Documentary Form	
Content	None
Presentation	None
Signs	None
Annotations	None
Context	I*nformation* that Producer stored and managed surveyed records in a SharePoint Team Services instance at http://lib.tufts.edu with read/write access for members of the ULC and its teams.
Medium	None

Ingest Guide Steps: Archive and Producer Actions		
Steps	**Archive Actions**	**Producer Actions**
A Negotiate Submission Agreement		
A1 Establish Relationship		
A1.1 Initiate Contact/Ingest Project	The Archive contacts the Director of the University Library Technology Services (ULTS) to initiate an Ingest Project concerning the records of University Library Council and its teams stored and managed on an instance of SharePoint Team Services (lib.tufts.edu).	The Director of the University Library Technology Services (ULTS) responds to the Archive's contact and agrees to initiate an Ingest Project.
A1.2 Identify Producer	The Archive identifies the ULTS as the Producer.	
A1.3 Has the Archive already defined its relationship with the Producer?	No, the Archive has not defined its relationship with the Producer.	
A1.4 Is this the Appropriate Archive?	Yes, the Archive determines it is the appropriate Archive for the ULC, its teams, and the ULTS. *The Archive skips Step A1.5.*	
A1.6 Collect and document information about Producer	The Archive collects information about ULC, its teams, and the ULTS and creates a **Producer Record** for each entity, and produces a **Producer Entry** in the Submission Agreement which references the appropriate **Producer Record**.	
A2 Define Project		
A2.1 Identify records at issue, agreeing upon scope of survey	The Archive and Producer identify the records of the ULC and its teams in the SharePoint instance at http://lib.tufts.edu managed by the ULTS as the records of this Ingest Project.	The Archive and Producer identify the records of the ULC and its teams in the SharePoint instance at http://lib.tufts.edu managed by the ULTS as the records of this Ingest Project.
A2.2 Does the Producer have custody/authority over the records identified in A2.1?	Yes, the Archive determines that the ULTS has authority over the records identified in Step A2.1 because it manages the records on behalf of the ULC and its teams and has the authority to execute their disposition. *The Archive skips Steps A2.3 through A2.6.*	
A3 Collect Information and Assess Value of Records		
A3.1 Conduct Records Survey, note attributes of records	The Archive conducts a survey and produces a **Survey Report**.	The Producer gives the Archive access to the SharePoint instance so it can conduct its survey.

A3.2 Judge authenticity of records	The Archive examines content of records in the SharePoint instance and determines that they are very probably authentic, and updates the **Survey Report**.	
A3.3 Should the Archive accession at least some of the records?	Yes, the Archive determines that the ULC and team records have archival value, and judges them to be authentic. The Archive updates the **Survey Report** to reflect this decision.	
A3.4 Determine the essential elements of the records that should be accessioned	The Archive examines records and determines that the essential elements of the minutes, reports, subject files, and charges are their content, textual formatting, and the information that they were managed in a SharePoint Team Services recordkeeping system with read and write access to members of the ULC and its teams. The Archive adds its determinations to the **Survey Report**.	
A4 Assess Record Type		
A4.1 Are all records identified as a Record Type?	Yes, the Archive determines that records surveyed include Meeting Minutes, Charges, Reports, and Subject Files. All are record types already known to the Archive. The Archive creates a **Record Type List** in the Submission Agreement which references **Record Type Records.** *The Archive skips Step A4.2.*	
A5 Assess Formats		
A5.1 Are any records in file formats that are not a preservation format?	Yes, the Archive determines the following: MS Word, MS Excel, and non-valid HTML are not preservation formats according to the **Formats Standards Policy**. For the Word, Excel, and non-valid HTML files the Archive produces a **Format Transformation Plan**.	
A5.2 Should Archive transform or natively handle these formats?	The Archive will transform the non-valid HTML, Word, and Excel files. *Archive skips Steps A5.3 through A5.4.*	
A5.5 Choose appropriate format	The Archive will transform the non-valid HTML into valid XHTML; it will transform the Word into PDF; it will transform the Excel into PDF. The Archive updates its **Format Transformation Plan** to reflect this decision.	
A5.6 Is format chosen in A5.5 a Preservation Format the Archive already uses?	Yes, the Archive determines that valid XHTML and PDF is a preservation format based on its **Formats Standards Policy**.	
A6 Assess Identifier Rules		

A6.1 Is there a Producer naming/ identification scheme that needs accommodation?	No, the Archive determines there is no **Producer Naming/ Identification Scheme** that it needs to preserve or accommodate. The *Archive skips Steps A6.2 through A6.3.*	
A6.4 Determine appropriate naming/ identification scheme(s).	The Archive decides to use the standard **Archive Naming/ Identification Scheme**. The Archive records its **Naming/Identification Scheme Decision**.	
A7 Assess Copyright		
A7.1 Determine copyright status of records in Ingest Project	The Archive determines that the records are under copyright of Tufts University. The Archive records the **Copyright status**.	
A7.2 Does the Archive need to acquire copyright or license for records?	No, the Archive determines that the University holds copyright. *Archive skips Steps A7.3 through A7.8.*	
A8 Assess Access Rights		
A8.1 Determine records' Records Security Profile	The Archive determines that the records are administrative records and are closed for 20 years from the date of creation and should have a **Records Security Profile** of closed access. The Archive documents this decision as a **Records Security Profile Decision** which references the **Records Security Profile**.	
A8.2 Does current security component meet the access control needs of the records?	Yes, the Archive determines the security component of the Preservation System meets the needs of the **Records Security Profile**. *Archive skips Steps A8.3 through A8.7.*	
A9 Assess Recordkeeping System		
A9.1 Has the Archive documented recordkeeping system as supporting feasible and trustworthy transfer?	No, the Archive has not previously examined the http://lib.tufts.edu instance of SharePoint and the Producer's method of managing the records on that SharePoint instance.	
A9.2 Can recordkeeping system support feasible and trustworthy transfer?	The Archive determines that the recordkeeping system can support the trustworthy transfer of records to the Archive, but not in a scalable manner. It can support a feasible transfer if the volume of records is low. The Archive produces a **Recordkeeping System Report.** *The Archive skips Step A9.3.*	
A9.4 Is Archive or Producer willing to take extraordinary measures to transfer records?	The Archive determines that it is willing to manually transfer the records from the recordkeeping system to the Archive because of the small volume of records involved in the Ingest Project. The Archive produces **SIP Creation**	

	Procedures for moving the records from the recordkeeping system to the Archive. *The Archive Steps A9.5 through A9.7.*	
A10 Assess Feasibility		
A10.1 Can the Archive feasibly accession the records?	Yes. The Archive determines that it is feasible to manage and preserve the records and its requirements without extraordinary effort or special accommodation. The Archive produces a **Preservation System Availability Statement**. *The Archive skips Steps A10.2 through A10.5.*	
A11 Finalize Submission Agreement		
A11.1 Add description of Metadata Encoding Rules to Submission Agreement.	The Archive chooses standard Dublin Core metadata encoding and documents its **Metadata Encoding Rules Decision**.	
A11.2 Add description of Transfer Procedures to Submission Agreement	The Producer has allowed the Archive to create an inactive copy of the http://lib.tufts.edu SharePoint instance on a server controlled by the Archive. The Archive will create the SIPs from the SharePoint instance it manages at a convenient time. The Archive documents its **Transfer Procedures Decision** in the Submission Agreement.	The Producer allows the Archive to create an inactive copy of the http://lib.tufts.edu SharePoint instance on a server controlled by the Archive.
A11.3 Add description of Validation Procedures to Submission Agreement	The Archive documents the validation procedures for HTML, Word, and Excel files in the **Validation Procedures Decision**, which becomes part of the **Submission Agreement**.	
A11.4 Add Transfer Schedule to Submission Agreement	The Archive selects an indefinite schedule, and documents this decision in the **Transfer Schedule Decision**.	
A11.5 Add SIP Creation Procedures to Submission Agreement	The Archive chooses its standard SIP format and documents its **SIP Creation Procedures Decision**.	
A11.6 Finalize Submission Agreement	The Archive draws up the **Draft Submission Agreement** based on its previous decisions.	
A11.7 Does Archive and Producer agree to and approve the Submission Agreement?	Yes, the Archive and Producer agree and produce the **Finalized Submission Agreement**. The Archive submits finalized submission agreement to the ingest system. The Archive's ingest application accepts and validates a machine-readable version of the **Finalized Submission Agreement.** *Archive skips Steps A11.8 through A11.10.*	Yes, the Producer and Archive agree and produce the **Finalized Submission Agreement**.
B **Transfer and Validation**		

B1 **Create and Transfer SIPs**		
B1.1 Producer prepares SIP according to Submission Agreement		The Producer allows the Archive to make an inactive copy the Producer's http://lib.tufts.edu instance of SharePoint on a server managed by the Archive. The Archive then extracts records from the SharePoint instance it controls and constructs the SIPs. The Archive signs the SIP with its own digital signature.
B1.2 Producer transfers the SIP to the Archive		The Archive places the SIP in the Ingest application drop-box.
B2 **Validate**		
B2.1 Archive receives SIP from Producer	The Ingest application accepts the SIP from the drop-box and produces **Documentation of Receipt** and delivers that to the Producer.	The Producer receives the **Documentation of Receipt**.
B2.2 Is SIP well formed?	The ingest application checks the SIP format. The SIP is well-formed. The Application updates the **SIP validity statement.**	
B2.3 Does SIP contain malicious code?	The ingest application scans the SIP components for viruses and other malicious code. All SIP components are clean. The application updates the **SIP validity statement**.	
B2.4 Is the submitter authorized to submit SIP to the Archive?	The ingest application validates the SIP signatures and validates identities against its database of certificates. The SIP is signed by an authorized person. The application updates the **SIP validity statement**.	
B2.5 Does SIP contain all necessary records components?	The ingest application checks all included records for completeness. All records in the SIP are complete. The application updates the **SIP validity statement**.	
B2.6 Do the record components in SIP validate?	The ingest application tests the record components for validity, where necessary. All record components validate. Application updates the **SIP validity statement**.	
B3 **Transform and Attach Metadata**		
B3.1 Do any of the records in SIP require transformation?	Yes, the Archive determines that the HTML files require tidying and transformation to valid XHTML and the Word and Excel files require transformation to PDF according to the **Format Transformation Plan** in the Submission Agreement.	
B3.2 Perform transformation on records that require transformation	The Archive tidies and transforms the HTML files into valid XHTML and transforms the Word and Excel files into PDF files, producing **Transformed Records**.	

B3.3 Attach to records metadata inferred from Submission Agreement	The ingest application attaches stock metadata from Submission Agreement, **Producer Record**, and **Record Type Records**, creating **Records with Attached Metadata**.	
B3.4 Attach to records the Records Security Profile defined by Submission Agreement	The ingest application attaches the **Record Security Profile** identified in the Submission Agreement, creating a **Records with Security Profile**.	
B4 **Formulate AIPs**		
B4.1 Formulate AIPs	The Ingest application creates an **AIP** for each record.	
B5 **Assess AIPs**		
B5.1 Are all of the records in the AIP part of accession described by Submission Agreement?	The Archive verifies that all of the records to be accessioned come from the SharePoint instance at http://lib.tufts.edu and produces an **AIP Validity Statement**. The ingest application accepts the **AIP Validity Statement**. *The Archive skips Steps B5.2 through B5.3.*	
B5.4 Is proper metadata attached to records in the AIP?	The Archive verifies that all of the records have sufficient and correct metadata and updates the **AIP Validity Statement**. The ingest application accepts the **AIP Validity Statement**. *The Archive skips Step B5.5.*	
B6 Formally Accession		
B6.1 Submit AIPs into Preservation Repository.	The ingest application submits the AIPs to the Preservation Repository.	
B6.2 Formally notify Producer that Archive has accepted and accessioned records described by Ingest Project.	The ingest application generates a **Transfer Notice** for the Producer and an entry in the **Accession Log**.	The Producer receives the **Transfer Notice**.

Fedora and the Preservation of University Records Project

2.3 Ingest Tools

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/

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:UA069.004.001.00008
(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

TABLE OF CONTENTS

OVERVIEW

Although it was not the original focus of research, the project team utilized the expertise of its members to undertake some development work into some of the specific tools needed to support the accessioning of electronic records into a preservation repository. Such tools are necessary to make Fedora more suitable for preservation of university electronic records and more compliant with OAIS model specifications. The work was undertaken separately at both Tufts and Yale under the supervision of Robert Dockins, the Project Analyst. Two particular tools are described below in detail.

TUFTS INGEST PROTOTYPE SYSTEM

In order to support the work described in "Ingest Projects," the Tufts project team developed a prototype ingest system which automates many steps of the ingest process. The system, known as the Tufts Ingest Prototype System (TIPS), allowed the project team to gain experience with the processes involved in production-scale ingests. The project team developed TIPS concurrently with the Ingest Guide, allowing theoretical knowledge and practical experience to inform the development of the Guide and the tool.

TIPS is available for use under the Mozilla Public License Version 1.1. It can be downloaded from http://dca.tufts.edu/features/nhprc/reports/tips/index.html. Archives can use TIPS to execute many of the tasks described in Section B Transfer and Validation of the Ingest Guide. However, this is not a production-ready tool with a polished user interface. It is a developer-oriented tool that can help archives develop scalable transfer and validation workflows within an ingest system. Although the project team used TIPS with a Fedora-based repository, it is not a Fedora-specific tool.

TIPS defines a Submission Information Package (SIP) format based on the popular info-zip compression format with an XML manifest. TIPS uses the manifest to group together related files into "digital objects," provides fixity information in the form of checksums, and allows the SIP creator to digitally sign the manifest. TIPS includes an API for creating and validating SIPs in this format. This SIP format is well-defined and an XML schema exists for the manifest.

The project staff created an XML format for submission agreements, although it is less well defined than the XML format for the SIP and there is no formal schema. Submission agreements are centered on the idea of "submission elements." Submission elements are sub-parts of a submission agreement representing a set of items in an accession that share certain properties, such as format types, record types, and access restrictions. Producers, sometimes with the help of archives, assign objects to particular submission elements during SIP creation. The decision about assignment to a submission element is largely an intellectual question. In the absence of rich metadata, it will probably need to be done manually; however, if a recordkeeping system has sufficient metadata, submission element selection may be automated.

The object components assigned to a particular submission element do not all have to have the same file format (however, they must all belong to a pre-determined set of file formats). To handle different format types, a submission element is associated with one or more "object profiles." Object profiles provide information about a format type, including the ability to recognize a file of that type and specific validation and transformation procedures for files of that format.

After a SIP is accepted by TIPS, each object is assigned to its submission element. Then, it is tested against each object profile in turn. The object is bound to the first profile it matches, which then determines the validation and transformation that object will undergo.

After a digital object is transformed and validated, it must be submitted to the preservation system. TIPS calls preservation systems “repositories” and defines an API for arbitrary repositories. We implemented this API for both Fedora 1.2.1 and Fedora 2.0 by binding to the API-M and API-A soap interfaces of Fedora.

Thus by writing object profiles, and combining them together into a submission agreement, the project team was able to define all the actions that should occur to an object before being submitted to a preservation repository.

TIPS does have a number of quirks. Most importantly, the project team constantly changed the design of the tool as the team built it concurrently with the development of the Ingest Guide, which itself underwent many revisions. There are a number of places were ideas were tried out and abandoned, but affected the way the project team shaped the code. Furthermore, object profiles are simply implemented as dynamic scripts that are run on demand. The scripting system is quite fragile and difficult to work with. The task of writing object profiles is complicated by the lack of a central repository of file format information with globally unique identifiers. Such a system would make it possible to register handlers that recognize, transform, and validate files of given, well-specified types; as it is the project team had to make do with what it had. Finally, TIPS is driven by an embedded workflow engine, with the idea that the operation of TIPS could be customized by writing custom workflows. However, the workflow integration greatly complicated TIPS for little additional benefit, and it interacts badly with the transaction management subsystem, leading to difficult-to-find bugs.

TIPS is therefore unwieldy and unsuited for large-scale use. It has been, however, a great help in discovering the issues that one will confront when attempting to build a highly customizable, highly automated system for ingesting electronic records in a trustworthy manner.

YALE EUDORA EMAIL INGEST TOOL

The Yale project team investigated the issues surrounding ingest of university records in the form of email, a problem of both file format and workflow common to a number of universities. In particular, the project team sought a solution to the scalability issues created by email stored in thousands of staff workstations using a proprietary software application that stores the email in a proprietary format.

Yale has a robust email operation, delivering over 500,000 messages everyday, excluding spam. The University has a diverse email environment with numerous email servers and many email software packages in widespread use. There is no single mandatory supported interface or application for utilizing the email services. AppleMail, Thunderbird, Eudora, Pine, Webmail, and the Outlook family of products are all widely used by faculty, staff and students on campus. These programs access the central email servers and store the resulting email messages in different manners that pose different issues for recordkeeping and preservation.

The most common method for connecting to email services from Yale staff workstations is using email application clients and the Post Office Protocol version 3 (POP3), an application-layer Internet standard protocol, to retrieve email from a remote server over a TCP/IP connection. POP3 allows users to retrieve email when connected and then to view and manipulate the retrieved messages without needing to stay connected. Although most client applications have an option to leave mail on server, Yale users employing POP3 clients generally connect, retrieve all messages, store them on their workstation as new messages, delete them from the server, and then disconnect. Far fewer Yale staff utilize, the Internet Message Access Protocol (IMAP) with their email applications. IMAP supports both *connected* and *disconnected* modes of operation. Those staff that do utilize IMAP generally leave messages on the server until the user explicitly deletes them. This is a reason that IMAP is more common in situations where multiple staff members share a mailbox. The fundamental difference between POP3 and IMAP when university records are concerned is that POP3 offers access to a mail drop; the mail exists on the server until it is collected by staff member's email client application. Even if the staff member leaves some or all messages on the server, the staff member's messages store is considered authoritative. In contrast, IMAP offers access to the central mail store; the staff member's client application may store local copies of the messages, but these are considered to be a temporary cache; the central server's store is authoritative.[1]

The project team was most interested in the issues posed by Yale staff using the Eudora email client application in a POP3 configuration. Other projects have focused on managing email at the central server[2] or with the Outlook family of applications.[3] Both of these solutions were

[1] Wikipedia contributors, "Post Office Protocol," *Wikipedia, The Free Encyclopedia*, <http://en.wikipedia.org/w/index.php?title=Post_Office_Protocol&oldid=77082274>.
[2] Marlan Green, Sue Soy, Stan Gunn, and Patricia Galloway, "Coming to TERM: Designing the Texas Email Repository Model," *D-Lib Magazine*, Volume 8, Number 9 (September 2002), <http://www.dlib.org/dlib/september02/galloway/09galloway.html>.
[3] Maureen Potter, "XML For Digital Preservation: XML Implementation Options for E-Mails," 2002 <http://www.digitaleduurzaamheid.nl/bibliotheek/docs/email-xml-imp.pdf#search=%22dutch%20testbed%20outlook%20email%22>.

insufficient for the current Yale environment because they either supposed a level of control over central recordkeeping that does not exist at Yale or require a level of recordkeeping intervention by record's creator that is not practical. It was determined that the most prominent university officers were using the Eudora email client application. The recordkeeping problem is that Eudora stores email not as individual files, but instead bound together with other files from the same mailbox. Every mailbox created in Eudora is stored as two different proprietary files that work together to give access to the individual emails. Because it is at this client application level that the email is set aside, organized, and stored as records, it is necessary to accession the email from the client application. This poses great sustainability problems. First, the file format of the mailbox files is proprietary and in no way would be considered a de jure or de facto standard. The decision was made that the file format would need to be converted to one more stable. The second issue is that the email is distributed across hundreds of computers on campus, requiring significant resources to maintain and/or transfer.

The Yale project team designed software to address the problems for Ingest created by the Yale use of Eudora. The concept of the operation is that the Archive enters into an agreement with Producer. They give the Archive access to their Eudora mailboxes, either on their workstation, or on a network folder. The Archive agrees to copy all messages from an agreed upon set of email folders (e.g. not inbox, outbox, sent, or trash). The Producer organizes their email folders and discards inconsequential messages (this operation would only apply to email that an employee sends or receives as part of his/her work at Yale). The Archive would periodically gain access to each Producer's Eudora mailboxes. Each mailbox file would be copied. Once copied, each mailbox file would be parsed to separate the out the mail messages. The separate mail messages would be saved as plain text and marked up in XML. Header information and other documentation would be copied into metadata. Attachments from the "Attachments" directory, are copied and linked to corresponding email message. All records would be transferred to either a compliant recordkeeping system or preservation system. The folder structure of the Producer's mailboxes is replicated in the recordkeeping and preservation system. When the operation is undertaken subsequently, messages already copied are ignored, and new messages are transferred. The tools were built as a series of scripts, in both Perl and Python, so that it could be easily configured and scheduled to run in an automated fashion. The path into Fedora utilized the Vital Batch Import module created by VTLS.[4] The toolset was designed in a modular fashion, relying on open source components, so that different features could be more easily swapped in and out during development. For example, the tool was utilized to output records to both the Yale's electronic recordkeeping application (Livelink), as well as the preservation application (Fedora).

While the development of this email ingest tool was helpful in understanding some of the issues of ingesting into Fedora, there are still a number of issues with the toolset in its current state. There is no user interface to the software. In order to configure or run the scripts a user must be comfortable enough working in the code itself. The scripts require access to the email folders of university staff members, often high-level staff, a level of access that the Yale University Archives is not normally granted. Also, the scripts must be either installed onto Producer workstations (which resulted in user permissions and scalability issues) or run from a central

[4] For more information on Fedora integration products offered by VTLS, see <http://www.vtls.com/Products/vital.shtml>.

administrative server (which resulted in processing and network traffic performance issues). These problems may be quite simple to resolve, but such work fell outside the scope of this project.

Fedora and the Preservation of University Records Project

3.1 Maintain Guide

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/

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:UA069.004.001.00009
(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

TABLE OF CONTENTS

OVERVIEW

Records with enduring value that have been created or ingested into a recordkeeping or preservation system[1] must be kept, stored, and protected from harm along with their accompanying metadata; in short, they must be maintained. This process is roughly equivalent to the Data Management and Archival Storage functions of the *OAIS* reference model[2] and to the Maintain Electronic Records process from the InterPARES Project's *Preservation Model*.[3] The Maintain Guide does not represent the entire preservation process. It instead represents a core subset of that larger process, excluding many key preservation activities that occur in the Preservation Planning and Administration functions of the *OAIS* model (e.g. file format transformation, monitoring changing technology, and setting policies). This guide is instead intended to provide a high-level view of the activities involved in the maintenance of the digital components of electronic records[4] and their accompanying metadata for the purpose of reproducing authentic copies of such records. The maintenance of electronic records is a necessary part, but not the whole, of electronic records preservation.

The Maintain Guide is based largely on the conceptual underpinnings of the records lifecycle model, presuming that a Producer will create, acquire, use, and manage records in a Recordkeeping System to suit its current business needs, and later the Archive will ingest some of those records into a separate Preservation System that the Archive administers. In this model, the Archive acts as a neutral third party in the recordkeeping process, acting on behalf of broader societal needs rather than on behalf of the Producer. As a neutral third party, the Archive has no stake in the content of the records and no reasons to alter records in its custody, and it should not allow anybody to alter the records either accidentally or on purpose. Many archivists have rejected the lifecycle model in favor of the records continuum concept, where recordkeeping is seen as a continuous process that is not time-based, separated into a series of clearly defined steps, or administered by completely separate juridical entities. Many Producers and Archives operate in a mixed world between these two models. For example, many Archives operate separately from a Producer but are part of same organization as the Producer and do not act as a neutral third party. The Maintain Guide should be useful to most Archives operating in a mixed lifecycle/continuum environment, particularly ones where the systems responsible for recordkeeping and preservation are physically and/or intellectually separated.

Electronic records are stored as digital components, which may be separate digital files or contained in a single digital file. The preservation of electronic records includes all of the activities and processes involved in the physical and intellectual protection and technical stabilization of digital resources through time in order to reproduce authentic copies of those

[1] See discussion of different recordkeeping environments in Part 1.3 of *Project Overview* <http://dca.tufts.edu/features/nhprc/reports/1_1final.pdf>.
[2] ISO 14721:2003, Space data and information transfer systems—Open Archival Information System—Reference Model.
[3] "A Model of the Preservation Function," Appendix 5 of *The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project* (San Miniato, Italy: Archilab, 2005).
[4] "Preservation Task Force Report," from *The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project* (San Miniato, Italy: Archilab, 2005) <http://www.interpares.org/book/interpares_book_f_part3.pdf>.

records. Each time an electronic record is delivered to a human user, the records components must be reassembled and presented in their original documentary form. The documentary form is a set of rules that structure a document's extrinsic and intrinsic elements in order to communicate its content, its administrative and documentary context, and its authority. An Archive will often not preserve an electronic record in the form it ingests the record, but rather migrate the content information through a series of format changes until it reproduces it for a user, all the while preserving the documentary form of the original. If an Archive attempts to preserve a contract with a signature that needs to look like a signature as an essential element of its documentary form, it will fail its preservation mission if it does not preserve the signature in a way that it looks like a signature. The documentary form of the rest of the contract might just be readable text in any form, so the Archive knows that it is allowed to change the contract's form from TIF to XML or some other file format. Reassembly is necessary because the electronic record is not stored in the same form in which it is presented to people. To maintain electronic records, the records' digital components, as well as information about them, must be stored, managed and maintained. This information includes a description of what digital components each record contains, how those components relate to each other and to the record itself or other records, and how the records components should be reassembled into authentic copies of the original records. In order to output authentic electronic records, an activity undertaken during the Access function, it is also necessary to maintain evidence that the currently held records components have not undergone unwanted changes and document any planned changes.[5]

Maintaining electronic records may be understood as a series of data protection actions necessary to maintain a minimum foundation of continuity. This minimum foundation will, in turn, enable long-term preservation. Data protection actions in themselves do not constitute long-term preservation, but they are necessary if electronic records are to survive long enough to allow long-term preservation actions to be undertaken.[6] The data protection actions are not the responsibility of the Archive alone. In order for the Producer to maintain reliable, accurate, and authentic electronic records to support its ongoing business operations, the Producer must undertake many of these same actions.[7]

This guide is aimed at a specific audience: managers of Archives, either professional archivists or other person(s) responsible for the long-term preservation of university electronic records. While this guide may be helpful to other communities, it should not be understood to replace standards or guidance covering the information processing, security, storage, or networking fields. The Maintain Guide demonstrates the difficulty of executing this task. No archivist or electronic records preservation officer should attempt to maintain university electronic records in isolation. There are a number of reasons to collaborate with others, not the least of which is the expense. It will be very expensive to set up and operate the necessary infrastructure. These expenses will likely dwarf the normal operating budgets of most university archives and will

[5] "How to Preserve Electronic Records," Appendix 6 of *The Long-term Preservation of Authentic Electronic Records*, p. 20-21.
[6] "Protecting Data," Chapter 16 of *Guidelines for the Preservation of Digital Heritage* (United Nations Educational, Scientific and Cultural Organization, 2003).
[7] For the purpose of this document we will concentrate on these same actions only for maintaining electronic records by the Archive. For more information on the requirements of the Producer, see "Recordkeeping System Requirements," Section IV of *Requirements for Trustworthy Recordkeeping Systems and the Preservation of Electronic Records in a University Setting* <http://dca.tufts.edu/features/nhprc/reports/1_5final.pdf>.

necessitate finding ways to utilize existing resources or sharing expenses (both development and operating expenses) across departments or even across institutions. There are other significant benefits to cooperating with others to maintain electronic records. An Archive can access a wider range of expertise than is likely to reside in any one university archives or with any one university archivist. This is particularly true of the technical expertise that exists in other units of the university. Working with existing services and drawing on existing expertise can only benefit the process. If no such expertise or services exist within the institution, those charged with preservation may need to work with outside vendors or perhaps seek collaborators at other institutions, just as libraries work together to share online public access catalogue development.[8]

The Maintain Guide has excluded any management—such as preservation planning—or subject-related decision-making activities from its purview in order to focus on the technical and procedural activities of maintaining data integrity. The Guide describes activities that an automated system or systems administrator or technician can execute without needing the subject or management knowledge of the records to undertake this work. Any maintain work that rises to the level of administration or preservation planning falls outside of the scope of the Guide.

The Maintain Guide assumes that it is in the best interest of most university archives to engage the services of its institutional information services (IS) department, which is dedicated to providing computing and/or storage services to the university's departments, or with an outside vendor to carry out many of these maintenance activities. Because the expense and technical complexity associated with these activities is often beyond the capacity of many university archives, such partnerships with internal IS departments or vendors should be an attractive solution for many Archives. It is up to each Archive to decide how best to delegate its responsibilities for maintenance of electronic records. The Maintain Guide is designed to help archivists understand these activities and to enable archivists to define the set of services needed from an internal IS unit or external vendor, potentially serving as the basis for negotiating service level agreements.

The event types described below are much the same as those managed by any typical information systems (IS) department. However, the nature of the response to these events and the activities specified do not necessarily follow the standard operating procedures of the typical IS department. It is expected that the continuing value assigned to the records during appraisal and the requirements inherent in reproducing authentic copies of electronic records may force those maintaining electronic records to undertake different and perhaps more expensive activities than most IS departments normally execute. It may be necessary to be emphatic about this point when negotiating services to achieve satisfactory results.

Background on Archival Storage

The Maintain Guide describes electronic records as being stored in two separate storage areas (these may be either physically or merely conceptually separate): the Records Components Store and the Administrative Metadata Store. The Records Components Store will be the archival storage location for the content bitstreams of the records components. It is presumed to be a

[8] See Chapter 11, "Working Together," in *Guidelines for the Preservation of Digital Heritage* (Paris: UNESCO, 2003).

large, reliable, stable storage area with moderate rates of read access and low rates of write access. It is not necessary that this storage be online (constant, very rapid access to data). It may be offline (infrequent access for backup purposes or long-term storage) and may or may not be the subject of regular backups. The Administrative Metadata Store will provide storage for the Archival Information Package (AIP) wrappers, Preservation Description Information (PDI), checksums and other associated metadata needed to keep track of the records and their associated components effectively. This store is presumed to be a smaller storage area with high rates of both read and write access. However, it is not necessary or desirable for the Administrative Metadata Store to be implemented as a relational database; querying capabilities are provided by the Data Management function (described below). It is imperative to have regular backups of this store in a safe location (in addition to high reliability base storage) in order to mitigate the possibility of archive-wide data loss due to malicious or unintentional administrative data alteration. It is likely that the Administrative Metadata Store represents a higher cost per unit of storage than the Records Components Store. If only one actual storage area is used, it is presumed to have the attributes of the Administrative Metadata Store, even though both the records components and administrative metadata will be stored there. For both stores, any backup systems used are considered a part of the storage system as a whole, and not a separate storage system.

The exact storage system strategy used depends on the nature of the Archive, its needs for trustworthiness, the value of the records it maintains, and the resources available. The primary properties of storage areas that impact the trustworthiness of the Archive are the *mean time to data loss* (MTTDL) and the *data loss rate*, which are measures of the reliability of a storage system.[9] Data storage experts can base estimates of these properties on the reliability of the base media used, the way in which that media is arranged into a storage system, and the types and schedules of backups. The MTTDL is a measure of the expected amount of time that will pass before data loss occurs. Longer MTTDL implies a more reliable storage system. The data loss rate refers to the expected amount of data lost per unit time. Lower data loss rates indicate a more reliable system.

Each Archive must determine the appropriate kind of storage by evaluating its reliability, cost, and performance characteristics. Furthermore, the Archive Administration should continue to monitor the storage options available to it and initiate changes in storage strategies when needed. These policy decisions fall strictly outside of the Maintain Electronic Records process, but they greatly affect its success.

Background on Data Management

The *OAIS* Data Management function "provides the services for populating, maintaining, and accessing both Descriptive Information which identifies and documents Archive holdings and administrative data used to manage the Archive.[10] Its primary value is the ability to promote discovery of records components with particular attributes and to generate reports about records.

[9]Peter M. Chen, David E. Lowell, "Reliability Hierarchies," *HotOS*, p. 168, *The Seventh Workshop on Hot Topics in Operating Systems*, 1999 <http://portal.acm.org/citation.cfm?id=822076.822439>.
[10] ISO 14721:2003, p. 4-2.

Thus, it is important that Data Management accurately describe the records components in Archival Storage. Data Management is presumed to have fewer of the storage reliability concerns that dominate Archival Storage, but a more constant need to rapidly access the data. The Data Management function will most likely be implemented by some database specifically designed to support queries, such as a relational or XML database. This database will need to be updated every time there is a change or addition to a record's metadata. With the exception of query statistics, all information in the Data Management function should be derivable from the archival data stored in Archival Storage. As query results from the data management database may be visible to Consumers, care is needed to ensure that any sensitive information about records is properly controlled.

Form of the Guide

The Maintain Guide is a prescriptive guide for an Archive to conduct a Trustworthy Maintain Electronic Records process. However, this process is a continual activity that lacks easily defined beginning and ending points. It does not lend itself well to a step-by-step process definition. Instead, most of the actions in this process occur in reaction to a specific event. These events can occur either in response to the passage of a specified period of time (a Scheduled Event Type) or to the action of another Archive function (an Irregular Event Type). Event Types may have preconditions that must be true for the event to occur. The Activities of an Event Type may cause other Event Types to occur. For example, a scheduled "test checksum" event may cause an irregular "checksum failed" event to occur.

The Maintain Guide prescribes a list of Activities that an Archive must follow in response to an event. Because these Activities are sequential, each set of Activities can be read as a general step-by-step guide. However, because the Guide does not fully prescribe all the details and decisions for the Activities, implementation of the Guide will vary from Archive to Archive. Capitalized words throughout the Guide identify keywords that are defined in the project glossary.

Recordkeeping Infrastructure or Natural or Juridical People are the actors undertaking every activity listed in all of the event types of the Maintain Guide. Such actors can consist of people or hardware or software that belong to the Archive, belong to a systems group at the Archive's institution, or belong to a third-party vendor. An Archive's People and Infrastructure may come from a combination of these in-house and outsourced locations. This Guide does not prescribe how an Archive should organize or administer its People and Infrastructure. The Archive will also have staff that undertakes Preservation Planning and Administration actions, but these activities fall outside of the scope of the Guide.

MAINTAIN GUIDE: SCHEDULED EVENT TYPES

Overview

These types of events may occur according to a predetermined schedule. The exact schedule is determined by the Archive Administration based on the needs and particular situation of the Archive. These services may need to be negotiated with a vendor or internal information services department. Suggested activities and guidelines are provided for each Event. The types of events are listed roughly in the order of their frequency.

Incremental Backup of Administrative Metadata

Description A data backup that is performed frequently. This may be a full backup (including all data objects, regardless of whether they have been modified since the last backup) or a cumulative incremental backup (including all data objects modified since the last full backup was copied). It may be stored on- or off-site. The backup schedule can affect the data loss rate for the Administrative Metadata Store and should be carefully considered. Data loss rate is the expected amount of data lost per unit of time. A greater frequency of backups will reduce the data loss rate because that will shorten the time between backups and therefore lessen the amount of data that is not backed-up at any given moment.

Suggested Schedule This can vary depending on the volume of data and activity at the Archive, but we suggest incremental backups occur at least weekly. Daily is ideal.

Preconditions

- None

Activities

1. Perform backup of administrative metadata
2. Store backup data in a secure location

Full Backup of Administrative Metadata

Description A less frequent data backup that is a full image of the Administrative Metadata Store. The exact state of the Administrative Metadata Store at the time of the backup can be completely restored using this backup only. This backup should be stored in a highly secure off-site location. Cycle the backups such that several full backup images going back in time are retained. This aids in recovery from accidental or malicious damage if it is discovered long after the damage occurred. A minimum of a full twelve months worth of backup images is recommended.

Suggested Schedule This can vary, but a full backup is suggested to be performed at least every three months.

Preconditions

- None

Activities

1. Perform full backup image of Administrative Metadata
2. Move full backup image to secure off-site location
3. Retain at least one year’s worth of full backup images at the secure off-site location

Incremental Backup of Records Component Store
Description A data backup that is taken frequently. This may be a full backup (including all data objects, regardless of whether they have been modified since the last backup) or a cumulative incremental backup (including all data objects modified since the last full backup was copied). It may be stored on- or off-site. The backup schedule can affect the data loss rate for the Records Component and should be carefully considered. Data loss rate is the expected amount of data lost per unit of time. A greater frequency of backups will reduce the data loss rate because that will shorten the time between backups and therefore lessen the amount of data that is not backed-up at any given moment.
Suggested Schedule This can vary depending on the volume of data and activity at the Archive, but we suggest incremental backups occur at least weekly. Daily is ideal.
Preconditions
- None

Activities
1. Perform backup of records components
2. Store backup data in a secure location

Full Backup of Records Component Store
Description A complete backup image of the Records Component Store. The exact state of the Records Component Store at the time of the backup can be completely restored using this backup only. This backup always maintains at least the previous backup image, and does not overwrite the same tapes each time.
Suggested Schedule Can vary, but a full backup image of the Records Component Store should be performed at least every 12 months.
Preconditions
- None

Activities
1. Perform a full backup image of the Records Component Store
2. Move full backup image to secure off-site location

Verify AIP Consistency
Description Check each AIP in the repository against some internal consistency criteria. This can be similar to a Cyclic Redundancy Check (CRC) or can be separately stored checksums, or some other appropriate mechanism.
Suggested Schedule Each record should have its AIP consistency checked as often as resources practically allow, but at least once between full backup images of the Administrative metadata store.
Preconditions
- None

Activities
1. Verify the internal consistency of each AIP
2. If any AIP fails the consistency check, go to **Respond to AIP Consistency Failure**

Verify Records Components

Description The AIPs for each record contain fixity information about records components (perhaps in the form of cryptographic checksums, message authentication codes, integrity check-values, modification detection codes, or message integrity codes). These fixity information values should be periodically calculated from the records components and verified against the existing fixity information values. In addition, if digital signatures are part of the fixity information, they should also be verified.

Suggested Schedule Each record should have its records components checksums verified as often as resources practically allow, but at least once between each full Records Components Store backup image.

Preconditions

- None

Activities

1. Compute checksums on records components and compare to checksums stored in PDI
2. If any checksums are not correct, go to **Respond to Checksum Failure**
3. Document "Verify Checksum" event in PDI

Check Access and Retention Status

Description Access and retention may be governed by a time interval. Records should be monitored to discover when such a time interval has elapsed so that Archive Administration and Preservation Planning, if necessary, can take appropriate action.

Suggested Schedule Each record should have its retention and access status checked often enough to ensure that the appropriate level of granularity is achieved. This may be daily, monthly, etc, depending on local policy.

Preconditions

- None

Activities

1. Identify all records governed by time-based expiration
2. Report all such records to Administration
3. Document "Retention or Access Period Expired" event in PDI

Report on Media Life

Description A report should be periodically generated that lists the types and service life of media. This report will aid Administration in deciding when to refresh media.

Suggested Schedule Media life reports should be generated frequently enough that Administration can make appropriate decisions about media refreshment. We anticipate that this will be once every one to six months.

Preconditions

- None

Activities

1. Generate a report listing the types and age of all media in the Records Component Store
2. Submit report to Administration

Hardware Test and Maintenance Window

Description The Preservation System Hardware Environment will require periodic maintenance and should be tested to ensure that hardware components still operate within specifications. Maintenance activities may involve security patches, filesystem defragmentation, and other low-risk activities. Tests should include stress tests to ensure the hardware and system software still operates within its intended parameters.

Suggested Schedule This can vary depending on the need of the Archives, but it should be done at least every six months.

Preconditions

- None

Activities

1. Activate Hot Spare, if available
2. Take down the server in question
3. Perform test suite and regular maintenance operations
4. Report test results to Administration
5. If no immediate problems are discovered, restore server to active service (or Hot Spare status)
6. Report test failure to Administration and take corrective action prescribed by Administration
7. Document test results and maintenance activities in repository history metadata

Security Audit

Description A periodic audit should be conducted of security practices related to all aspects of the Preservation System. This audit should be conducted by independent professional auditors. It should include a review of security procedures and protocols, system software security practices, and potential social engineering problems.

Suggested Schedule A security audit should be performed every 36 months.

Preconditions

- None

Activities

1. Hire an independent auditor to review security procedures and protocols, adherence to procedures, physical security, and other security practices
2. Report security audit results to Administration
3. Document security audit results in repository history metadata

MAINTAIN GUIDE: IRREGULAR EVENT TYPES

Overview

These events occur according to some external stimulus or as the result of the actions of a scheduled event. They are irregular because their exact timing cannot be anticipated. Some Event Types preconditions must be true for the Event Type to occur.

Digital Object Accession

Description This Event occurs when an object has successfully completed the ingest process and needs to be maintained in the Archive.

Preconditions

- Object passed through Ingest

Activities

1. Generate Storage Identifier(s)
2. Place records component(s) in Records Component Storage
3. Document Storage Identifier(s) in AIP
4. Add “Object Accession” event to PDI history
5. Place AIP and PDI in Administrative Metadata Storage
6. Update Data Management Database
7. Schedule Events based on accession date

Retrieval Request

Description This Event occurs whenever an Archive needs to obtain a copy of a record component (either for a consumer or for some internal function of the Archive). The initiator of this event can be a Customer (requesting a Dissemination Information Package (DIP)), a member of the Archive (doing some sort of review), or an internal maintenance operation (such as verifying checksums).

Preconditions

- Initiator has read permission for the record component

Activities

1. Look up the Storage Identifier for the record component
2. Retrieve record component from Records Component Storage
3. Provide record component to requesting archive function
4. Update retrieval statistics

Query Request

Description This Event occurs when an Archive needs to run a query against record metadata. This might be for a Customer search, a regular report for Administration, or a maintenance operation.

Preconditions

- None

Activities
1. Perform requested query
2. Filter results set as necessary according to initiator's permissions
3. Provide results set to requesting archive function
4. Update query statistics

Metadata Update Request
Description This Event occurs whenever record metadata that is *not* part of the administrative metadata is updated. This can occur when a member of the Archive creates or modifies descriptive metadata, when technical metadata is created or derived from the records, or when supporting records (such as Representation Information (RI), Record Type Records or Producer Records) are updated.
Preconditions
- Initiator has write permission for the record

Activities
1. Add a new metadata bitstream or version existing bitstream as appropriate
2. Update AIP with any new Storage Identifiers and fixity information
3. Update PDI with "Metadata Update Event"
4. Update Data Management Database

Format Transform Request
Description This Event occurs when Administration has decided to transform a file from one format into another. For example, this could be a metadata crosswalk or a content transformation.
Preconditions
- Transformation process tested and approved by Preservation Planning
- Transformation approved by Administration

Activities
1. Identify all records components or metadata bitstreams that are represented in the affected format
2. For each bitstream, retrieve the bitstream and perform the transformation
3. Validate the file formats of the transformation outputs; report any validation failures to Administration and take corrective action prescribed by Administration
4. Generate a Storage Identifier for the newly created bitstream
5. Store the bitstream in the Records Component Store
6. Update AIP to include the new bitstream
7. Update PDI with a "Format Transformation" event
8. Update Data Management Database

Remove Record Component Request
Description This Event occurs when the Archive Management decides to remove a metadata or content bitstream from a record. This can occur when a format becomes obsolete, when the

preservation goals for a record have changed, or in response to other Preservation or disposition actions.

Preconditions

- Preservation Planning has approved the removal
- Administration has approved the removal

Activities

1. Confidentially destroy the records component
2. Update AIP to remove the record component
3. Update PDI to add "Record Component Removed" event
4. Update Data Management Database

Preservation Application Hardware Environment Replacement

Description This Event occurs when the Archive Management has determined that the Preservation Application needs to be migrated to new hardware. This decision may be prompted by poor hardware test results, increased demand, or it may simply be a preventive measure to replace aging hardware.

Preconditions

- Replacement authorized by Archive Administration

Activities

1. Continue to maintain Hardware Environment
2. Acquire the new Hardware Environment
3. Set up the new Hardware Environment with a new installation of the Preservation Application and all needed system and utility software. Perform system-level configuration, including networking setup.
4. Set up new Administrative Metadata Store and Data Management Database (and Records Components Store, if necessary)
5. Place the Preservation Application in stasis on the old Hardware Environment
6. Update all PDI to include a "Begin Hardware Environment Replacement" event
7. Copy all AIPs and PDI from the old Administrative Metadata Store to the new Administrative Metadata Store. Also copy Records Components Store if necessary
8. Build the new Data Management Database from the new Administrative Metadata Store
9. Test a representative sample of AIPs on the new system to ensure full functionality
10. After passing tests, update PDI on new server with "End Hardware Environment Replacement" event
11. Update Repository History record with "Hardware Environment Replacement" event
12. Perform full backup image of the new Administrative Metadata Store
13. Remove Preservation Application from stasis on new Hardware Environment
14. Re-designate new Hardware Environment as active
15. Confidentially destroy data on old Hardware Environment

New AIP Format and/or New Preservation Application

Description This Event occurs whenever the Archive decides to change the Preservation Application or an AIP format. It is anticipated that these events will usually coincide.

Preconditions

- AIP transformation tested and approved by Preservation Planning
- New Preservation Application tested and approved by Preservation Planning
- Changes authorized by Archive Administration

Activities

1. Continue to maintain the old Hardware Environment
2. Acquire a new Hardware Environment
3. Set up the new Hardware Environment with a new installation of the new Preservation Application and all needed system and utility software. Perform system-level configuration, including networking setup.
4. Set up new Administrative Metadata Store and Data Management database (and Records Component Store, if necessary)
5. Place Preservation Application in stasis on old Hardware Environment
6. Update all PDI to include "Begin Preservation Application Change" event
7. Transform all AIPs and PDI and submit them to the new repository. Also copy Records Component Store if necessary
8. Build the new Data Management database from the new Administrative Metadata Store
9. Test a representative sample of AIPs on the new system to ensure full functionality; report any test failures to Administration and take corrective action prescribed by Administration
10. After passing tests, update PDI on new Hardware Environment with "End Preservation Application Change" event
11. Update Repository History with "Change Preservation Application" event
12. Perform full backup image of the new Administrative Metadata store
13. Remove Preservation Application from stasis on new Hardware Environment
14. Re-designate new Hardware Environment as active
15. Confidentially destroy data on old Hardware Environment

New Records Component Store

Description This Event occurs when the Archive has decided to move to a new Storage Hardware Environment for its Records Components Store.

Preconditions

- New Storage Hardware Environment researched and approved by Preservation Planning
- Change approved by Archive Management

Activities

1. Acquire the new Storage Hardware Environment
2. Set up the new Storage Hardware Environment
3. Test new Storage Hardware Environment
4. Copy the Records Component Store when necessary; See **Preservation Application Hardware Environment Replacement.**

Add Additional Representation Information

Description This Event occurs when a new preservation file format is added that requires new Representation Information (RI) or when a shift in the Knowledge Base of the designated community requires new RI for file formats that was removed from the community's Knowledge Base.

Preconditions
- RI approved by Preservation Planning

Activities
1. Accession the digital objects representing the new RI (see **Digital Object Accession**)
2. Update any objects which require links to this RI (see **Metadata Update Request**). This list of objects is provided by Preservation Planning.

Change Standard Computing Platform

Description This Event occurs when Preservation Planning determines that the Standard Computing Platform (SCP) needs to be changed. Preservation Planning provides the new SCP definition.

Precondition
- Change approved by Preservation Planning

Activities
1. Update SCP record (see **Metadata Update Request**)
2. Find all records that are no longer grounded in the Knowledge Base
3. Report all such records to Preservation Planning

Refresh Records Component Media

Description This Event occurs when one of the primary media elements of the Records Components Store is refreshed. This can occur preventively or because errors have been detected on the media.

Preconditions
- Archive Administration approves the refresh

Activities
1. Prepare new media for use
2. Test new media for manufacturing defects
3. Copy records components onto new media
4. Perform a bit-level comparison between old and new media
5. If bit-level test succeeds, redesignate new media as active storage for affected records components
6. Update PDI for all affected records with “Media Refresh” event
7. Document when media becomes active
8. Confidentially destroy data on old media
9. Discard or recycle old media

Respond to Checksum Failure

Description This Event occurs when checksum verification fails for a record component. Such an Event usually occurs following an automated **Verify Records Components** Event.

Preconditions
- Checksum failure has occurred

Activities
1. Mark the record as containing compromised data

2. Alert Administration of this event
3. Search alternate storage locations to find the record component backups
4. If an intact, independently stored record component is discovered, then go to **Verify Records Components**, verify component checksum
5. If component checksum does not match, proceed to next appropriate independently stored record component
6. If no intact datastream is found, go to **Respond to Data Loss: Record Component Store**
7. If component checksum matches, then replace current record component with alternately stored component in the active Record Component Store, and document that the Archive has repaired the record in its PDI
8. Update record PDI "Record Component Repaired" event or "Record Component Corrupted" event

Respond to Media Failure: Record Component Store
Description This Event occurs when one of the primary media elements of the Records Components Store has completely failed.
Preconditions
- Media failure has occurred in the Record Component Store

Activities
1. Mark all records with affected components as having corrupted data
2. Look for alternate storage locations for the data stream (such as backups)
3. Prepare and test new media
4. Restore onto new media all found datastreams that positively match their existing integrity information
5. Document that the Archive has repaired the records in their PDI
6. If any record components could not be repaired, report to Preservation Planning and go to **Respond to Data Loss: Record Component Store**

Respond to Data Loss: Record Component Store
Description This Event occurs following a checksum failure or media failure event that affects the Records Component Store in which the record component is unable to be restored. The loss of a record component is a major detriment because it reduces the trustworthiness of the Archive. All reasonable efforts should be made to avoid such a loss.
Preconditions
- Unrecoverable data loss has occurred in the Record Component Store

Activities
1. Notify Archive Administration and Preservation Planning of the data loss
2. Document data loss in repository history metadata
3. Document data loss in the PDI of the affected records

Respond to AIP Consistency Failure
Description This Event occurs when an AIP fails an internal consistency check. Such events usually occur following an automated consistency check event.

Preconditions

- AIP consistency failure has occurred

Activities

1. Place Preservation Application in stasis
2. Check all AIPs for consistency. Assume all media upon which consistency failures have occurred has failed; go to **Respond to Media Failure: Administrative Metadata Store**

Respond to Media Failure: Administrative Metadata Store

Description This Event occurs when one of the primary media elements of the Administrative Metadata Store has completely failed.

Preconditions

- Administrative Metadata Store media has failed

Activities

1. Place Preservation Application in stasis
2. Acquire a new Preservation Application Hardware Environment
3. Set up a new Administrative Metadata Store
4. Set up the new Preservation Application Hardware Environment with a new instance of the Preservation Application and all needed system and utility software
5. Copy all savable AIPs and PDI to the new Administrative Data Store
6. Reconstruct missing AIPs from backup images
7. Document missing or corrupted AIPs that cannot be restored from backup images
8. Document the media failure in the repository history metadata
9. Report results of the reconstruction to Archive Administration, who will decide what further actions to take. If any missing or corrupted AIPs cannot be restored from backup images, Archive Administration will probably want to undertake **Respond to Data Loss: Administrative Metadata Store**

Respond to Data Loss: Administrative Metadata Store

Description This Event occurs following a media failure if the Archive Administration is unable to reconstruct a usable AIP for one or more records. Administrative Metadata Store data loss is a catastrophic event for an Archive because it fundamentally undermines the trustworthiness of the Archive. All possible efforts should be made to avoid such a loss[11].

Preconditions

- Unrecoverable data loss has occurred in the Administrative Metadata Store

Activities

[11] Losing administrative metadata is generally worse than losing records components. If an Archive loses the records components but has the administrative metadata then it can at least document the record's past existence, provenance, custody, and its (unintended) destruction. The Archive will probably also be able to describe the function of the records when they were in their active environment.

If an Archive loses its administrative metadata but corresponding records components still exist the Archive cannot demonstrate their provenance, custody, or other essential qualities that give the user the reasonable ability to judge the records as authentic because the preservation system now lacks the administrative metadata it needs to be trustworthy. In addition, without the administrative metadata it would be very difficult for the Preservation System managers to locate with certainty the records components they are looking for, never mind preserve them over time.

1. Determine all records components "orphaned" by the lost AIP(s)
2. Generate a new "record fragment" AIP for each of the record components, including as much information as can be reconstructed or gathered from the records components, including at least the format type of the components
3. Document information about the data loss in the PDI for the new AIPs
4. Document the data loss in the repository history metadata

Respond to Unintentional Data Damage
Description This Event occurs when a mistake is made while handling electronic records which that results in an unintended change or deletion of records components or administrative metadata.
Preconditions
- An instance of unintentional data damage has been discovered

Activities
1. Place Preservation Application in stasis
2. Identify the scope and nature of the damage
3. Report to Archive Administration concerning the scope and nature of the damage; Administration will decide the appropriate corrective action
4. Take corrective action prescribed by Administration; if data loss occurs, undertake **Respond to Data Loss: Administrative Metadata Store** or **Respond to Data Loss: Records Component Store**
5. After taking the Preservation Application off stasis, record damage and corrective actions in repository history metadata and the PDI of all affected records

Respond to Security Breach
Description This Event occurs whenever the Archive discovers that an unauthorized person has gained access to any of the hardware which runs the repository.
Preconditions
- Unauthorized activity discovered

Activities
1. Take Preservation System offline; do *not* activate Hot Spares
2. Analyze all repository hardware to determine what machines have been compromised and to discover the nature and scope of the attack, and what actions the attacker took while he or she had access to the Hardware Environment
3. Perform internal consistency check of all Administrative Metadata
4. Perform checksum verification of all records components
5. Compare Administrative Metadata to a known-good backup image (taken before the attack occurred), and compile a list of all changes between the current image and the backup
6. Report all findings to Archive Administration which determines if the attacker was:
   a. A Squatter (only using computing resources)
   b. A Vandal (intending to do indiscriminate damage)
   c. An attacker with motive against the records (intending to alter or destroy records in particular)
7. Administration decides appropriate corrective actions

8. Perform prescribed actions; if data loss occurs, undertake **Respond to Data Loss: Administrative Metadata Store** or **Respond to Data Loss: Records Component Store**
9. Document security breach in repository history metadata and PDI of all records

Fedora and the Preservation of University Records Project

3.2 Checklist of Fedora's Ability to Support Maintain Activities

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/

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:UA069.004.001.00010
(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

TABLE OF CONTENTS

OVERVIEW

This Checklist builds directly on the Maintain Guide, describing the abstract services that support the prescribed responses to the set of scheduled and irregular event types presented by the Guide. The Checklist describes services in the most generic and abstract sense; this notion of service is intended to include "manual" services performed by people, "automated" services implemented in software, as well as various combinations of the two. Some services are most easily implemented as manual processes, some as software services, and others can be reasonably implemented in a variety of ways.

Understanding the Checklist depends significantly on reading the Maintain Guide. The Guide presents a set of scheduled and irregular event types that support various Maintain steps. The event types are non-sequential. Each event type is triggered by various circumstances or conditions. The Guide describes the nature of each event type, the conditions for their occurrences, and a set of steps an Archive (or its designee) needs to undertake when an event type occurs. An Archive will need to utilize at least some of the abstract services the Checklist describes in order to undertake the steps set out by the Maintain Guide in a scaleable manner.

The Checklist also analyzes Fedora's ability to support the abstract services identified. This analysis examines if Fedora currently utilizes the abstract services as part of its core architecture, in its service framework, or as an external service that it communicates with. The Checklist discusses current and potential future development work for abstract services that do not currently work with Fedora in some capacity. This future development work often refers to the activities of the Fedora Preservation Services Working Group, a body within the Fedora community charged to develop a "general definition of preservation services for the Fedora service framework [and]… recommend enhancements to the [core] Fedora repository service as well [as] develop specifications for new preservation-support services for the Fedora Service Framework."[1] This Checklist measures its analysis of Fedora against version 2.1 (released February 2006).

Form of the Checklist

After providing an overview and a description of Fedora the Checklist lists nineteen abstract services, providing for each service a description, an analysis of Fedora's support for the service, and a list of event type steps described in the Maintain Guide that the service supports. In Appendix A, the Checklist has a crosswalk of the Maintain Guide event type steps to the abstract services that support the steps.

Fedora

Fedora is open repository application.[2] 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. It is a developer-oriented, rather than user-oriented, application that allows repository managers to build (or use existing) services on top of Fedora, including ones such as search services that interface with users.

[1] "Working Group: Preservation," <http://www.fedora.info/wiki/index.php/Working_Group:_Preservation>
[2] Fedora stands for Flexible Extensible Digital Object Repository Architecture. For more information on Fedora see <http://www.fedora.info>

Fedora is essentially a set of repository services. At its kernel, are the services that formulate the Fedora Core that is surrounded by services in the Fedora Service Framework. Services in the Core and the Framework also work with external services that reside outside a Fedora repository. The Fedora Core[3] is a core repository application that ingests digital objects and its data-streams; validates the object wrapper (usually FOXML) and the presence of declared data-streams; stores (internally or externally), registers, indexes, and manages the objects and data-streams; enables search of the objects and dissemination of the data-streams; and enforces access polices. These functions are exposed through four web service APIs:

1) Repository management interface (API-M) that provides write access to the repository
2) Repository access interface (API-A) that provides read access to the repository
3) Basic repository search index
4) RDF-based search of the Core's resource index.

The Fedora Service Framework allows for a more modular preservation system, because services are implemented on top of the Fedora Core, allowing their configuration or replacement with another service.[4] The Fedora Service Framework, which was implemented with version 2.1 (released February 2006), surrounds the Fedora Core and allows the smooth integration of the Core with new services. These services operate independently of the core repository application. Managers of Fedora instances are free to choose which services to implement. This keeps Fedora agile and better able to respond to emerging needs by extending its services while leveraging existing resources and avoiding the need to do costly and risky core overhauls.[5]

If configured properly, a Fedora repository instance can communicate with external services. For example, an external workflow system can communicate with a specialized ingest service in the Framework to undertake batch ingests. An integrity service could work with an external service such as JHOVE to validate the format of data-streams ingested into a repository. This allows Fedora to extend its services further by leveraging the work of resources completely outside the repository.[6]

[3] In this document, Fedora Core refers to the core services for the Flexible Extensible Digital Object Repository Architecture (Fedora) developed by Cornell University and University of Virginia Library. See <http://www.fedora.info> and <http://www.fedora.info/wiki/index.php/Fedora_Core_Repository_Service>. This does *not* have anything to do with Red Hat releases of their community-oriented Linux distribution.
[4] "Fedora Core Repository Service," <http://www.fedora.info/wiki/index.php/Fedora_Core_Repository_Service>
[5] "Fedora Service Framework," <http://www.fedora.info/wiki/index.php/Fedora_Service_Framework>
[6] Ibid.

ABSTRACT SERVICES

AIP Module

Description

This module helps the Archive manage Archival Information Packages (AIPs) within its Preservation System. AIPs consist of Content Information, Preservation Description Information, and Packaging Information.[7] The module facilitates the semi-automated management of AIPs. For example, when the Archive transforms the format of thousands of records to a new format, the AIP Module facilitates the addition of the new bitstreams to the AIPs.

Fedora Support

Fedora manages AIPs as part of its *Core* services, undertaking the creation, modification, and deletion of digital objects and its data-streams, executing the activities of ingest, validation, management, storage, register, index, search, and disseminate. Fedora exposes these activities through its Fedora Management service, known as API-M Methods. In addition, Preservation System managers might utilize management tools, like FEZ and ELATED. Currently these services provide some of the management functionality needed in an AIP Module.[8] The API-M Methods, FEZ, or ELATED cannot support all the specialized needs of an AIP Module. For example, they are not explicitly designed to address the discovery of all records components "orphaned" by lost AIP(s) (Respond to Data Loss: Administrative Metadata Store) or the ability to generate a new "record fragment" AIP for each orphaned record component (Respond to Data Loss: Administrative Metadata Store).[9]

Supports Event Type Steps

Format Transform Request: Step 6
Remove Record Component Request: Step 2
Preservation Application Hardware Environment Replacement: Step 9
New AIP Format and/or New Preservation Application: Step 9
Respond to Data Loss: Administrative Metadata Store: Steps 1, 2

Alerting Service

Description

A service that alerts a repository manager or service that an event has occurred that may require the execution of a subsequent Maintain activity. The service provides the infrastructure for communicating event occurrences to the proper services or managers. For example if Storage Management Module generates a report on the age of storage media, it uses the Alerting Service to send that report to Administration. Archive Administration and Preservation Planning could configure an Alerting Service to feed

---

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

[8] "FEZ development at the University of Queensland Library," <http://www.library.uq.edu.au/escholarship/fezdev.pdf> and "Elated: a general-purpose web-based client for the Repository System," < http://elated.sourceforge.net/ >. Also see "Tools," <http://www.fedora.info/tools/index.shtml>.

[9] Administrators of a Fedora-based Preservation System may use a variety of techniques to discover "orphaned" records components or generate "records fragments" but modules specifically designed to undertake these tasks do not exit.

them reports on appropriate Maintain activities to help ensure that they are properly monitoring the records in the Preservation System and can respond to events in a timely manner.

Fedora Support

Fedora *Core* does not include an Alerting Service, nor does one currently exist as a *Framework* resource. While tools exist that can execute as least some of the Alerting Service function, no one has utilized them with Fedora to date. The Fedora Preservation Services Working Group is currently engaged in developing an Alerting Service that would be part of the Framework and may require some small adjustments to the Core. The Fedora Development Team has begun work on message-enabling the *Core* and the *Service Framework*, which lay the architectural foundation for the Alerting Service. The Working Group envisions the Alerting Service as part of a larger Event Management Service that documents, encodes, and facilitates management of events that occur within the repository. The Alerting Service may in future serve as the underlying communication architecture between a broad range of services in the Fedora *Framework* and *Core*.[10]

Supports Event Type Steps

Check Access and Retention Status: Step 2
Report on Media Life: Step 2
Hardware Test and Maintenance Window: Steps 4, 6
Digital Object Accession: Step 7
Format Transformation Request: Step 1
New AIP Format and/or New Preservation Application: Step 9
Change Standard Computing Platform: Step 3
Respond to Checksum Failure: Step 2
Respond to Media Failure: Record Component Store: Step 6
Respond to Data Loss: Record Component Store: Step 1
Respond to Media Failure: Administrative Metadata Store: Step 9
Respond to Unintentional Data Damage: Step 3
Respond to Security Breach: Step 6

Data Backup Protocol

Description A set of rules and procedures for administering a data backup process. This usually includes defining backup types (full or incremental), the frequency of backups, the storage of backup data, the cycling and retention of backup data, and the responsibilities for administering the data backup process. The Data Backup Protocol is closely related to the Storage Management Module and may be a subset of that Module.

Fedora Support

Currently, neither Fedora *Core* nor the *Framework* manages data backup processes, nor does an *External* data backup service built specifically to work with Fedora exist. Managing a Data Backup Protocol completely separate from Fedora does not cause problems per se, but it would be helpful to connect data backup administration with the digital objects in the repository. For example, retention and disposition metadata may have a significant impact on the cycling and retention of backup media.

Supports Event Type Steps

Incremental Backup of Administrative Metadata: Steps 1, 2

[10] Fedora Group: Preservation <http://www.fedora.info/wiki/index.php/Working_Group:_Preservation>.

Full Backup of Administrative Metadata: Steps 1, 2, 3
Incremental Backup of Records Component Store: Steps 1, 2
Full Backup of Records Component Store: Steps 1, 2
Preservation Application Hardware Environment Replacement: Step 12
New AIP Format and/or New Preservation Application: Step 12
Respond to Checksum Failure: Steps 3, 4, 5, 6, 7
Respond to Media Failure: Record Component Store: Step 2

Data Management Database

Description

This service corresponds closely to the OAIS data management functional entity. It is primarily responsible for enabling the discovery of objects in the repository. As such, it usually takes the form of some manner of a database that supports a query language. This service differs from the search service in that the search service is concerned with the execution of queries and presentation of results, whereas the Data Management Database is concerned with maintaining the data stores which supports such queries.

Fedora Support

Fedora 2.1 comes with two mechanisms for discovering objects in the repository. The first mechanism is called "Basic Search." It is a simple text-based search of the basic properties of the FOXML that wrap Fedora objects and the Dublin Core metadata. The basic search is very simplistic.[11] The second native mechanism for discovering Fedora objects is the RDF triple store. Every Fedora object can have a distinguished metadata stream which contains RDF statements. These RDF statements are maintained in a triple-store which supports arbitrary iTQL queries. It is additionally possible to maintain external search indexes. However, currently maintaining such an external index requires either 1) using an external application which manages both ingest and the index or 2) keeping the index in sync with the repository manually. However, an Alerting Service (see below) would allow one to automatically maintain such an external index.

Support Event Type Steps

Digital Object Accession: Step 6
Metadata Update Request: Step 4
Format Transformation Request: Step 8
Remove Record Component Request: Step 4
Preservation Application Hardware Environment Replacement: Step 8
New AIP Format and/or New Preservation Application: Step 8

Format Transformation Service

Description

This service performs formats transformations of records components, moving a record component from one file format environment to another. This service should be able to undertake batch transformations.

Fedora Support

[11] The relational database underlying a Fedora repository has considerably more information in it than is exposed by the Basic Search interface. Most notably, Basic Search does not show all of the registry and storage tables, plus other control tables. Preservation System administrators can make standard SQL queries directly on these tables. However, Fedora does not expose this information via a public API.

Fedora *Core* does not include a Transformation Service at the individual records component level. While Preservation System administrators can use the Fedora API-M to modify a digital object metadata stream while the digital object remains in place in the repository, it is not a transformation service that transforms data-stream formats—from a MS Word file to a PDF/A file, for example.[12] The Fedora Preservation Services Working Group is considering the possibilities of adopting some kind of Transformation Service to work with Fedora, most likely as an *External* service. Preservation System managers should be able to successfully manage their own Format Transformation Services that only communicate with their Fedora repositories as *External* services.

Supports Event Type Steps

Format Transformation Request: Step 2

Format Validation Service

Description

This service checks and validates the file formats of digital objects. JHOVE is a prominent example of a Format Validation Service. Ideally, a Preservation System should have a Format Validation Service that can validate all of its Preservation Formats and formats it ingests. The Service should also be able to batch-validate. Thus, a Validation Service should have the ability to add new formats to its validating repertoire.

Fedora Support

While Fedora *Core* does validate the digital object wrapper (usually FOXML files) and can confirm the existence of all declared data-streams, it does not include a Validation Service for individual records components. For example, while it can confirm the existence of a data-stream that belongs to a digital object in a Fedora repository, it cannot confirm the format of that data-stream—determining, for example, if a file really is a PDF-A file as it claims. The Fedora Preservation Services Working Group is examining the possibility of developing a Validation Service within a broader preservation integrity service that would call on External validations services such as JHOVE or GDFR. While system managers can successfully run this service completely separate from Fedora, it would help managers monitor the long-term integrity of the digital objects in their repositories.

Supports Event Type Steps

Format Transformation Request: Step 3

Integrity Checking Protocol

Description

This protocol is a set of rules and procedures that defines the methods, timing, and responsibilities for checking the integrity of records components in a Preservation System. In defining the methods, the Protocol incorporates an integrity checking service(s) that are composed of an integrity checking tool(s), such as checksums.

Fedora Support

At the moment, the Fedora *Core* does not come packaged with an Integrity Checking Protocol or tools for checking the integrity of data-streams or digital objects. However, checksum for datastreams is already specified in the Fedora data model (FOXML). Currently, the Fedora Development Team is working on enabling support for checksums

[12] An administrator of a Fedora-based Preservation System could currently build a disseminator to transform data-streams in MS Word to a PDF/A formatted Dissemination Information Package.

in *Core*. This would allow a Preservation System administrator to configure the repository core of his or her Preservation System to automatically calculate checksums for datastreams as part of completing any API-M transaction. The Development Team is also creating a new Fedora API-M method that will allow clients to make requests to (1) calculate and store the checksum of a datastream on demand, and (2) run a comparison of a checksum to the current state of datastream content to see if there is a change. The Development Team expects to complete this work by the fourth quarter of 2006. A manager of a Fedora-based Preservation System can successfully manage an Integrity Checking Protocol that is external to the Fedora repository. Fedora will never define the protocol of the methods, timing, and responsibilities of the integrity checking—that is something that Preservation System managers will have to develop on their own or adopt from external integrity checking standards or best practices. The Fedora Preservation Services Working Group is currently investigating integrity checking tools as a *Framework* service that will allow managers to more easily report on an integrity checking event in order to document the occurrence of a check—and the results of that check—in digital objects' metadata or repository log files. This work is closely related to the Working Group's development of its Alerting Service.

Supports Event Type Steps

Verify AIP Consistency: Step 1
Verify Records Components: Step 1
Respond to AIP Consistency Failure: Step 2
Respond to Unintentional Data Damage: Step 2
Respond to Security Breach: Step 3

Knowledge Base Module

Description

This module facilitates an Archive's tracking and defining a Designated Community's Knowledge Base. The Knowledge Base of a Designated Community is the Community's language ability, subject knowledge, and capabilities of its Standard Computing Platform that it needs to understand records in a Preservation System. Archives attach Representation Information to records to ensure that members of its Designated Community can understand those records with their Knowledge Base. As a Designated Community's Knowledge Base changes over time, an Archive must adjust its Representation Information accordingly.[13]

Fedora Support

Neither Fedora *Core* nor the Service *Framework* includes a Knowledge Base Module and there is no external module that Fedora connects to either. The concept of Knowledge Base has not matured greatly in the digital preservation community so no such module yet exists. Fedora could use its de facto semantic web capabilities, as manifested in the Resource Index, to establish a relationship between objects in a repository and the technical capabilities of the Designated Community to view the objects. Fedora's innate RDF support provides significant flexibility for expressing a wide variety of objects and their relationships, such as object properties and their relationships to external entities like the Knowledge Base of a Designated Community. The Fedora Development Team has started work on providing a more robust and simpler triplestore capabilities in the

[13] According to ISO 14721:2003, a Knowledge Base is, "a set of information, incorporated by a person or system, that allows that person or system to understand received information.

Core and Service *Framework*, which would give Preservation System administrators a more robust and flexible environment for characterizing, tracking, and modifying the Knowledgebase Base of a Designated Community. However, this does not provide direct guidance on how to define and characterize the Knowledge Base of a Designated Community over time.

Supports Event Type Steps

Change Standard Computing Platform: Step 2

PDI Module

Description

This module helps the Archive manage Preservation Description Information (PDI) which is a component of every record (AIPs) within its Preservation System. PDI is the provenance, reference, fixity, and context information that is needed to preserve the records in a Preservation System.[14] The module facilitates the semi-automated management of PDI. For example, when the Archive ingests thousands of records (SIPs) and turns them into thousands of AIPs, this module helps to generate the necessary PDI to be part of those AIPs. The PDI Module is closely related to the AIP Module and in some cases could be a component of the AIP Module.

Fedora Support

Just as with AIPs, Fedora essentially manages PDI as part of its *Core* services, undertaking the creation, modification, and deletion of digital object metadata. Much of Fedora's ability to support the proper administration of PDI will depend on a Preservation System manager's configuration of metadata. This metadata will need to properly capture and manage the appropriate provenance, reference, fixity, and contextual information.

Supports Event Type Steps

Verify Records Components: Steps 1, 3
Check Access and Retention Status: Step 3
Hardware Test and Maintenance Window: Steps 6, 7
Security Audit: Step 3
Digital Object Accession: Step 4
Metadata Update Request: Step 2, 3
Format Transformation Request: Step 7
Remove Record Component Request: Step 3
Preservation Application Hardware Environment Replacement: Steps 6, 10, 11
New AIP Format and/or New Preservation Application: Steps 6, 10, 11
Refresh Records Component Media: Step 6
Respond to Checksum Failure: Steps 1, 8
Respond to Media Failure: Record Component Store: Steps 1, 5
Respond to Data Loss: Record Component Store: Step 3
Respond to Media Failure: Administrative Metadata Store: Step 1
Respond to Data Loss: Administrative Metadata Store: Steps 3, 4
Respond to Unintentional Data Damage: Step 5
Respond to Security Breach: Step 4, 9

[14]ISO 14721:2003, Space data and information transfer systems – Open archival information system – Reference model.

Persistent Identifier Manager

Description

This resource assigns persistent identifiers to records ingested into the Preservation System and manages these identities over time, ensuring they persist and do not conflict with other digital objects.

Fedora Support

Through the API-M Methods the Fedora services allow managers to assign persistent identifiers to objects during ingest into the repository. Fedora also allows for the management of those identifiers over time. Assigning and managing persistent identifiers is a *Core* service. However, it is important to note that Fedora does not manage persistent identifiers across multiple repositories. For example, Fedora does not resolve potential conflicts resulting from duplicate unique identifiers found in a federation of Fedora repositories.

Supports Event Type Steps

Digital Object Accession: Steps 1, 3
Format Transformation Request: Steps 1, 4

Repository History

Description

This is a service that generates a log of events in the life of the repository core of a Preservation System. These events may include significant changes in repository hardware or software, changes in administration, or data loss events. They would not include regular activities such as adding new digital objects to the repository core. Ideally, the log would exist as a special digital object associated with the repository and as data-streams that are part of each digital object in the Preservation System, mostly likely service as a part of the provenance information of an object's PDI.

Fedora Support

Fedora *Core* does not include a Repository History, nor does one currently exist as a *Framework* resource. While the manager of a Fedora-based Preservation System can manually create logs of significant repository events, a service for automatically capturing and logging this information does not exit. The Fedora Preservation Services Working Group is examining the feasibility of developing such a service.

Supports Event Type Steps

Hardware Test and Maintenance Window: Step 7
Security Audit: Step 3
Respond to Data Loss: Record Component Store: Step 2
Respond to Media Failure: Administrative Metadata Store: Step 8
Respond to Security Breach: Step 9

Repository Stasis

Description

This is a service that allows a Preservation System administrator to prevent all changes to a System's repository core—both to the repository itself and to the digital objects it manages. In short, stasis freezes the repository. This would also manage who has permission to put a repository core in stasis and take it out of stasis.

Fedora Support

Fedora does not have a repository stasis module as part of its *Core* functionality *per se*.

However, the Fedora XACML policy module can act as a *defacto* stasis module, allowing Fedora to prevent modifications. A Preservation System administrator can express policies in a variety of ways to halt changes, such as rejecting all calls to API-M methods. Naturally, an administrator can also shut down the Fedora application entirely; however, this prevents access in addition to modification, which may not be acceptable in some cases.

Supports Event Type Steps

Hardware Test and Maintenance Window, Steps 1
Preservation Application Hardware Environment Replacement: Steps 5, 13
New AIP Format and/or New Preservation Application: Steps 5, 13
Respond to AIP Consistency Failure: Step 1
Respond to Media Failure: Administrative Metadata Store: Step 1
Respond to Unintentional Data Damage: Step 1
Respond to Security Breach: Step 1

Request Service Manager

Description

This service manages a request for a records component in a Preservation System with the retrieval of the appropriate records component and delivery of the component to the requestor. The service should also generate statistical data on the requests it receives and services.

Fedora Support

Fedora supports this service in its *Core* with its Dissemination function. Currently the Core does not automatically generate statistical data on requests and disseminations, but such functionality is currently being examined as part of the research into the Alerting Service being undertaken by the Fedora Preservation Services Working Group.

Supports Event Type Steps

Retrieval Request: Steps 1, 2, 3, 4

Retention and Disposition Module

Description

This service manages the retention and disposition of records in a Preservation System—determining how long records must be kept in the System and what should ultimately happen to the records. In managing the retention period of records, the module should be able to alert Preservation System administrators of records with retention periods about to expire. In managing records disposition, the module should be able locate and execute the disposition—which is often to destroy—of all the components of the appropriate records. It should also be able to provide administrators with reports on disposition activities.

Fedora Support

Fedora does not currently support the definition of record types with retention and disposition rules as part of its *Core*, *Framework*, or *External* services. However, the Fedora Community may be able to articulate record types as content models, building on the community's current work to formalize content model rules.

Supports Event Type Steps

Check Access and Retention Status: Steps 1, 2
Remove Record Component Request: Step 1

Search Service

Description

This service conducts searches against the metadata of records in the Preservation System and provides query results to the requestor. This service should appropriately limit result sets to the requestor's permissions. In addition, the service should also be able to generate statistical data on the queries it conducts.

Fedora Support

Fedora supports this service in its *Core* with Registry, Access, and ResourceIndex functions within the *Core.* The Fedora Search Working Group has recently developed the Generic Search Service and made it available for downloading. [15] The Fedora Development Team has integrated the Generic Search Service into the official Fedora CVS repository and the service will be part of the Fedora 2.2 release as part of the Service Framework in the fourth quarter of 2006.

Supports Event Type Steps

Query Request: Steps 1, 2, 3, 4

Security Audit

Description

This is a set of rules and procedures for systematically checking the effectiveness and reliability of the Security Protocol. Ideally, the Security Audit should be conducted by entities that do not manage or administer the Security Protocol.

Fedora Support

The security audit rules, procedures, and processes are usually managed externally from the Preservation System and repository core it is auditing. Therefore, there is no pressing need to have the audit be a Fedora *Core* or *Framework* service.

Supports Event Type Steps

Security Audit: Steps 1, 2, 3

Security Protocol

Description

This is a set of rules and procedures for protecting the Preservation System from security breaches, attacks, and unauthorized access. A Security Protocol usually incorporates a wide range of measures from network security to physically securing workstations. A Security Protocol also manages user access to records in a Preservation System.

Fedora Support

Fedora can manage access permissions and rules as part of its Core service employing XACML metadata. XACML is an XML-based language that enables a wide variety of access control and security policies to be expressed. The Fedora XACML enforcement module will enforce the rules expressed in such policies. Fedora started utilizing XACML with version 2.1. Fedora also has support for user authentication, via plug-in Tomcat modules, and it also has support for Secure Sockets Layer (SSL) so it can transmit data-streams over a network securely. Network and workstation security activities naturally fall outside the purview of Fedora.

Supports Event Type Steps

[15] "Working Group: Search," <http://www.fedora.info/wiki/index.php/Working_Group:_Search>. Also see "Fedora Generic Search Service Development," <http://defxws2006.cvt.dk/fedoragsearch/>, which has a prototype of the service available for downloading.

Security Audit: Steps 1, 2, 3
Respond to Security Breach: Step 7

Storage Management Module

Description

A module that manages the storage of records components that are in and managed by the repository core of the Preservation System. The module should be able to track and report where records components are stored, the type and age of storage media, and test storage media performance. The module should facilitate a Preservation System administrator's efficient storage of records components. The Storage Management Module is closely related to and can encompass the Data Backup Protocol.

Fedora Support

Fedora is deployed by default with a file system storage module as part of its *Core.* This default module manages the simple storage of data-streams but does not track and manage the type, age, and performance of storage media, and it turn does not manage which data-streams should be stored on which media. However, Fedora addresses this limitation by having its storage module written with a generic interface, allowing it to plug into different backend storage management systems. This allows Preservation System administrators to replace the default storage module with alternatives, such as commercial hierarchical storage systems. Fedora has released a beta version of a plug-in for the Storage Resource Broker (SRB).

Supports Event Type Steps

Report on Media Life: Steps 1, 2
Hardware Test and Maintenance Window: Steps 1, 5
Digital Object Accession: Steps 1, 5
Metadata Update Request: Step 1
Format Transformation Request: Step 5
Remove Record Component Request: Step 1
Preservation Application Hardware Environment Replacement: Step 7
New AIP Format and/or New Preservation Application: Step 7
Refresh Records Component Media: Steps 1, 2, 3, 4, 5, 7, 8, 9
Respond to Media Failure: Record Component Store: Steps 3, 4

System Administration Protocol

Description

A set of rules and procedures for managing the hardware and networks that support the repository core of the Preservation System. This protocol would usually mirror most standard system administration rules and procedures. The Administration Protocol may also include a preservation capabilities reporting function that reports to administrators the current capabilities of the repository, including storage capacity and system performance, among others.

Fedora Support

Fedora does not offer this Protocol as a *Core*, *Framework*, or *External* service. A System Administration Protocol is not a natural fit as a Fedora service and will probably always exist separately from Fedora.

Support Event Type Steps

Hardware Test and Maintenance Window: Step 3
Preservation Application Hardware Environment Replacement: Steps 1, 2, 3, 4, 14, 15

New AIP Format and/or New Preservation Application: Steps 1, 2, 3, 4, 14, 15
New Records Component Store: Steps 1, 2, 3, 4
Respond to Media Failure: Administrative Metadata Store: Steps 2, 3, 4, 5, 6
Respond to Unintentional Data Damage: Step 2

APPENDIX A: EVENT TYPE STEPS TO ABSTRACT SERVICES CROSSWALK

Scheduled Event Types

Incremental Backup of Administrative Metadata

Steps	Abstract Services
1. Perform backup of administrative metadata	Data Backup Protocol
2. Store backup data in a secure location	Data Backup Protocol

Full Backup of Administrative Metadata

Steps	Abstract Services
1. Perform full backup image of Administrative Metadata	Data Backup Protocol
2. Move full backup image to secure off-site location	Data Backup Protocol
3. Retain at least one year's worth of full backup images at the secure off-site location	Data Backup Protocol

Incremental Backup of Records Component Store

Steps	Abstract Services
1. Perform backup of records components	Data Backup Protocol
2. Store backup data in a secure location	Data Backup Protocol

Full Backup of Records Component Store

Steps	Abstract Services
1. Perform a full backup image of the Records Component Store	Data Backup Protocol
2. Move full backup image to secure off-site location	Data Backup Protocol

Verify AIP Consistency

Steps	Abstract Services
1. Verify the internal consistency of each AIP	Integrity Checking Protocol
2. If any AIP fails the consistency check, go to **Respond to AIP Consistency Failure**	*See **Respond to AIP Consistency Failure***

Verify Records Components

Steps	Abstract Services
1. Compute checksums on records components and compare to checksums stored in PDI	PDI Module; Integrity Checking Protocol
2. If any checksums are not correct, go to **Respond to Checksum Failure**	*See **Respond to Checksum Failure***
3. Document "Verify Checksum" event in PDI	PDI Module

Check Access and Retention Status

Steps	Abstract Services
1. Identify all records governed by time-based expiration	Retention and Disposition Module
2. Report all such records to Administration	Retention and Disposition Module; Alerting Service
3. Document "Retention or Access Period Expired" event in PDI	PDI Module

Report on Media Life

Steps	Abstract Services
1. Generate a report listing the types and age of all media in the Records Component Store	Storage Management Module
2. Submit report to Administration	Storage Management Module; Alerting Service

Hardware Test and Maintenance Window

Steps	Abstract Services
1. Activate Hot Spare, if available	Storage Management Module
2. Take down the server in question	Repository Stasis
3. Perform test suite and regular maintenance operations	System Administration Protocol
4. Report test results to Administration	Alerting Service
5. If no immediate problems are discovered, restore server to active service (or Hot Spare status)	Storage Management Module
6. Report test failure to Administration and take corrective action prescribed by Administration	PDI Module; Alerting Service
7. Document test results and maintenance activities in repository history metadata	PDI Module; Repository History

Security Audit

Steps	Abstract Services
1. Hire an independent auditor to review security procedures and protocols, adherence to procedures, physical security, and other security practices	Security Protocol; Security Audit
2. Report security audit results to Administration	Security Protocol; Security Audit
3. Document security audit results in repository history metadata	Security Protocol; PDI Module; Repository History

Irregular Events Types

Digital Object Accession

Steps	Abstract Services
1. Generate Storage Identifier(s)	Persistent Identifier Manager
2. Place records component(s) in Records Component Storage	Storage Management Module
3. Document Storage Identifier(s) in AIP	Persistent Identifier Manager
4. Add "Object Accession" event to PDI history	PDI Module
5. Place AIP and PDI in Administrative Metadata Storage	Storage Management Module
6. Update Data Management Database	Data Management Database
7. Schedule Events based on accession date	Alerting Service

Retrieval Request

Steps	Abstract Services
1. Look up the Storage Identifier for the record component	Request Service Manager
2. Retrieve record component from Records Component Storage	Request Service Manager
3. Provide record component to requesting archive function	Request Service Manager
4. Update retrieval statistics	Request Service Manager

Query Request

Steps	Abstract Services
1. Perform requested query	Search Service
2. Filter results set as necessary according to initiator's permissions	Search Service
3. Provide results set to requesting archive function	Search Service
4. Update query statistics	Search Service

Metadata Update Request

Steps	Abstract Services
1. Add a new metadata bitstream or version existing bitstream as appropriate	Storage Management Module
2. Update AIP with any new Storage Identifiers and fixity information	PDI Module
3. Update PDI with "Metadata Update Event"	PDI Module
4. Update Data Management Database	Data Management Database

Format Transform Request

Steps	Abstract Services
1. Identify all records components or metadata bistreams that are represented in the affected format	Persistent Identifier Manager, Alerting Service
2. For each bitstream, retrieve the bitstream and perform the transformation	Format Transformation Service
3. Validate the file formats of the transformation outputs; report any validation failures to Administration and take corrective action prescribed by Administration	Format Validation Service
4. Generate a Storage Identifier for the newly created bitstream	Persistent Identifier Manager
5. Store the bitstream in the Records Component Store	Storage Management Module

6. Update AIP to include the new bitstream	AIP Module
7. Update PDI with a "Format Transformation" event	PDI Module
8. Update Data Management Database	Data Management Database

Remove Record Component Request

Steps	Abstract Services
1. Confidentially destroy the records component	Retention and Disposition Module, Storage Management Module
2. Update AIP to remove the record component	AIP Module
3. Update PDI to add "Record Component Removed" event	PDI Module
4. Update Data Management Database	Data Management Database

Preservation Application Hardware Environment Replacement

Steps	Abstract Services
1. Continue to maintain Hardware Environment	System Administration Protocol
2. Acquire the new Hardware Environment	System Administration Protocol
3. Set up the new Hardware Environment with a new installation of the Preservation Application and all needed system and utility software	System Administration Protocol
4. Set up new Administrative Metadata Store and Data Management Database (and Records Components Store, if necessary)	System Administration Protocol
5. Place the Preservation Application in stasis on the old Hardware Environment	Repository Stasis
6. Update all PDI to include a "Begin Hardware Environment Replacement" event	PDI Module
7. Copy all AIPs and PDI from the old Administrative Metadata Store to the new Administrative Metadata Store. Also copy Records Components Store if necessary	Storage Management Module
8. Build the new Data Management Database from the new Administrative Metadata Store	Data Management Database
9. Test a representative sample of AIPs on the new system to ensure full functionality	AIP Module
10. After passing tests, update PDI on new server with "End Hardware Environment Replacement" event	PDI Module
11. Update Repository History record with "Hardware Environment Replacement" event	PDI Module
12. Perform full backup image of the new Administrative Metadata Store	Data Backup Protocol
13. Remove Preservation Application from stasis on new Hardware Environment	Repository Stasis
14. Re-designate new Hardware Environment as active	System Administration Protocol
15. Confidentially destroy data on old Hardware Environment	System Administration Protocol

New AIP Format and/or New Preservation Application

Steps	Abstract Services
1. Continue to maintain the old Hardware Environment	System Administration Protocol
2. Acquire a new Hardware Environment	System Administration Protocol
3. Set up the new Hardware Environment with a new installation of the new Preservation Application and all needed system and utility software	System Administration Protocol
4. Set up new Administrative Metadata Store and	System Administration Protocol

Data Management database (and Records Component Store, if necessary)	
5. Place Preservation Application in stasis on old Hardware Environment	Repository Stasis
6. Update all PDI to include "Begin Preservation Application Change" event	PDI Module
7. Transform all AIPs and PDI and submit them to the new repository. Also copy Records Component Store if necessary	Storage Management Module
8. Build the new Data Management database from the new Administrative Metadata Store	Data Management Database
9. Test a representative sample of AIPs on the new system to ensure full functionality; report any test failures to Administration and take corrective action prescribed by Administration	AIP Module; Alerting Service
10. After passing tests, update PDI on new Hardware Environment with "End Repository Application Change" event	PDI Module
11. Update Repository History with "Change Repository Application" event	PDI Module
12. Perform full backup image of the new Administrative Metadata store	Data Backup Protocol
13. Remove Preservation Application from stasis on new Hardware Environment	Repository Stasis
14. Re-designate new Hardware Environment as active	System Administration Protocol
15. Confidentially destroy data on old Hardware Environment	System Administration Protocol

New Records Component Store

Steps	Abstract Services
1. Acquire the new Storage Hardware Environment	System Administration Protocol
2. Set up the new Storage Hardware Environment	System Administration Protocol
3. Test new Storage Hardware Environment	System Administration Protocol
4. Copy the Records Component Store when necessary; See **Preservation Application Hardware Environment Replacement**	System Administration Protocol; *Also see* ***Preservation Application Hardware Environment Replacement***

Add Additional Representation Information

Steps	Abstract Services
1. Accession the digital objects representing the new RI (see **Digital Object Accession**)	*See* ***Digital Object Accession***
2. Update any objects which require links to this RI (see **Metadata Update Request**). This list of objects is provided by Preservation Planning.	S*ee* ***Metadata Update Request***

Change Standard Computing Platform

Steps	Abstract Services
1. Update SCP record (see **Metadata Update Request**)	*See* ***Metadata Update Request***
2. Find all records that are no longer grounded in the Knowledge Base	Knowledge Base Module
3. Report all such records to Preservation Planning	Alerting Service

Refresh Records Component Media

Steps	Abstract Services
1. Prepare new media for use	Storage Management Module
2. Test new media for manufacturing defects	Storage Management Module
3. Copy records components onto new media	Storage Management Module
4. Perform a bit-level comparison between old and new media	Storage Management Module
5. If bit-level test succeeds, redesignate new media as active storage for affected records components	Storage Management Module
6. Update PDI for all affected records with "Media Refresh" event	PDI Module
7. Document when media becomes active	Storage Management Module
8. Confidentially destroy data on old media	Storage Management Module
9. Discard or recycle old media	Storage Management Module

Respond to Checksum Failure

Steps	Abstract Services
1. Mark the record as containing compromised data	PDI Module
2. Alert Administration of this event	Alerting Service
3. Search alternate storage locations to find the record component backups	Data Backup Protocol
4. If an intact, independently stored record component is discovered, then go to **Verify Records Components**, verify component checksum.	Data Recovery Protocol; *See also* ***Verify Records Components***
5. If component checksum does not match, proceed to next appropriate independently stored record component.	Data Recovery Protocol
6. If no intact datastream is found, go to **Respond to Data Loss: Record Component Store**	Data Recovery Protocol; *See also* ***Respond to Data Loss: Record Component Store***
7. If component checksum matches, then replace current record component with alternately stored component in the active Record Component Store, and document that the Archive has repaired the record in its PDI	Data Recovery Protocol
8. Update record PDI "Record Component Repaired" event or "Record Component Corrupted" event	PDI Module

Respond to Media Failure: Record Component Store

Steps	Abstract Services
1. Mark all records with affected components as having corrupted data	PDI Module
2. Look for alternate storage locations for the data stream (such as backups)	Data Backup Protocol
3. Prepare and test new media	Storage Management Module
4. Restore onto new media all found datastreams that positively match their existing integrity information	Storage Management Module
5. Document that the Archive has repaired the records in their PDI	PDI Module
6. If any record components could not be repaired, report to Preservation Planning and go to **Respond to Data Loss: Record Component Store**	Alerting Service; *Also see* ***Respond to Data Loss: Record Component Store***

Respond to Data Loss: Record Component Store

Steps	Abstract Services
1. Notify Archive Administration and Preservation Planning of the data loss	Alerting Service
2. Document data loss in repository history metadata	Repository History
3. Document data loss in the PDI of the affected records	PDI Module

Respond to AIP Consistency Failure

Steps	Abstract Services
1. Place Preservation Application in stasis	Repository Stasis
2. Check all AIPs for consistency. Assume all media upon which consistency failures have occurred has failed; go to **Media Failure: Administrative Metadata Store**	Integrity Checking Protocol

Respond to Media Failure: Administrative Metadata Store

Steps	Abstract Services
1. Place Preservation Application in stasis	Repository Stasis
2. Acquire a new Preservation Application Hardware Environment	System Administration Protocol
3. Set up a new Administrative Metadata Store	System Administration Protocol
4. Set up the new Preservation Application Hardware Environment with a new instance of the Preservation Application and all needed system and utility software	System Administration Protocol
5. Copy all savable AIPs and PDI to the new Administrative Data Store	System Administration Protocol
6. Reconstruct missing AIPs from backup images	System Administration Protocol
7. Document missing or corrupted AIPs that cannot be restored from backup images	PDI Module
8. Document the media failure in the repository history metadata	Repository History
9. Report results of the reconstruction to Archive Administration, who will decide what further actions to take. If any missing or corrupted AIPs cannot be restored from backup images, Archive Administration will probably want to undertake **Respond to Data Loss: Administrative Metadata Store**	Alerting Service; *See also* ***Respond to Data Loss: Administrative Metadata Store***

Respond to Data Loss: Administrative Metadata Store

Steps	Abstract Services
1. Determine all records components "orphaned" by the lost AIP(s)	AIP Module
2. Generate a new "record fragment" AIP for each of the record components, including as much information as can be reconstructed or gathered from the records components, including at least the format type of the components	AIP Module
3. Document information about the data loss in the PDI for the new AIPs	PDI Module
4. Document the data loss in the repository history metadata	PDI Module

Respond to Unintentional Data Damage

Steps	Abstract Services
1. Place Preservation Application in stasis	Repository Stasis
2. Identify the scope and nature of the damage	Integrity Checking Protocol
3. Report to Archive Administration concerning the scope and nature of the damage; Administration will decide the appropriate corrective action	Alerting Service
4. Take corrective action prescribed by Administration; if data loss occurs, undertake **Respond to Data Loss: Administrative Metadata Store** or **Respond to Data Loss: Records Component Store**	*See **Respond to Data Loss: Administrative Metadata Store** or **Respond to Data Loss: Records Component Store***
5. After taking the Preservation Application off stasis, record damage and corrective actions in repository history metadata and the PDI of all affected records	PDI Module

Respond to Security Breach

Steps	Abstract Services
1. Take Preservation System offline; do *not* activate Hot Spares	Repository Stasis
2. Analyze all repository hardware to determine what machines have been compromised and to discover the nature and scope of the attack, and what actions the attacker took while he or she had access to the Hardware Environment	System Administration Protocol
3. Perform internal consistency check of all Administrative Metadata	Integrity Checking Protocol
4. Perform checksum verification of all records components	PDI module
5. Compare Administrative Metadata to a known-good backup image (taken before the attack occurred), and compile a list of all changes between the current image and the backup	Data Recovery Protocol
6. Report all findings to Archive Administration which determines if the attacker was: a. A Squatter (only using computing resources) b. A Vandal (intending to do indiscriminate damage) c. An attacker with motive against the records (intending to alter or destroy records in particular)	Alerting Service
7. Administration decides appropriate corrective actions	Security Protocol
8. Perform prescribed actions; if data loss occurs, undertake **Respond to Data Loss: Administrative Metadata Store** or **Respond to Data Loss: Records Component Store**	*See **Respond to Data Loss: Administrative Metadata Store** or **Respond to Data Loss: Records Component Store***
9. Document security breach in repository history metadata and PDI of all records	PDI Module; Repository History

Fedora and the Preservation of University Records Project

4.1 Analysis of Fedora's Ability to Support Preservation Activities

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/

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:UA069.004.001.00011
(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

TABLE OF CONTENTS

OVERVIEW

This report provides an analysis of Fedora's ability to support preservation activities.[1] This analysis is not as detailed as the Checklist of Fedora's Ability to Support Maintain Activities but is rather a broader examination of Fedora's ability to support the full scope of preservation activities, and not just Data Management and Archival Storage, which is the extent of the Checklist's scope. Rather than comprehensive analyze of Fedora's ability to support every preservation function, this report discusses two general capabilities of Fedora that cut across many preservation functions: security architecture and policy enforcement and resource records. The project team found it fruitful to carefully analyze Fedora's ability to support maintain services because Fedora is a repository architecture that at its core maintains digital objects. As a repository architecture, it presents a platform constructing a highly configurable repository. Thus, preservation functions such as Administration and Preservation Planning become highly dependent on their particular implementation, making a carefully mapped analysis of Fedora's ability to support these functions less fruitful (although still useful).

This report also examines Fedora's planned development and shift from a grant-funded project to a repository architecture maintained by a sustainable, community-supported organization and its ability to continue to support preservation activities. This report also discusses the difficulty of analyzing Fedora as a complete preservation system.

[1] See <www.fedora.info> for more information on Fedora.

ANALYSIS OF FEDORA AS A PRESERVATION SYSTEM

As described in the Project Overview, this project set out to prove the hypothesis that the flexibility and extensibility of the Fedora software would allow it to serve as a Preservation System. Over the course of the project, the project team's focus shifted from asking whether Fedora could serve as a preservation system to working on developing requirements for recordkeeping and preservation, the Ingest Guide, and the Maintain Guide. The focus changed in large part because the project team realized that it was asking the wrong question. Like many other archivists the team was looking for easy solutions to the very difficult problems posed by the long term preservation of electronic records. In hindsight, it seems obvious that no existing software application could serve on its own as a trustworthy preservation system. Preservation is the act of physically and intellectually protecting and technically stabilizing 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. To accomplish this, the preservation system requires natural and juridical people, institutions, applications, infrastructure, and procedures.[2] As a result, Fedora cannot serve as the entire preservation system, but only as a preservation application, which is just a portion of the entire system. Without the appropriate people, infrastructure, policies, and procedures, even the best preservation application cannot ensure preservation.

In serving as the repository application of a preservation system, a Fedora instance (or instances) would be only one of many components that comprise a preservation system. Large portions of ingest and access activities and all preservation planning decisions, among other activities, would occur outside of the Fedora instance. Even though some preservation policies many be articulated and managed in Fedora, an institution still has to formulate these policies—they are not preset in Fedora. Rather than serving as an out-of-box repository solution, Fedora is a repository architecture upon which an institution can build a repository in many different ways. As a result, the suitability of Fedora as the basis of a preservation system depends significantly on its implementation.

The question we should have asked was: "Can a Fedora repository, surrounded by the proper preservation policies, tools, and Fedora services, serve as the basis of a trustworthy preservation system?" We feel the answer to this question is yes. The Fedora core provides a promising basis for a preservation system. Its agnostic view towards file formats and object types enables it to manage essentially any type of file. It has the ability to manage objects with complex—including hierarchical—relationships with its use of RDF or METS metadata. It can manage multiple bitstreams for a single object, which can enable archivists to track and store the original bitstream of an ingested record and the bitstreams of subsequent transformations. It has versioning and persistent identifier capabilities for all content objects, metadata, and disseminators. With Extensible Access Control Markup Language (XACML), an institution can articulate policies to help manage access to records. Fedora is a transparent system and Fedora objects are articulated in XML (usually FOXML or METS), making it feasible to migrate records out of Fedora.

[2] A preservation system has the same components as those of a recordkeeping system.

PLANNED FEDORA DEVELOPMENT

Future Management of Fedora Application

The Fedora Project, run jointly by Cornell University and the University of Virginia, currently manages the Fedora code base. The project is supported by a three-year Andrew W. Mellon Foundation grant that lasts through the end of 2007. This builds on initial design of Fedora at Cornell in 1997, a subsequent prototype implementation at the University of Virginia, and a 2001 Mellon grant that supported Cornell and Virginia's release of Fedora 1.0 in 2003.[3] The Fedora development team is currently developing a plan for moving Fedora beyond the grant-funded project stage to a sustainable, community-based organization that will take the responsibility for maintaining and developing the Fedora software and nurturing an active and growing Fedora community. The project directors have created an advisory board to help the project transition to sustainable organization led by a board of directors. The project directors have also established—or is in the process of establishing—the Outreach Committee to help foster growth of the Fedora community and manage Fedora's promotion; the Architecture Committee, which will manage and develop the specifications for the Fedora Service Framework; and three working groups, Preservation Services, Search Services, and Workflow Services that each investigate and undertake service development in their respective domains.[4]

This transition for Fedora being managed by a grant-funded project team to an organization supported by an active community is a crucial junction for Fedora's long-term ability to support preservation activities. To successfully serve as the repository core of preservation systems—many of which will demand a rigorous and high performance repository—Fedora will need a robust, clean, and well-constructed code base. As new technologies emerge, Fedora will need the organizational infrastructure to ensure that its software code is current and that these updates are smoothly added to existing code. Without a well-tended code base, Fedora's performance will lag and its ability to support scalable preservation activities will suffer.

Its object-oriented architecture and recently developed Service Framework (see below) has enabled Fedora to gracefully incorporate new services developed by the Fedora community. This ability to support new services should allow Fedora to serve as a repository core that can stay current with emerging preservation technologies, techniques, standards, and metadata schemas. However, this depends on a community developing the necessary Framework services that can embody—or at least communicate with—these new technologies, techniques, standards, and schemas. The Fedora community will also have to ensure that existing external tools needed for preservation activities, such as format validators and checksums, are able to become services within the Framework, guaranteeing that they will work smoothly with Fedora. The Fedora community will probably need some degree of leadership, either from a board of directors, the Architecture Committee, or the Preservation Services Working Group to provide a roadmap of preservation needs and priorities for new Fedora services.

The difficulty of moving beyond the project phase is a problem faced by all open-source endeavors and is not unique to Fedora. Like all open-source community-based efforts, Fedora

[3] "History," <http://www.fedora.info/about/history.shtml>.
[4] "The Future of Fedora," <http://www.fedora.info/community/fedorafuture.shtml>.

has the benefit of a diverse community building a range of services. For example, the work of the Search Services and Workflow Services working groups grow directly out of the efforts of particular members of the Fedora community. The community has to date developed a variety of Fedora tools and services.[5] If successful, this will grow into a rich array of tools and services that will allow Fedora users to select the appropriate resource from a sufficiently rich menu of options instead of building their own tools and services. On the other hand, like all open-source efforts, Fedora has the problem of finding an appropriate, dedicated entity to ensure continuity and sustainability. The Fedora project team aims to create that needed entity with its work to create an advisory board and later board of trustees, the committees, and the working groups.

Fedora Service Framework

As of version 2.1 (released February 2006), Fedora development will continue within the Fedora Service Framework, which establishes a structure in which new services that support a Fedora repository instance exist outside and independently of the repository.[6] New functionality for Fedora can be developed as a distinct stand-alone service that will interface gracefully with the core Fedora repository services. This allows new functionality to be developed in a more flexible, modular manner. Most importantly, it does not overburden the core repository software with endless new functionality. The project team expects the Fedora Service Framework to be the focal point of new development for the remainder of the Fedora project and beyond. Both the core development team and the Fedora community will contribute new services to the framework.[7] Two such services that have already been developed are Directory Ingest and OAI Provider. Members of Fedora community are currently developing additional open-source services. The Fedora core development team is also thinking of its own sustainability as an organization and is encouraging users and developers to continue working together to improve the application into the future. As a way of moving towards a development consortium that can sustain Fedora after the grant-funded project ends, the development team has initiated two groups that are of particular importance for preservation, one working on preservation services, the other on workflow.

Fedora Preservation Services Working Group

The Fedora Preservation Services Working Group is currently investigating and developing services to support preservation activities.[8] In 2006 it has focused much of its efforts on creating a messaging service that will support other preservation services in the Fedora Service Framework. The messaging service would serve as a generalized solution for sending messages to repository managers or machines about preservation-related events. The Fedora Development Team has begun work on message-enabling the Fedora Core and the Fedora Service Framework, which lay the architectural foundation for this messaging service. In addition, the Working Group has also spent time examining a variety of other services and their suitability for the Fedora Service Framework or as an external service that smoothly communications with Framework services or the Fedora Core. These services include format transformation, format validation, integrity checking, and repository histories.

[5] "Tools," <http://www.fedora.info/tools/index.shtml>.
[6] "Fedora Service Framework," <www.fedora.info/download/2.1/userdocs/server/features/serviceframework.htm> and <www.fedora.info/wiki/index.php/Fedora_Core_Repository_Service>.
[7] Internal Report from Thornton Staples to project staff, October 10, 2005.
[8] "Working Group: Preservation," <www.fedora.info/wiki/index.php/Working_Group_Preservation>.

Fedora Workflow Services Working Group

The Workflow Services Working Group has been formed in order to design and build a prototype set of business process or workflow orchestration services that could be used to load electronic records into a Fedora repository in a more automated manner. Such automation will be necessary for Archives to reduce the staff time required to process large volumes of electronic records.[9] There are several candidates for a standard to describe business processes using XML, including Business Process Execution Language (BPEL). This should enable an engine to "orchestrate" a business process by executing the steps, by messaging a human to begin a step and waiting for a specific response, or by running a computer program and waiting for the response. Complicated processes can be built up, allowing for concurrent steps or restricting them to be run in series.[10]

[9] "Working Group: Workflow," <http://www.fedora.info/wiki/index.php/Working_Group:_Workflow>.
[10] Internal Report from Thornton Staples to project staff, October 10, 2005.

FEDORA CAPABILITIES IN SUPPORT OF PRESERVATION ACTIVITIES

Most Fedora capabilities that support preservation activities center on maintain activities. The project team evaluates these capabilities in Checklist of Fedora's Ability to Support Maintain Activities. The project team examines two Fedora capabilities that cut across most preservation function: security architecture and policy enforcement and resource records.

Security Architecture and Policy Enforcement

Fedora provides a pluggable authentication module using Tomcat's standard approach to authentication, as well as a new access control module that enforces policies written in eXtensible Access Control Markup Language (XACML), an emerging standard that is beginning to be adopted on many fronts. Fedora has two plug-in modules for authentication: (1) a standard module that authenticates using a file of user identity and role information (i.e., tomcat-users.xml) and (2) an LDAP module to obtain user attributes from an LDAP directory. Fedora also provides an XACML-based policy enforcement module for authorization purposes. The choice of XACML allows institutions to record XML-encoded access control policies in Fedora, rather than in idiosyncratic database or file formats. XACML is very flexible and allows the specification of extremely fine-grained policies.

Fedora supports repository-wide policies, as well as object-specific policies. Policies can be written to permit or deny access to any Fedora API action based on attributes of the user, attributes of digital objects, and attributes of the environment (e.g., current date/time). This means that one can easily shut down all write-access to the repository by setting the policy for API-M at the highest level to deny all, while leaving the policy for API-A to allow all. This would result in a condition in which no changes could be made to the data in the repository, but read-access could continue.

Also, fine-grained object policies can be written to control access to a particular object as a whole, as well as its specific datastreams and disseminations. For example, an object could be set to have no public access until a certain date, or an image object could be set to allow free access to a thumbnail version while restricting all other versions to a specific group, such as a particular University community.

In a workflow that allows web-based users to submit records to a repository, administrators could change the permissions for an object as it progresses through the workflow. For example, once the submission is deemed complete by the originator, all content datastreams could be restricted from being changed, while the metadata datastreams could allow catalogers to update them.

Resource Records

The Ingest and Maintain Guides rely on resource records including resources like format information, record type information, submission agreements, producer metadata, retention schedules, knowledge base metadata, and the history of the repository itself. These supporting records can also be represented in a straightforward manner as Fedora objects with content models describing the kinds of information and abilities expected from such objects. Fedora also

allows datastream versioning, which would be a very important capability for resource records whose content may change periodically. Encoded Archival Context can likely be adapted as a standard way for encoding producer records, but some way to encode information for the remaining supporting record types will have to be developed.

Fedora and the Preservation of University Records Project

4.2 Conclusions and Future Directions

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/

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:UA069.004.001.00012
(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

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 severly

[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).