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