The Consultative Committee for Space Data Systems

Draft Recommendation for
Space Data System Practices

AUDIT AND CERTIFICATION OF TRUSTWORTHY DIGITAL REPOSITORIES

DRAFT RECOMMENDED PRACTICE

CCSDS 652.0-R-1

RED BOOK
October 2009

The Consultative Committee for Space Data Systems

Draft Recommendation for
Space Data System Practices

AUDIT AND CERTIFICATION OF TRUSTWORTHY DIGITAL REPOSITORIES

DRAFT RECOMMENDED PRACTICE

CCSDS 652.0-R-1

RED BOOK

October 2009

AUTHORITY

Issue:	Red Book, Issue 1
Date:	October 2009
Location:	Not Applicable

(WHEN THIS RECOMMENDED PRACTICE IS FINALIZED, IT WILL CONTAIN THE FOLLOWING STATEMENT OF AUTHORITY:)

This document has been approved for publication by the Management Council of the Consultative Committee for Space Data Systems (CCSDS) and represents the consensus technical agreement of the participating CCSDS Member Agencies. The procedure for review and authorization of CCSDS documents is detailed in the *Procedures Manual for the Consultative Committee for Space Data Systems*, and the record of Agency participation in the authorization of this document can be obtained from the CCSDS Secretariat at the address below.

This document is published and maintained by:

CCSDS Secretariat
Space Communications and Navigation Office, 7L70
Space Operations Mission Directorate
NASA Headquarters
Washington, DC 20546-0001, USA

STATEMENT OF INTENT

(WHEN THIS RECOMMENDED PRACTICE IS FINALIZED, IT WILL CONTAIN THE FOLLOWING STATEMENT OF INTENT:)

The Consultative Committee for Space Data Systems (CCSDS) is an organization officially established by the management of its members. The Committee meets periodically to address data systems problems that are common to all participants, and to formulate sound technical solutions to these problems. Inasmuch as participation in the CCSDS is completely voluntary, the results of Committee actions are termed **Recommendations** and are not in themselves considered binding on any Agency.

CCSDS Recommendations take two forms: **Recommended Standards** that are prescriptive and are the formal vehicles by which CCSDS Agencies create the standards that specify how elements of their space mission support infrastructure shall operate and interoperate with others; and **Recommended Practices** that are more descriptive in nature and are intended to provide general guidance about how to approach a particular problem associated with space mission support. This **Recommended Practice** is issued by, and represents the consensus of, the CCSDS members. Endorsement of this **Recommended Practice** is entirely voluntary and does not imply a commitment by any Agency or organization to implement its recommendations in a prescriptive sense.

No later than five years from its date of issuance, this **Recommended Practice** will be reviewed by the CCSDS to determine whether it should: (1) remain in effect without change; (2) be changed to reflect the impact of new technologies, new requirements, or new directions; or (3) be retired or canceled.

In those instances when a new version of a **Recommended Practice** is issued, existing CCSDS-related member Practices and implementations are not negated or deemed to be non-CCSDS compatible. It is the responsibility of each member to determine when such Practices or implementations are to be modified. Each member is, however, strongly encouraged to direct planning for its new Practices and implementations towards the later version of the Recommended Practice.

FOREWORD

This document is a technical Recommendation to use as the basis for providing audit and certification of the trustworthiness of digital repositories. It provides a detailed specification of criteria by which digital repositories shall be audited.

The OAIS Reference Model (reference [1]) contained a roadmap which included the need for a certification standard. The initial work was to be carried out outside CCSDS and then brought back into CCSDS to take into the standard.

In 2003, Research Libraries Group (RLG) and the National Archives and Records Administration (NARA) created a joint task force to specifically address digital repository certification. That task force published the TRAC document (reference [B3]) on which this standard is based.

Through the process of normal evolution, it is expected that expansion, deletion, or modification of this document may occur. This Recommended Practice is therefore subject to CCSDS document management and change control procedures, which are defined in the *Procedures Manual for the Consultative Committee for Space Data Systems*. Current versions of CCSDS documents are maintained at the CCSDS Web site:

http://www.ccsds.org/

Questions relating to the contents or status of this document should be addressed to the CCSDS Secretariat at the address indicated on page i.

At time of publication, the active Member and Observer Agencies of the CCSDS were:

Member Agencies

- Agenzia Spaziale Italiana (ASI)/Italy.
- British National Space Centre (BNSC)/United Kingdom.
- Canadian Space Agency (CSA)/Canada.
- Centre National d'Etudes Spatiales (CNES)/France.
- China National Space Administration (CNSA)/People's Republic of China.
- Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR)/Germany.
- European Space Agency (ESA)/Europe.
- Russian Federal Space Agency (RFSA)/Russian Federation.
- Instituto Nacional de Pesquisas Espaciais (INPE)/Brazil.
- Japan Aerospace Exploration Agency (JAXA)/Japan.
- National Aeronautics and Space Administration (NASA)/USA.

Observer Agencies

- Austrian Space Agency (ASA)/Austria.
- Belgian Federal Science Policy Office (BFSPO)/Belgium.
- Central Research Institute of Machine Building (TsNIIMash)/Russian Federation.
- Centro Tecnico Aeroespacial (CTA)/Brazil.
- Chinese Academy of Sciences (CAS)/China.
- Chinese Academy of Space Technology (CAST)/China.
- Commonwealth Scientific and Industrial Research Organization (CSIRO)/Australia.
- CSIR Satellite Applications Centre (CSIR)/Republic of South Africa.
- Danish National Space Center (DNSC)/Denmark.
- European Organization for the Exploitation of Meteorological Satellites (EUMETSAT)/Europe.
- European Telecommunications Satellite Organization (EUTELSAT)/Europe.
- Geo-Informatics and Space Technology Development Agency (GISTDA)/Thailand.
- Hellenic National Space Committee (HNSC)/Greece.
- Indian Space Research Organization (ISRO)/India.
- Institute of Space Research (IKI)/Russian Federation.
- KFKI Research Institute for Particle & Nuclear Physics (KFKI)/Hungary.
- Korea Aerospace Research Institute (KARI)/Korea.
- Ministry of Communications (MOC)/Israel.
- National Institute of Information and Communications Technology (NICT)/Japan.
- National Oceanic and Atmospheric Administration (NOAA)/USA.
- National Space Organization (NSPO)/Chinese Taipei.
- Naval Center for Space Technology (NCST)/USA.
- Scientific and Technological Research Council of Turkey (TUBITAK)/Turkey.
- Space and Upper Atmosphere Research Commission (SUPARCO)/Pakistan.
- Swedish Space Corporation (SSC)/Sweden.
- United States Geological Survey (USGS)/USA.

PREFACE

This document is a draft CCSDS Recommended Practice. Its 'Red Book' status indicates that the CCSDS believes the document to be technically mature and has released it for formal review by appropriate technical organizations. As such, its technical contents are not stable, and several iterations of it may occur in response to comments received during the review process.

Implementers are cautioned **not** to fabricate any final equipment in accordance with this document's technical content.

DOCUMENT CONTROL

Document	Title	Date	Status
CCSDS 652.0-R-1	Audit and Certification of Trustworthy Digital Repositories, Draft Recommended Practice, Issue 1	October 2009	Current draft

CONTENTS

Section Page

1 INTRODUCTION 1-1

1.1 PURPOSE AND SCOPE 1-1
1.2 APPLICABILITY 1-1
1.3 RATIONALE 1-1
1.4 STRUCTURE OF THIS DOCUMENT 1-2
1.5 SECURITY CONSIDERATIONS 1-3
1.6 DEFINITIONS 1-3
1.7 CONFORMANCE 1-6
1.8 REFERENCES 1-6

2 OVERVIEW OF AUDIT AND CERTIFICATION CRITERIA 2-1

2.1 A TRUSTWORTHY DIGITAL REPOSITORY 2-1
2.2 EVIDENCE 2-1
2.3 RELEVANT STANDARDS, BEST PRACTICES, & CONTROLS 2-1

3 ORGANIZATIONAL INFRASTRUCTURE 3-1

3.1 GOVERNANCE & ORGANIZATIONAL VIABILITY 3-1
3.2 ORGANIZATIONAL STRUCTURE & STAFFING 3-3
3.3 PROCEDURAL ACCOUNTABILITY & PRESERVATION POLICY FRAMEWORK 3-5
3.4 FINANCIAL SUSTAINABILITY 3-10
3.5 CONTRACTS, LICENSES, & LIABILITIES 3-11

4 DIGITAL OBJECT MANAGEMENT 4-1

4.1 INGEST: ACQUISITION OF CONTENT 4-1
4.2 INGEST: CREATION OF THE AIP 4-6
4.3 PRESERVATION PLANNING 4-16
4.4 AIP PRESERVATION 4-19
4.5 INFORMATION MANAGEMENT 4-23
4.6 ACCESS MANAGEMENT 4-24

5 INFRASTRUCTURE AND SECURITY RISK MANAGEMENT 5-1

5.1 TECHNICAL INFRASTRUCTURE RISK MANAGEMENT 5-1
5.2 SECURITY RISK MANAGEMENT 5-12

ANNEX A SECURITY CONSIDERATIONS (NORMATIVE) A-1
ANNEX B REFERENCES (INFORMATIVE) B-1

1 INTRODUCTION

1.1 PURPOSE AND SCOPE

The main purpose of this document is to define a CCSDS Recommended Practice on which to base an audit and certification process for assessing the trustworthiness of digital repositories. The scope of application of this document is the entire range of digital repositories.

1.2 APPLICABILITY

This document is meant primarily for those responsible for auditing digital repositories and also for those who work in or are responsible for digital repositories seeking objective measurement of the trustworthiness of their repository. Some institutions may also choose to use these metrics during a design or redesign process for their digital repository.

1.3 RATIONALE

In 1996 the Task Force on Archiving of Digital Information (reference [B1]) declared, ‘a critical component of digital archiving infrastructure is the existence of a sufficient number of trusted organizations capable of storing, migrating, and providing access to digital collections’. The task force saw that ‘trusted’ or trustworthy organizations could not simply identify themselves. To the contrary, the task force declared, ‘a process of certification for digital archives is needed to create an overall climate of trust about the prospects of preserving digital information’.

Work in articulating responsible digital archiving infrastructure was furthered by the development of the Open Archival Information System (OAIS) Reference Model (reference [1]). Designed to create a consensus on ‘what is required for an archive to provide permanent or indefinite long-term preservation of digital information’, the OAIS addressed fundamental questions regarding the long-term preservation of digital materials that cut across domain-specific implementations. The reference model (ISO 14721) provides a common conceptual framework describing the environment, functional components, and information objects within a system responsible for the long-term preservation of digital materials. Long before it became an approved standard in 2002, many in the cultural heritage community had adopted OAIS as a model to better understand what would be needed from digital preservation systems.

Institutions began to declare themselves ‘OAIS-compliant’ to underscore the trustworthiness of their digital repositories. However, there was no established understanding of ‘OAIS-compliance’ beyond being able to apply OAIS terminology to describe their archive, despite there being a compliance section in OAIS which specifies the need to support the model of information and fulfilling the mandatory responsibilities.

Claims of trustworthiness are easy to make but are thus far difficult to justify or objectively prove. Establishing more clear criteria detailing what a trustworthy repository is and is not has become vital.

In 2002, Research Libraries Group (RLG) and Online Computer Library Center (OCLC) jointly published Trusted Digital Repositories: Attributes and Responsibilities (reference [B2]), which further articulated a framework of attributes and responsibilities for trusted, reliable, sustainable digital repositories capable of handling the range of materials held by large and small cultural heritage and research institutions. The framework was broad enough to accommodate different situations, technical architectures, and institutional responsibilities while providing a basis for the expectations of a trusted repository. The document has proven to be useful for institutions grappling with the long-term preservation of cultural heritage resources and has been used in combination with the OAIS as a digital preservation planning tool. As a framework, this document concentrated on high-level organizational and technical attributes and discussed potential models for digital repository certification. It refrained from being prescriptive about the specific nature of rapidly emerging digital repositories and archives and instead reiterated the call for certification of digital repositories, recommending the development of certification program and articulation of auditable criteria.

OAIS included a Roadmap for follow-on standards which included ‘standard(s) for accreditation of archives’. It was agreed that RLG and National Archives and Records Administration (NARA) would take this particular topic forward and the later published the TRAC (reference [B3]) document which combined ideas from OAIS (reference [1]) and TDR.

The current document follows on from TRAC in order to produce an ISO standard.

1.4 STRUCTURE OF THIS DOCUMENT

This document is divided into informative and normative sections and annexes.

Sections 0-2 of this document are informative and give a high-level view of the rationale, the conceptual environment, some of the important design issues and an introduction to the terminology and concepts.

- Section 0 gives purpose and scope, rationale, a view of the overall document structure, and the acronym list, glossary, and reference list for this document.
- Section 2 provides an overview of audit and certification criteria, ideas about evidence to support claims, and a discussion of related standards.

  Metrics are empirically derived and consistent measures of effectiveness. When evaluated together, metrics can be used to judge the overall suitability of a repository to be trusted to provide a preservation environment that is consistent with the goals of the OAIS. Separately, individual metrics or measures can be used to identify possible weaknesses or pending declines in repository functionality.

– Sections 3 to 5 provide the normative metrics against which a digital repository may be judged. These sections provide metrics grouped as follows:

  - covers Organizational Infrastructure;
  - covers Digital Object Management;
  - covers Infrastructure and Security Risk Management.

  Each section groups metrics into one or more subsections.

– Annex B provides Informative References.

1.5 SECURITY CONSIDERATIONS

Additional security considerations are dealt with in normative annex A.

1.6 DEFINITIONS

1.6.1 ACRONYMS AND ABBREVIATIONS

AIP	Archival Information Package
CCSDS	Consultative Committee for Space Data Systems
DIP	Dissemination Information Package
ISO	International Organization for Standardization
OAIS	Open Archival Information System
PDI	Preservation Description Information
SIP	Submission Information Package
UML	Unified Modeling Language
XML	Extensible Markup Language

1.6.2 TERMINOLOGY

Digital preservation interests a range of different communities, each with a distinct vocabulary and local definitions for key terms. A glossary is included in this document, but it is important to draw attention to the usage of several key terms.

In general, key terms in this document have been adopted from the OAIS Reference Model. One of the great strengths of the OAIS Reference Model has been to provide a common terminology made up of terms ‘not already overloaded with meaning so as to reduce conveying unintended meanings’ (reference [1]). Because the OAIS has become a foundational document for digital preservation, the common terms are well understood and are therefore used within this document.

The OAIS Reference Model uses ‘digital archive’ to mean the organization responsible for digital preservation. In this document, the term ‘repository’ or phrase ‘digital repository’ is

used to convey the same concept in all instances except when quoting from the OAIS. It is important to understand that in all instances in this document, 'repository' and 'digital repository' are used to convey digital repositories and archives that have, or contribute to, long-term preservation responsibilities and functionality. This document uses the OAIS concept of the 'Designated Community'. A repository may have a single, generalized 'Designated Community' (e.g., every citizen of a country), while other repositories may have several, distinct Designated Communities with highly specialized needs, each requiring different functionality or support from the repository; this document uses the term Designated Community to cover this second case also.

Finally, this document names criteria that, combined, evaluate the trustworthiness of digital repositories and archives.

1.6.2.1 Glossary

Unless otherwise indicated, other definitions are taken from the OAIS Reference Model (reference [1]).

Access Policy: Written statement, authorized by the repository management, that describes the approach to be taken by the repository for providing access to objects accessioned into the repository. The Access Policy may distinguish between different types of access rights, for example between system administrators, Designated Communities and general users.

Practice: Actions conducted to execute procedures. Practices are measured by logs or other evidence that record actions completed.

Preservation Implementation Plan: A written statement, authorized by the management of the repository, that describes the services to be offered by the repository for preserving objects accessioned into the repository in accordance with the Preservation Policy.

NOTE – The relationship between these terms is motivated as follows. A repository is assumed to have an overall Repository Mission Statement, part of which will be concerned with preservation. The Preservation Strategic Plan states how the mission will be achieved, in general terms with goals and objectives. The Preservation Policy then declares the range of approaches that the repository will employ to ensure preservation (that is, to implement the Preservation Strategic Plan), and finally the Preservation Implementation Plan translates those into services that the repository must carry out.

Preservation Policy: Written statement, authorized by the repository management, that describes the approach to be taken by the repository for the preservation of objects accessioned into the repository. The Preservation Policy is consistent with the Preservation Strategic Plan.

Preservation Strategic Plan: A written statement, authorized by the management of the repository, that states the goals and objectives for achieving that part of the mission of the

repository concerned with preservation. Preservation Strategic Plans may include long-term and short-term plans.

Procedure: A written statement that specifies actions required to complete a service or to achieve a specific state or condition. Procedures specify how various aspects of the relevant Preservation Implementation Plans are to be fulfilled.

Producer: The role played by those persons, or client systems, who provide the information to be preserved. This can include other OAISes or internal OAIS persons or systems (from reference [1]).

Provider (or Submitter): A person or system that submits a digital object to the repository. The Provider can be the Producer.

Repository Mission Statement: A written statement, authorized by the management of the repository, that, among other things, describes the commitment of the organization for the stewardship of digital objects in its custody.

1.6.3 NOMENCLATURE

The following conventions apply throughout this Recommended Practice:

a) the words ‘shall’ and ‘must’ imply a binding and verifiable specification;

b) the word ‘should’ implies an optional, but desirable, specification;

c) the word ‘may’ implies an optional specification;

d) the words ‘is’, ‘are’, and ‘will’ imply statements of fact.

1.6.4 CONVENTIONS

The following conventions apply:

– The term Designated Community may include multiple Designated Communities.

– A metric may have one or more sub-metrics which pick out particularly important aspects of the broader parent metric. In some cases these sub-metrics also have their own sub-sub-metrics.

– Each metric has one or more of the following informative pieces of text associated with it:

  - Supporting text: giving an explanation of why the metric is important;
  - Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement: providing examples of the evidence which might be examined to test whether the repository satisfies the metric;
  - Discussion: clarifications about the intent of the metric.

1.7 CONFORMANCE

An archive that conforms to this Recommended Practice shall have satisfied the auditor on each of the requirements.

1.8 REFERENCES

The following documents contain provisions which, through reference in this text, constitute provisions of this Recommended Practice. At the time of publication, the editions indicated were valid. All documents are subject to revision, and users of this Recommended Practice are encouraged to investigate the possibility of applying the most recent editions of the documents indicated below. The CCSDS Secretariat maintains a register of currently valid CCSDS documents.

[1] *Reference Model for an Open Archival Information System (OAIS).* Recommendation for Space Data System Standards, CCSDS 650.0-B-1. Blue Book. Issue 1. Washington, D.C.: CCSDS, January 2002. [Also published as ISO 14721:2003.]

NOTE – Informative references are listed in annex B.

2 OVERVIEW OF AUDIT AND CERTIFICATION CRITERIA

This section provides an overview of some of the key concepts that are incorporated in the design of the metrics in this Recommended Practice.

2.1 A TRUSTWORTHY DIGITAL REPOSITORY

At the very basic level, the definition of a trustworthy digital repository must start with 'a mission to provide reliable, long-term access to managed digital resources to its Designated Community, now and into the future' (reference [B2]). Expanding the definition has caused great discussion both within and across various groups, from the broad digital preservation community to the data archives or institutional repository communities.

A trustworthy digital repository will understand threats to and risks within its systems. Constant monitoring, planning, and maintenance, as well as conscious actions and strategy implementation will be required of repositories to carry out their mission of digital preservation. All of these present an expensive, complex undertaking that depositors, stakeholders, funders, the Designated Community, and other digital repositories will need to rely on in the greater collaborative digital preservation environment that is required to preserve the vast amounts of digital information generated now and into the future. Communicating audit results to the public—transparency—will engender more trust, and additional objective audits, potentially leading towards certification, will promote further trust in the repository and the system that supports it. Finally, attaining trustworthy status is not a one-time accomplishment, achieved and forgotten. To retain trustworthy status, a repository will need to undertake a regular cycle of audit and/or certification.

2.2 EVIDENCE

As noted in 1.6.4 each metric has associated with it informative text under the heading *Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement:* providing examples of the evidence which might be examined to test whether the repository satisfies the metric. These examples are illustrative rather than prescriptive, and the lists of possible evidence are not exhaustive.

2.3 RELEVANT STANDARDS, BEST PRACTICES, & CONTROLS

Numerous documents and standards include pieces that are applicable or related to this work. These standards are important to acknowledge and embrace as complementary audit tools. A few examples:

- The ISO 9000 family of standards addresses quality assurance components within an organization and system management that, while valuable, were not specifically developed to gauge the trustworthiness of organizations operating digital repositories.

- Similarly, ISO 17799:2005 was developed specifically to address data security and information management systems. Like ISO 9000, it has some very valuable components to it but it was not designed to address the trustworthiness of digital repositories. Its requirements for information security seek data security compliance to a very granular level, but do not address organizational, procedural, and preservation planning components necessary for the long-term management of digital resources.
- ISO 15489-1:2001 and ISO 15489-2:2001 define a systematic and process-driven approach that governs the practice of records managers and any person who creates or uses records during their business activities, treats information contained in records as a valuable resource and business asset, and protects/preserves records as evidence of actions. Conformance to ISO 15489 requires an organization to establish, document, maintain, and promulgate policies, procedures, and practices for records management, but, by design, addresses records management specifically rather than applying to all types of repositories and archives.
- Finally, ISO 14721:2003, the Open Archival Information System Reference Model, provides a high-level reference model or framework identifying the participants in digital preservation, their roles and responsibilities, and the kinds of information to be exchanged during the course of deposit and ingest into and dissemination from a digital repository.

It is important to acknowledge that there is real value in knowing whether an institution is certified to related standards or meets other controls that would be relevant to an audit.

Certainly, an institution that has undertaken any kind of certification process—even if none of the evaluated components overlap with a digital repository audit—will be better prepared for digital repository certification. And those that have achieved certification in related standards will be able to use those certifications as evidence during the digital repository audit.

3 ORGANIZATIONAL INFRASTRUCTURE

3.1 GOVERNANCE & ORGANIZATIONAL VIABILITY

3.1.1 The repository shall have a mission statement that reflects a commitment to the preservation of, long term retention of, management of, and access to digital information.

Supporting Text

This is necessary in order to ensure commitment to preservation, retention, management and access at the repository's highest administrative level.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Mission statement or charter of the repository or its parent organization that specifically addresses or implicitly calls for the preservation of information and/or other resources under its purview; a legal, statutory, or government regulatory mandate applicable to the repository that specifically addresses or implicitly requires the preservation, retention, management and access to information and/or other resources under its purview.

Discussion

The repository's or its parent organization's mission statement should explicitly address preservation. If preservation is not among the primary purposes of an organization that houses a digital repository then preservation may not be essential to the organization's mission. In some instances a repository pursues its preservation mission as an outgrowth of the larger goals of an organization in which it is housed, such as a university or a government agency, and its narrower mission may be formalized through policies explicitly adopted and approved by the larger organization. Government agencies and other organization may have legal mandates that require they preserve materials, in which case these mandates can be substituted for mission statements, as they define the purpose of the organization. Mission statements should be kept up to date and continue to reflect the common goals and practices for preservation.

3.1.2 The repository shall have a Preservation Strategic Plan that defines the approach the repository will take in the long-term support of its mission.

Supporting Text

This is necessary in order to help the repository make administrative decisions, shape policies and allocate resources in order to successfully preserve its holdings.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Preservation Strategic Plan; meeting minutes; documentation of administrative decisions which have been made.

Discussion

The strategic plan should be based on the organization's established mission, and on its defined values, vision and goals. Strategic plans typically cover a particular finite time period, normally in the 3-5 year range.

3.1.2.1 The repository shall have an appropriate, formal succession plan, contingency plans, and/or escrow arrangements in place in case the repository ceases to operate or the governing or funding institution substantially changes its scope.

Supporting Text

This is necessary in order to preserve the information content entrusted to the repository by handing it on to another custodian in the case that the repository ceases to operate.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Written and credible succession and contingency plan(s); explicit and specific statement documenting the intent to ensure continuity of the repository, and the steps taken and to be taken to ensure continuity; escrow of critical code, software, and metadata sufficient to enable reconstitution of the repository and its content in the event of repository failure; escrow and/or reserve funds set aside for contingencies; explicit agreements with successor organizations documenting the measures to be taken to ensure the complete and formal transfer of responsibility for the repository's digital content and related assets, and granting the requisite rights necessary to ensure continuity of the content and repository services.

Discussion

A repository's failure threatens the long-term sustainability of a repository's information content. It is not sufficient for the repository to have an informal plan or policy regarding where its data goes should a failure occur. A formal plan with identified procedures needs to be in place.

3.1.2.2 The repository shall monitor its organizational environment to determine when to execute its formal succession plan, contingency plans, and/or escrow arrangements.

Supporting Text

This is necessary in order to ensure that the repository can recognize when it is necessary to execute those plans.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Administrative policies, procedures, protocols, requirements; budgets and financial analysis documents; fiscal calendars; business plan(s); any evidence of active monitoring and preparedness.

Discussion

The management of a repository should have formal procedures in place to periodically check on the viability of the repository. This periodic check should be used to determine if, or when, to execute the repository's formal succession plan, contingency plans, and/or escrow arrangements.

3.1.3 The repository shall have a Collection Policy or other document that specifies the type of information it will preserve, retain, manage and provide access to.

Supporting Text

This is necessary in order that the repository has guidance on acquisition of digital content it will preserve, retain, manage and provide access to.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Collection policy and supporting documents; Preservation Policy, mission, goals and vision of the repository.

Discussion

The collection policy can be used to understand what the repository holds, what it does not hold, and why. The collection policy supports the broader mission of the repository. Without such a policy the repository is likely to collect in a haphazard manner, or store large amounts of low-value digital content. The collection policy helps the organization to identify what digital content it will and will not accept for ingestion. In an organization with a broader mission than preservation of digital content the collection policy helps to define the role of the repository within the larger organizational context.

3.2 ORGANIZATIONAL STRUCTURE & STAFFING

3.2.1 The repository shall have identified and established the duties that it needs to perform and shall have appointed staff with adequate skills and experience to fulfill these duties.

Discussion

Staffing of the repository should be by personnel with the required training and skills to carry out the activities of the repository. The repository should be able to document through

development plans, organizational charts, job descriptions and related policies and procedures that the repository is defining and maintaining the skills and roles that are required for the sustained operation of the repository.

3.2.1.1 The repository shall have identified and established the duties that it needs to perform.

Supporting Text

This is necessary in order to ensure that the repository can complete all tasks associated with the long-term preservation and management of the data objects.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

A staffing plan; competency definitions; job descriptions; staff professional development plans; certificates of training and accreditation; plus evidence that the repository reviews and maintains these documents as requirements evolve.

Discussion

Preservation depends upon a range of activities from maintaining hardware and software to migrating content and storage media to negotiating intellectual property rights agreements. In order to ensure long-term sustainability, a repository must be aware of all required activities and demonstrate that it can successfully complete them. The repository can achieve these aims by, for example, identifying the competencies and skill sets required to carry out its activities over time—e.g., archival training, technical skills, and legal expertise.

3.2.1.2 The repository shall have the appropriate number of staff to support all functions and services.

Supporting Text

This is necessary in order to ensure repository staffing levels are adequate for preserving the digital content and providing a secure, quality repository.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Organizational charts; definitions of roles and responsibilities; comparison of staffing levels to industry benchmarks and standards.

Discussion

The repository should determine the appropriate number and level of staff that corresponds to requirements and commitments. The repository should also demonstrate how it evaluates staff effectiveness and suitability to support its functions and services.

3.2.1.3 The repository shall have in place an active professional development program that provides staff with skills and expertise development opportunities.

Supporting Text

This is necessary to ensure that staff skill sets evolve as the repository technology and preservation procedures change.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Professional development plans and reports; training requirements and training budgets, documentation of training expenditures (amount per staff); performance goals and documentation of staff assignments and achievements, copies of certificates awarded.

Discussion

Technology and general practices for digital preservation will continue to change, as will the requirements of its Designated Community, so the repository must ensure that its staff's skill sets evolve. Ideally the repository will meet this requirement through a lifelong learning approach to developing and retaining staff.

3.3 PROCEDURAL ACCOUNTABILITY & PRESERVATION POLICY FRAMEWORK

Documentation assures stakeholders (consumers, producers, and contributors of digital content) that the repository is meeting its requirements and fully performing its role as a trustworthy digital repository. A repository must create documentation that reflects its Mission Statement and Strategic Plan and captures its normal activities. This entails documenting all repository processes, decision-making, and goal setting. Documentation is provided so that the activities of the repository will be understood by stakeholders and management. It ensures that repository policies and procedures are carried out in approved, consistent ways, resulting in long-term preservation and access to digital content in its care. Certification, the clearest indicator of a repository's sound and standards-based practice, is facilitated by procedural accountability and documentation.

3.3.1 The repository shall have defined its Designated Community and associated knowledge base(s) and shall have these definitions appropriately accessible.

Supporting Text

This is necessary in order that it is possible to test that the repository meets the needs of its Designated Community.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

A written definition of the Designated Community.

Discussion

The Designated Community is defined as '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. A Designated Community is defined by the archive and this definition may change/evolve over time' (OAIS Glossary, reference [1]).

Examples of Designated Community definitions include:

- General English-reading public educated to high school and above, with access to a Web Browser (HTML 4.0 capable).
- For GIS data: GIS researchers—undergraduates and above—having an understanding of the concepts of Geographic data and having access to current (2005, USA) GIS tools/computer software, e.g., ArcInfo (2005).
- Astronomer (undergraduate and above) with access to FITS software such as FITSIO, familiar with astronomical spectrographic instruments.
- Student of Middle English with an understanding of TEI encoding and access to an XML rendering environment.
    - Variant 1: Cannot understand TEI;
    - Variant 2: Cannot understand TEI and no access to XML rendering environment;
    - Variant 3: No understanding of Middle English but does understand TEI and XML.
- The repository has defined the external parties, and its assets, owners and uses. Two groups: the publishers of scholarly journals and their readers, each of whom have different rights to access material and different services offered to them.

3.3.2 The repository shall have Preservation Policies in place to ensure its Preservation Strategic Plan will be met.

Supporting Text

This is necessary in order to ensure that the repository can fulfill that part of its mission related to preservation.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Preservation Policies; Repository Mission Statement.

Discussion

Repository policies show how the repository fulfills the requirements of the repository's preservation strategic plan. For example, a preservation strategic plan may contain a

requirement that the repository 'comply with current preferred preservation standards'. The preservation policy might then require that the repository 'monitor current preservation standards and ensure repository compliance with the preferred preservation standards'. In another example the repository may be required by the strategic plan to keep its data understandable. The preservation policy might then include information about the expected level of understandability by the repository's Designated Community for each Archival Information Package.

3.3.2.1 The repository shall have mechanisms for review, update, and ongoing development of its Preservation Policies as the repository grows and as technology and community practice evolve.

Supporting Text

This is necessary in order that the repository has up-to-date, complete policies and procedures in place that reflect the current requirements and practices of its community(ies) for preservation.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Current and past written documentation in the form of Preservation Policies, Preservation Strategic Plans and Preservation Implementation Plans, procedures, protocols, and workflows; specifications of review cycles for documentation; documentation detailing reviews, surveys and feedback. If documentation is embedded in system logic, functionality should demonstrate the implementation of policies and procedures.

Discussion

Preservation Policies capture organizational commitments and intents for staffing, security and other preservation-related concerns. Preservation Implementation Plans address preservation activities and practices such as transfer, submission, quality control, storage management, metadata management, and access and rights management. The repository may find it beneficial to maintain all versions of the preservation policies (e.g., outdated versions are clearly identified and maintained in some organized way) in order to document the results of monitoring for new developments, showing the repository's responsiveness to prevailing standards and practice, emerging requirements, and standards that are specific to the domain, if appropriate; and similar developments. Qualified staff and peers are an important part of the review process, as they help to update and expand these documents. The policies should be understandable by the repository staff in order for them to carry out their work. Preservation Policies and procedures must be demonstrated to be understandable and implementable.

3.3.3 The repository shall have a documented history of the changes to its operations, procedures, software, and hardware.

Supporting Text

This is necessary in order to provide an 'audit trail' through which stakeholders can identify and trace decisions made by the repository.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Capital equipment inventories; documentation of the acquisition, implementation, update, and retirement of critical repository software and hardware file retention and disposal schedules and policies, copies of earlier versions of policies and procedures; minutes of meetings.

Discussion

This documentation may include decisions about the organizational and technical infrastructure. Documentation of or interviews with appropriate staff who can explain repository practices and workflow should be available.

3.3.4 The repository shall commit to transparency and accountability in all actions supporting the operation and management of the repository that affect the preservation of digital content over time.

Supporting Text

This is necessary because transparency, in the sense of being available to anyone who wishes to know, is the best assurance that the repository operates in accordance with accepted standards and practices.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Reports of financial and technical audits and certifications; disclosure of governance documents, independent program reviews, and contracts and agreements with providers of funding and critical services.

Discussion

If the repository uses software to capture information about its history, it should be able to demonstrate these tracking tools. Where appropriate, the history is linked to relevant preservation strategies and describes potential effects on preserving digital content. This requirement does not mean that the organization must make information which would make it vulnerable to competitors available, but rather that the organization commits to disclosing its methods for preserving digital content at least to the Designated Community or other appropriate stakeholder in order to demonstrate that is meeting all current preservation requirements.

3.3.5 The repository shall define, collect, track, and appropriately provide its information integrity measurements.

Supporting Text

This is necessary in order to provide documentation that it has developed or adapted appropriate measures for ensuring the integrity of its holding.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Written definition or specification of the repository's integrity measures (for example, computed checksum or hash value); documentation of the procedures and mechanisms for monitoring integrity measurements and for responding to results of integrity measurements that indicate digital content is at risk; an audit process for collecting, tracking and presenting integrity measurements; Preservation Policy and workflow documentation.

Discussion

The mechanisms to measure integrity will evolve as technology evolves. The repository may provide documentation that it has developed or adapted appropriate measures for ensuring the integrity of its holdings. If protocols, rules and mechanisms are embedded in the repository software, there should be some way to demonstrate the implementation of integrity measures.

3.3.6 The repository shall commit to a regular schedule of self-assessment and external certification.

Supporting Text

This is necessary in order to ensure the repository continues to be trustworthy and there is no threat to its content.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Completed, dated checklists from self-assessments and/or third-party audits; certificates awarded for compliance with relevant ISO standards; timetables and evidence of adequate budget allocations for future certification.

Discussion

A one-time check on trustworthiness is not adequate because many things will change over time.

3.4 FINANCIAL SUSTAINABILITY

3.4.1 The repository shall have short- and long-term business planning processes in place to sustain the repository over time.

Supporting Text

This is necessary in order to ensure the viability of the repository over the period of time it has promised to provide access to its contents for its Designated Community.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Up-to-date, multi-year strategic, operating and/or business plans; audited annual financial statements; financial forecasts with multiple budget scenarios; contingency plans; market analysis.

Discussion

An annual business planning process is commonly accepted as the standard for most organizations.

3.4.2 The repository shall have financial practices and procedures which are transparent, compliant with relevant accounting standards and practices, and audited by third parties in accordance with territorial legal requirements.

Supporting Text

This is necessary in order to guard against malfeasance or other untoward activity that might threaten the economic viability of the repository.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Demonstrated dissemination requirements for business planning and practices; citations to and/or examples of accounting and audit requirements, standards, and practice; audited annual financial statements.

Discussion

The repository cannot simply claim transparency, but should show that it adjusts its business practices to keep them transparent, compliant, and auditable. Confidentiality requirements may prohibit making information about the repository's finances public, but the repository should be able to demonstrate that it is satisfying the needs of its Designated Community.

3.4.3 The repository shall have an ongoing commitment to analyze and report on risk, benefit, investment, and expenditure (including assets, licenses, and liabilities).

Supporting Text

This is necessary in order to demonstrate that the repository has identified and documented these categories, and actively manages them, including identifying and responding to risks, describing and leveraging benefits, specifying and balancing investments, and anticipating and preparing for expenditures.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Risk management documents that identify perceived and potential threats and planned or implemented responses (a risk register); technology infrastructure investment planning documents; cost/benefit analyses; financial investment documents and portfolios; requirements for and examples of licenses, contracts, and asset management; evidence of revision based on risk.

Discussion

The repository should have a goal of maintaining an appropriate balance between risk and benefits, investment and return.

3.5 CONTRACTS, LICENSES, & LIABILITIES

3.5.1 The repository shall have and maintain appropriate contracts or deposit agreements for digital materials that it manages, preserves, and/or to which it provides access.

Supporting Text

This is necessary in order to ensure that the repository has the rights and authorizations needed to enable it to collect and preserve digital content over time, make that information available to its Designated Community, and defend those rights when challenged.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Properly signed and executed deposit agreements and licenses; policies on third-party deposit arrangements; definitions of service levels and permitted uses; repository policies on the treatment of ‘orphan works’ and copyright dispute resolution; reports of independent risk assessments of these policies; procedures for regularly reviewing and maintaining agreements, contracts, and licenses.

Discussion

Repositories may need to show evidence that their contracts are being followed. This is especially important for those with third-party deposit arrangements. These arrangements

may require the repository to guarantee that relevant contracts, licenses, or deposit agreements express rights, responsibilities, and expectations of each party. Contracts and formal deposit agreements should be legitimate; that is, they need to be countersigned and current. When the relationship between depositor and repository is less formal (e.g., a faculty member depositing work in an academic institution's preservation repository), documentation articulating the repository's capabilities and commitments should be provided to each depositor. Repositories engaged in Web harvesting may find this requirement difficult because of the way in which Web-based information is harvested/captured for long-term preservation, and so contracts or deposit agreements are rarely required. Some repositories capture, manage, and preserve access to this material without written permission from the content creators. Others go through the very time-consuming and costly process of contacting content owners before capturing and ingesting information. Ideally, agreements are tracked, linked, managed, and made accessible in a contracts database.

3.5.1.1 The repository shall have contracts or deposit agreements which specify and transfer all necessary preservation rights, and those rights transferred shall be documented.

Supporting Text

This is necessary in order to have sufficient control of the information for preservation and limit the repository's exposure to liability or legal and financial harm.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Contracts, deposit agreements; specification(s) of rights transferred for different types of digital content (if applicable); policy statements on requisite preservation rights.

Discussion

Because the right to change or alter digital information is often restricted by law to the creator, it is important that digital repository contracts and agreements address the need to be able to work with and potentially modify digital objects to keep them accessible. Repository agreements with depositors must specify and/or transfer to the repository certain rights enabling appropriate and necessary preservation actions for the digital objects within the repository. Because legal negotiations can take time, potentially preventing or slowing the ingest of digital objects at risk, it is acceptable for a digital repository to take in or accept digital objects even with only minimal preservation rights using an open-ended agreement and then deal with expanding to detailed rights later.

3.5.1.2 The repository shall have specified all appropriate aspects of acquisition, maintenance, access, and withdrawal in written agreements with depositors and other relevant parties.

Supporting Text

This is necessary in order to ensure that the respective roles of repository, producers and contributors in the depositing of digital content and transfer of responsibility for preservation are understood and accepted by all parties.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Properly executed submission agreements, deposit agreements, and deeds of gift; written standard operating procedures.

Discussion

The deposit agreement specifies all aspects of these issues that are necessary for the repository to carry out its function. There may be a single agreement covering all deposits, or specific agreements for each deposit, or a standard agreement supplemented by special conditions for some deposits. These special conditions may add to the standard agreement or override some aspects of the standard agreement. Agreements may need to cover restrictions on access and will need to cover all property rights in the digital objects. Agreements may place responsibilities on depositors, such as ensuring that Submission Information Packages (SIPs) conform to some pre-agreed standards, and may allow repositories to refuse SIPs that do not meet these standards. Other repositories may take responsibility for fixing errors in SIPs. The division of responsibilities must always be clear. Agreements, written or otherwise, may not always be necessary. The burden of proof is on the repository to demonstrate that it does not need such agreements because, for instance, it has a legal mandate for its activities. An agreement should include, at a minimum, property rights, access rights, conditions for withdrawal, level of security, level of finding aids, SIP definitions, time, volume, and content of transfers. One example of a standard to follow for this is the CCSDS/ISO Producer-Archive Interface Methodology Abstract Standard (reference [B4]).

3.5.1.3 The repository shall have written policies that indicate when it accepts preservation responsibility for contents of each set of submitted data objects.

Supporting Text

This is necessary in order to avoid misunderstandings between the repository and producer/depositor as to when and how the transfer of responsibility for the digital content occurs.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Properly executed submission agreements, deposit agreements, and deeds of gift; confirmation receipt sent back to producer/depositor.

Discussion

If this requirement is not met, there is a risk that, for example, the original is erased before the repository has taken responsibility for the submitted data objects. Without the understanding that the repository has already taken preservation responsibility for the SIP, there is the risk that the producer/depositor may make changes to the data and these would not be properly preserved since they had already been ingested by the repository. For example, for convenience the repository could receive a copy of raw science data from the instrument at the same time the science team gets it, but the science team would have responsibility for it until they turn over responsibility to the final repository. Repositories that report back to their depositors generally will mark this acceptance with some form of notification (for example, confirmation receipts) to the depositor. (This may depend on repository responsibilities as designated in the depositor agreement.) A repository may mark the transfer by sending a formal document, often a final signed copy of the transfer agreement, back to the depositor signifying the completion of the transformation from SIP to AIP process. Other approaches are equally acceptable. Brief daily updates may be generated by a repository that only provides annual formal transfer reports.

3.5.1.4 The repository shall have policies in place to address liability and challenges to ownership/rights.

Supporting Text

This is necessary in order to minimize potential liability and challenges to the rights of the repository.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

A definition of rights, licenses and permissions to be obtained from producers and contributors of digital content; citations to relevant laws and regulations; policy on responding to challenges; documented track record for responding to challenges in ways that do not inhibit preservation; records of relevant legal advice sought and received.

Discussion

The repository's Preservation Policies and Preservation Implementation Plans and mechanisms should be vetted by appropriate institutional authorities and/or legal experts to ensure that responses to challenges adhere to relevant laws and requirements, and that the potential liability for the repository is minimized.

3.5.2 The repository shall track and manage intellectual property rights and restrictions on use of repository content as required by deposit agreement, contract, or license.

Supporting Text

This is necessary in order to allow the repository to track, act on, and verify rights and restrictions related to the use of the digital objects within the repository.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

A Preservation Policy statement that defines and specifies the repository's requirements and process for managing intellectual property rights; depositor agreements; samples of agreements and other documents that specify and address intellectual property rights; documentation of monitoring by repository over time of changes in status and ownership of intellectual property in digital content held by the repository; results from monitoring, metadata that captures rights information.

Discussion

The repository should have a mechanism for tracking licenses and contracts to which it is obligated. Whatever the format of the tracking system, it must be sufficient for the institution to track, act on, and verify rights and restrictions related to the use of the digital objects within the repository.

4 DIGITAL OBJECT MANAGEMENT

4.1 INGEST: ACQUISITION OF CONTENT

4.1.1 The repository shall identify the Content Information and the Information Properties that the repository will preserve.

Supporting Text

This is necessary in order to make it clear to funders, depositors and users what responsibilities the repository is taking on and what aspects are excluded. It is also a necessary step in defining the information which is needed from the information producers or depositors.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Mission statement; submission agreements/deposit agreements/deeds of gift; workflow and Preservation Policy documents, including written definition of properties as agreed in the deposit agreement/deed of gift; written processing procedures; documentation of properties to be preserved.

Discussion

This process begins in general with the repository's mission statement and may be further specified in pre-accessioning agreements with producers or depositors (e.g., producer-archive agreements) and made very specific in deposit or transfer agreements for specific digital objects and their related documentation. For example, one repository may only commit to preserving the textual content of a document and not its exact appearance on a screen. Another may wish to preserve the exact appearance and layout of textual documents, while others may choose to keep the units of the measurement of data fields and to normalize the data during the ingest process. If unique identifiers are associated with digital objects before ingest, they may also be properties that need to be preserved.

4.1.1.1 The repository shall have a procedure(s) for identifying those Information Properties that it will preserve.

Supporting Text

This is necessary to establish a clear understanding with depositors, funders, and the repository's Designated Communities how the repository determines and checks what the characteristics and properties of preserved items will be over the long term. These procedures will be necessary to confirm authenticity or to identify erroneous claims of authenticity of the preserved digital record.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Definitions of the Information Properties which should be preserved; submission agreements/deposit agreements, Preservation Policies, written processing procedures, workflow documentation.

Discussion

These procedure(s) document the methods and factors a repository uses to determine the aspects of different types of Content Information for which it accepts preservation responsibility to its designated communities. For example, a repository's procedure may be to use file formats in order to determine the properties it will preserve unless otherwise specified in a deposit agreement. In this case, the repository would be able to demonstrate provenance for objects that may have been the same file format when received but are preserved differently over the long term.

4.1.1.2 The repository shall have a record of the Content Information and the Information Properties that it will preserve.

Supporting Text

This is necessary in order to identify in writing the Content Information of the records for which it has taken preservation responsibility and the Information Properties it has committed to preserve for those records based on their Content Information.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Preservation Policies, processing manuals, collection inventories or surveys, logs of Content Information types, acquired preservation strategies, and action plans.

Discussion

The repository must demonstrate that it establishes and maintains an understanding of its digital collections sufficient to carry out the preservation necessary to persist the properties to which it has committed. The repository can use this information to determine the effectiveness of its preservation activities over time.

4.1.2 The repository shall clearly specify the information that needs to be associated with specific Content Information at the time of its deposit.

Supporting Text

This is necessary in order that there is a clear understanding of what needs to be acquired from the Producer.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Transfer requirements; producer-archive agreements; workflow plans to produce the AIP.

Discussion

For most types of digital objects to be ingested, the repository should have written criteria, prepared by the repository on its own or in conjunction with other parties, that specify exactly what digital object(s) are transferred, what documentation is associated with the object(s), and any restrictions on access, whether technical, regulatory, or donor-imposed. These criteria document what information the repository and its designated communities may expect for digital object(s) upon deposit. The depositor may be a harvesting process created by the repository. The level of precision in these specifications will vary with the nature of the repository's collection policy and its relationship with creators. For instance, repositories engaged in Web harvesting, or those that rescue digital materials long after their creators have abandoned them, cannot impose conditions on the creators of material, since they are not 'depositors' in the usual sense of the word. But Web harvesters can, for instance, decide which metadata elements from the HTTP transactions that captured a site are to be preserved along with the site's files, and this still constitutes 'information associated with the digital material'. They may also choose to record the information or decisions—whether taken by humans or by automated algorithms—that led to the site's being captured. The repository can check what it receives from the producer based on the specifications.

4.1.3 The repository shall have adequate specifications enabling recognition and parsing of the SIPs.

Supporting Text

This is necessary in order to be sure that the repository is able to extract information from the SIPs.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Packaging Information for the SIPs; Representation Information for the SIP Content Data, including documented file format specifications; published data standards; documentation of valid object construction.

Discussion

The repository must be able to determine what the contents of a SIP are with regard to the technical construction of its components. For example, the repository needs to be able to recognize a TIFF file and confirm that it is not simply a file with a filename ending in 'TIFF'. Another example, would be a website for which the repository would need to be able to recognize and test the validity of the variety of file types (e.g., HTML, images, audio, video, CSS, etc.) that are part of the website. This is necessary in order to confirm: 1) the SIP

is what the repository expected; 2) the Content Information is correctly identified; and 3) the properties of the Content Information to be preserved have been appropriately selected.

4.1.4 The repository shall have mechanisms to appropriately verify the identity of the Producer of all materials.

Supporting Text

This is necessary in order to avoid providing erroneous provenance to the information which is preserved.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Legally binding submission agreements/deposit agreements/deeds of gift, evidence of appropriate technological measures; logs from procedures and authentications.

Discussion

The repository's written standard operating procedures and actual practices must ensure the digital objects are obtained from the expected depositor. Examples of a Producer include persons, organizations, corporate entities, or harvesting processes. Different repositories will adopt different levels of proof needed; the Designated Community should have the opportunity to review the evidence.

4.1.5 The repository shall have an ingest process which verifies each SIP for completeness and correctness.

Supporting Text

This is necessary in order to detect and correct errors in the SIP when created and potential transmission errors between the depositor and the repository.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Appropriate Preservation Policy and Preservation Implementation Plan documents and system log files from system(s) performing ingest procedure(s); logs or registers of files received during the transfer and ingest process; documentation of standard operating procedures, detailed procedures, and/or workflows; format registries; definitions of completeness and correctness.

Discussion

Information collected during the ingest process must be compared with information from some other source to verify the correctness of the data transfer and ingest process. Other sources will include technical and descriptive metadata obtained prior to ingest and may also include expectations set by the depositor, the object producer, a format registry, or the

repository's own expectations. The extent to which a repository can determine correctness will depend on what it knows about the SIP and what tools are available for verifying correctness. It can mean simply checking that file formats are what they claim to be (TIFF files are valid TIFF format, for instance), or can imply checking the content. This might involve human checking in some cases, such as confirming that the description of a picture matches the image. This allows the repository to demonstrate that its preserved objects have completely and correctly copied what it intended to copy from the SIPs. It also allows the repository to document reasons for other SIP-related actions such as rejecting the transfer, suspending processing until the missing information is received, or simply reporting the errors. Similarly, the definition of 'completeness' should be appropriate to a repository's activities. If an inventory of files was provided by a producer as part of pre-ingest negotiations, one would expect checks to be carried out against that inventory. Whatever checks are carried out must be consistent with the repository's own documented definition and understanding of completeness and correctness. One thing that a repository might want to do is check for network drop out or other corruption during the transmission process.

4.1.6 The repository shall obtain sufficient control over the Digital Objects to preserve them.

Supporting Text

This is necessary in order to ensure that the most basic type of preservation, namely bit preservation, is assured.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Documents showing the level of physical control the repository actually has. A separate database/metadata catalog listing all of the digital objects in the repository and metadata sufficient to validate the integrity of those objects (file size, checksum, hash, location, number of copies, etc.)

Discussion

The repository must obtain complete control of the bits of the digital objects conveyed with each SIP. For example, in cases where SIPs only reference digital objects, the repository must also reference the digital objects or preserve them if the current repository is not committed to such preservation.

4.1.7 The repository shall provide the producer/depositor with appropriate responses at agreed points during the ingest processes.

Supporting Text

This is necessary in order to ensure that the producer can verify that there are no inadvertent lapses in communication which might otherwise allow loss of SIPs.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Submission agreements/deposit agreements/deeds of gift; workflow documentation; standard operating procedures; evidence of 'reporting back' such as reports, correspondence, memos, or emails.

Discussion

Based on the initial processing plan and agreement between the repository and the producer/depositor, the repository must provide the producer/depositor with progress reports at agreed points throughout the ingest process. Repository responses can range from nothing at all to predetermined, periodic reports of the ingest completeness and correctness, error reports and any final transfer of custody document. Producers/Depositors can request further information on an ad hoc basis when the previously agreed upon reports are insufficient.

4.1.8 The repository shall have contemporaneous records of actions and administration processes that are relevant to content acquisition.

Supporting Text

This is necessary to ensure that such documentation, which may be needed in an audit, is captured and is accurate and authentic.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Written documentation of decisions and/or action taken; preservation metadata logged, stored, and linked to pertinent digital objects, confirmation receipts sent back to providers.

Discussion

These records should be created on or about the time of the actions they refer to and are related to actions taken during the Ingest: Acquisition of Content process (4.1). The records may be automated or may be written by individuals, depending on the nature of the actions described. Where community or international standards are used, the repository must demonstrate that all relevant actions are carried through.

4.2 INGEST: CREATION OF THE AIP

4.2.1 The repository shall have for each AIP or class of AIPs preserved by the repository an associated definition that is adequate for parsing the AIP and fit for long-term preservation needs.

Supporting Text

This is necessary to ensure that the AIP and its associated definition, including appropriate Packaging Information, can always be found, processed and managed within the archive.

4.2.1.1 The repository shall be able to identify which definition applies to which AIP.

Supporting Text

This is necessary to ensure that the appropriate definition is used when parsing/interpreting an AIP.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Documentation clearly linking each AIP, or class of AIPs, to its definition.

Discussion

The repository may use any method for associating the definitions and the AIPs that provides for the continued and continuous linkage of the two entities.

4.2.1.2 The repository shall have a definition of each AIP that is adequate for long-term preservation, enabling the identification and parsing of all the required components within that AIP.

Supporting Text

This is necessary in order to explicitly show that the AIPs are fit for their intended purpose, that each component of an AIP has been adequately conceived and executed and the plans for the maintenance of each AIP are in place. (See 4.3, Preservation Planning, below.)

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Demonstration of the use of the definitions to extract Content Information and PDI from AIPs.

Discussion

Documentation should identify each class of AIP and describe how each is implemented within the repository. Implementations may, for example, involve some combination of files, databases, and/or documents. Documentation shall relate the AIP component's contents to the related preservation needs of the repository, with enough detail for the repository's providers and consumers to be confident that the significant properties of AIPs will be preserved. Documentation should clearly show that AIP components such as Representation Information and Provenance can be managed and kept up to date. The repository should clearly identify when new versions of AIPs need to be created in order to keep them fit for purpose. The external dependencies of the AIP should also be recorded.

Definitions should exist for each AIP, or class of AIP if there are many instances of the same type. Repositories that store a wide variety of object types may need a specific definition for each AIP they hold, but it is expected that most repositories will establish class descriptions that apply to many AIPs. It must be possible to determine which definition applies to which

AIP. It may also be necessary for the definitions to say something about the semantics or intended use of the AIPs if this could affect long-term preservation decisions. For example, two repositories might both preserve only digital still images, both using multi-image TIFF files as their preservation format. Repository 1 consists entirely of real-world photographic images intended for viewing by people and has a single definition covering all of its AIPs. (The definition may refer to a local or external definition of the TIFF format.) Repository 2 contains some images, such as medical x-rays, that are intended for computer analysis rather than viewing by the human eye, and other images that are like those in Repository 1. Repository 2 should perhaps define two classes of AIPs, even though it only uses one storage format for both. A future preservation action may depend on the intended use of the image—an action that changes the bit-depth of the image in a way that is not perceivable to the human eye may be satisfactory for real-world photographs but not for medical images, for example. An AIP contains these key components: the primary data object to be preserved, its supporting Representation Information (format and meaning of the format elements), and the various categories of Preservation Description Information (PDI) that also need to be associated with the primary data object: Fixity, Provenance, Context, and Reference. There should be a definition of how these categories of information are linked.

4.2.2 The repository shall have a description of how AIPs are constructed from SIPs.

Supporting Text

This is necessary in order to ensure that the AIP(s) adequately represents the information in the SIP(s).

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Process description documents; documentation of the SIP-AIP relationship; clear documentation of how AIPs are derived from SIPs.

Discussion

In some cases, the AIP and SIP will be almost identical apart from packaging and location, and the repository need only state this. In other cases, complex transformations (e.g., data normalization) may be applied to objects during the ingest process, and a precise description of these actions may be necessary to reflect how the AIP(s) has been adequately transformed from the information in the SIP(s). The AIP construction description should include documentation that gives a detailed description of the ingest process for each SIP to AIP transformation, typically consisting of an overview of general processing being applied to all such transformations, augmented with description of different classes of such processing and, when applicable, with special transformations that were needed.

Some repositories may need to produce these complex descriptions case by case. Under such circumstances case diaries or logs of actions taken to produce each AIP should be created and maintained. In these cases, documentation should be mapped to individual AIPs, and the mapping should be available for examination. Other repositories that can run a more

production-line approach may have a description for how each class of incoming objects is transformed to produce the AIP. It must be clear which definition applies to which AIP. If, to take a simple example, two separate processes each produce a TIFF file, it must be clear which process was applied to produce a particular TIFF file.

4.2.3 The repository shall document the final disposition of all SIPs.

In particular the following aspect must be checked.

4.2.3.1 The repository shall follow documented procedures if a SIP is not incorporated into an AIP or discarded and shall indicate why the SIP was not incorporated or discarded.

Supporting Text

This is necessary in order to ensure that the SIPs received have been dealt with appropriately, and in particular have not been accidentally lost.

Examples of Ways the Repository can Demonstrate it is Meeting these Requirements

System processing files; disposal records; donor or depositor agreements/deeds of gift; provenance tracking system; system log files; process description documents; documentation of SIP relationship to AIP; clear documentation of how AIPs are derived from SIPs; documentation of standard/process against which normalization occurs; documentation of normalization outcome and how the resulting AIP is different from the SIP(s).

Discussion

The timescale of this process will vary between repositories from seconds to many months, but SIPs must not remain in an unprocessed limbo-like state forever. The accessioning procedures and the internal processing and audit logs should maintain records of all internal transformations of SIPs to demonstrate that they either become AIPs (or part of AIPs) or are disposed of. Appropriate descriptive information should also document the provenance of all digital objects.

4.2.4 The repository shall have and use a convention that generates persistent, unique identifiers for all AIPs.

In particular the following aspects must be checked.

4.2.4.1 The repository shall uniquely identify each AIP within the repository.

4.2.4.1.1 The repository shall have unique identifiers.

4.2.4.1.2 The repository shall assign and maintain persistent identifiers of the AIP and its components so as to be unique within the context of the repository.

4.2.4.1.3 Documentation shall describe any processes used for changes to such identifiers.

4.2.4.1.4 The repository shall be able to provide a complete list of all such identifiers and do spot checks for duplications.

4.2.4.1.5 The system of identifiers shall be adequate to fit the repository's current and foreseeable future requirements such as numbers of objects.

Supporting Text

This is necessary in order to ensure that each AIP can be unambiguously found in the future. This is also necessary to ensure that each AIP can be distinguished from all other AIPs in the repository.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Documentation describing naming convention and physical evidence of its application (e.g., logs).

4.2.4.2 The repository shall have a system of reliable linking/resolution services in order to find the uniquely identified object, regardless of its physical location.

Supporting Text

This is necessary in order that actions relating to AIPs can be traced over time, over system changes, and over storage changes.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Documentation describing naming convention and physical evidence of its application (e.g., logs).

Discussion

A repository needs to ensure that there is in place an accepted, standard naming convention that identifies its materials uniquely and persistently for use both in and outside the repository. The 'visibility' requirement here means 'visible' to repository managers and auditors. It does not imply that these unique identifiers need to be visible to end users or that they serve as the primary means of access to digital objects. Ideally, the unique ID lives as

long as the AIP; if it does not, there must be traceability. Subsection 4.2.1 requires that the components of an AIP be suitably bound and identified for long-term management, but places no restrictions on how AIPs are identified with files. Thus, in the general case, an AIP may be distributed over many files, or a single file may contain more than one AIP. Therefore identifiers and filenames may not necessarily correspond to each other. Documentation must represent these relationships.

4.2.5 The repository shall have access to necessary tools and resources to provide authoritative Representation Information for all of the digital objects it contains.

In particular the following aspects must be checked.

4.2.5.1 The repository shall have tools or methods to identify the file type of all submitted Data Objects.

4.2.5.2 The repository shall have tools or methods to determine what Representation Information is necessary to make each Data Object understandable to the Designated Community.

4.2.5.3 The repository shall have access to the requisite Representation Information.

4.2.5.4 The repository shall have tools or methods to ensure that the requisite Representation Information is persistently associated with the relevant Data Objects.

Supporting Text

This is necessary in order to ensure that the repository's digital objects are understandable to the Designated Community.

Examples of Ways the Repository can Demonstrate it is Meeting these Requirements

Subscription or access to registries of Representation Information (including format registries); viewable records in local registries (with persistent links to digital objects); database records that include Representation Information and a persistent link to relevant digital objects.

Discussion

These tools and resources can be held internally or can be shared via, for example, a trusted set of registries. However, this requirement does not demand that each repository has such tools and resources, merely that it has access to them. For example a repository may access

external registries.[1] Any such registry is a specialized type of repository, which itself must be certified/trustworthy. The repository may use these types of standardized, authoritative information sources to identify and/or verify the Representation Information components of Content Information and PDI. This will reduce the long-term maintenance costs to the repository and improve quality control. Sometimes there is both general Representation Information (e.g., format information) and specific Representation Information (e.g., meanings of individual fields within a dataset). Often the general information will be available in an external repository, but the local repository may need to maintain the instance-specific information. It is likely that many repositories would wish to keep local copies of relevant Representation Information; however, this may not be practical in all cases. Even where a repository strives to keep all such information locally there may be, for example, a schedule of updates which means that until an update is performed, the local Representation Information is incomplete. This may be regarded as a kind of local caching of, for example, the Representation Information held in registries. Alternatively one may say that in these cases, the use of international registries is not meant to replace local registries but instead serve as a resource to verify or obtain independent, authoritative information about any and all Representation Information. Good practice suggests that any locally held Representation Information should also be made available to other repositories via a trusted registry. In addition any item of Representation Information should itself have adequate Representation Information to ensure that the Designated Community can understand and use the data object being preserved.

4.2.6 The repository shall have documented processes for acquiring Preservation Description Information (PDI) for its associated Content Information and acquire PDI in accordance with the documented processes.

In particular the following aspects must be checked.

4.2.6.1 The repository shall have documented processes for acquiring PDI.

4.2.6.2 The repository shall execute its documented processes for acquiring PDI.

4.2.6.3 The repository shall ensure that the PDI is persistently associated with the relevant Content Information.

Supporting Text

This is necessary in order to ensure that an auditable trail to support claims of authenticity is available, that unauthorized changes to the digital holdings can be detected, and that the digital objects can be identified and placed in their appropriate context.

---

[1] The Unified Digital Formats Registry (UDFR, http://www.gdfr.info/udfr.html) and the UK Digital Curation Centre's Registry Repository of Representation Information (RRORI, http://registry.dcc.ac.uk) are two emerging examples.

Examples of Ways the Repository can Demonstrate it is Meeting these Requirements

Standard operating procedures; manuals describing ingest procedures; viewable documentation on how the repository acquires and manages Preservation Description Information (PDI); creation of checksums or digests, consulting with Designated Community about Context.

Discussion

PDI is needed not only by the repository to help ensure the Content Information is not corrupted (Fixity) and is findable (Reference Information), but to help ensure the Content Information is adequately understandable by providing a historical perspective (Provenance Information) and by providing relationships to other information (Context Information). The extent of such information needs is best addressed by members of the Designated Community(ies). The PDI must be permanently associated with Content Information.

4.2.7 The repository shall ensure that the Content Information of the AIPs is understandable for their Designated Community at the time of creation of the AIP.

In particular the following aspects must be checked.

4.2.7.1 Repository shall have a documented process for testing understandability for their Designated Communities of the Content Information of the AIPs at their creation.

4.2.7.2 The repository shall execute the testing process for each class of Content Information of the AIPs.

4.2.7.3 The repository shall bring the Content Information of the AIP up to the required level of understandability if it fails the understandability testing.

Supporting Text

This is necessary in order to ensure that one of the primary tests of preservation, namely that the digital holdings are understandable by their Designated Community, can be met. (See 4.3 for additional requirements for understandability beyond ingest.)

Examples of Ways the Repository can Demonstrate it is Meeting these Requirements

Test procedures to be run against the digital holdings to ensure their understandability to the defined Designated Community; records of such tests being performed and evaluated; evidence of gathering or identifying Representation Information to fill any intelligibility gaps which have been found; retention of individuals with the discipline expertise.

Discussion

This requirement is concerned with the understandability of the AIP. If the ingested material is not understandable, the repository needs to ingest or make available additional information

to make sure that the AIPs are understandable to the Designated Community(ies). For example, if documents are written in a dying language and the Designated Community is no longer able to understand the language the documents are written in, the repository would need to provide additional documentation that would allow the Designated Community to understand the documents (e.g., translations of the documents in a language the Designated Community could understand or dictionaries that would allow the Designated Communities to translate the documents into a language its members understand).

4.2.8 The repository shall verify each AIP for completeness and correctness at the point it is created.

Supporting Text

This is necessary in order to ensure that what is maintained over the long term is as it should be and can be traced to the information provided by the Producers.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Description of the procedure that verifies completeness and correctness of the AIPs; logs of the procedure.

Discussion

The repository should be sure that the AIPs it creates are as they are expected to be by checking them against the associated definition for each AIP or class of AIP (see 4.2.1) and the description of how AIPs are constructed from SIPs (see 4.2.2). If the repository has a standard process to verify SIPs for both completeness and correctness and a demonstrably correct process for transforming SIPs into AIPs, then it simply needs to demonstrate that the initial checks were carried out successfully and that the transformation process was carried out without indicating errors. On the other hand repositories that must create unique processes for many of their AIPs will also need to generate unique methods for validating the completeness and correctness of AIPs. This may include performing tests of some sort on the content of the AIP that can be compared with tests on the SIP. Such tests might be simple (counting the number of records in a file, or performing some simple statistical measure), but they might be complex. Documentation should describe how the completeness and correctness of AIPs is ensured, starting with receipt from the producer and continuing through AIP creation and supporting long-term preservation. Example approaches include the use of checksums, testing that checksums are still correct at various points during ingest and preservation, logs that such checks have been made, and any special tests that may be required for a particular AIP instance or class.

4.2.9 The repository shall provide an independent mechanism for verifying the integrity of the repository collection/content.

Supporting Text

This is necessary to enable the audit of the integrity of the collection as a whole.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Documentation provided for 4.2.1 through 4.2.4; documented agreements negotiated between the producer and the repository (see 4.1.1-4.1.8); logs of material received and associated action (receipt, action, etc.) dates; logs of periodic checks.

Discussion

It is the responsibility of the repository to choose the appropriate mechanism for checking the completeness and correctness of its collections. In general, it is likely that a repository that meets all the previous criteria will satisfy this one without needing to demonstrate anything more. As a separate requirement, it demonstrates the importance of being able to audit the integrity of the collection as a whole. For example, if a repository claims to have all e-mail sent or received by The Yoyodyne Corporation between 1985 and 2005, it has been required to show that:

- the content it holds came from Yoyodyne's e-mail servers;
- it is all correctly transformed into a preservation format;
- each monthly SIP of e-mail has been correctly preserved, including original unique identifiers such as Message-IDs.

However, it may still have no way of showing whether this really represents all of Yoyodyne's email. For example, if there is a three-day period with no messages in the repository, is this because Yoyodyne was shut down for those three days, or because the e-mail was lost before the SIP was constructed? This case could be resolved by the repository's amending its description of the collection, but other cases may not be so straightforward. A familiar mechanism from the world of traditional materials in libraries and archives is an accessions or acquisitions register that is independent of other catalog metadata. A repository should be able to show, for each item in its accessions register, which AIP(s) contain content from that item. Alternatively, it may need to show that there is no AIP for an item, either because ingest is still in progress, or because the item was rejected for some reason. Conversely, any AIP should be able to be related to an entry in the acquisitions register.

4.2.10 The repository shall have contemporaneous records of actions and administration processes that are relevant to AIP creation.

Supporting Text

This is necessary in order to ensure that there is omitted from the record nothing relevant that might be needed to provide an independent means to verify that all AIPs have been properly created in accord with the documented procedures (see 4.2.1 through 4.2.9). It is the responsibility of the repository to justify its practice in this respect.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Written documentation of decisions and/or action taken with timestamps; preservation metadata logged, stored, and linked to pertinent digital objects.

Discussion

These records must be created on or about the time of the actions they refer to and are related to actions associated with AIP creation. The records may be automated or may be written by individuals, depending on the nature of the actions described. Where community or international standards are used, the repository must demonstrate that all relevant actions are carried through.

4.3 PRESERVATION PLANNING

4.3.1 The repository shall have documented preservation strategies relevant to its holdings.

Supporting Text

This is necessary in order that it is clear how the repository plans to ensure the information will remain available and usable for future generations and to provide a means to check and validate the preservation work of the repository.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Documentation identifying each preservation risk identified and the strategy for dealing with that risk.

Discussion

These documented preservation strategies will describe how the repository will act upon identified risks, as part of the preservation strategic plan. These preservation strategies and the preservation strategic plan will typically address the degradation of storage media, the obsolescence of media drives, and the obsolescence or inadequacy of Representation Information (including formats) as the knowledge base of the Designated Community changes, and safeguards against accidental or intentional digital corruption. For example, if

migration is the chosen approach to some of these issues, there also needs to be Preservation Policies on what triggers a migration and what types of migration are expected to solve the preservation risk identified. The preservation strategy will describe the range of activities that need to be done in case of a migration.

4.3.2 The repository shall have mechanisms in place for monitoring its preservation environment.

Supporting Text

This is necessary so that the repository can react to changes and thereby ensure that the preserved information remains understandable and usable by the Designated Community.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Surveys of the Designated Community of the repository.

Discussion

The repository should show that it has some active mechanism to ensure that the preserved information remains understandable and usable by the Designated Community and that it has mechanisms in place for monitoring and notification when Representation Information (including formats) approaches obsolescence or is no longer viable. For most repositories, the concern will be with the Representation Information used to preserve information, which may include information on how to deal with a file format or software that can be used to render or process it. Sometimes the format needs to change because the repository can no longer deal with it. Sometimes the format is retained and the information about what software is needed to process it needs to change. If the mechanism depends on an external registry, the repository must demonstrate how it uses the information from that registry.

4.3.2.1 The repository shall have mechanisms in place for monitoring and notification when Representation Information is inadequate for the Designated Community to understand the data holdings.

Supporting Text

This is necessary in order to ensure that the preserved information remains understandable and usable by the Designated Community.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Subscription to a Representation Information registry service; subscription to a technology watch service, surveys amongst its Designated Community members, relevant working processes to deal with this information.

Discussion

The repository must show that it has some active mechanism to warn of impending obsolescence. Obsolescence is determined largely in terms of the knowledge base of the Designated Community.

4.3.3 The repository shall have mechanisms to change its preservation plans as a result of its monitoring activities.

Supporting Text

This is necessary in order for the repository to be prepared for changes in the external environment that may make its current preservation plans a bad choice as the time to implement draws near.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Preservation Plans tied to formal or informal technology watch(es); preservation planning or processes that are timed to shorter intervals (e.g., not more than five years); proof of frequent Preservation Policies and Preservation Plans updates; sections of Preservation Policies that address how plans may be updated and that address how often the plans are required to be reviewed and reaffirmed or updated.

Discussion

The repository should demonstrate or describe how it reacts to information from monitoring, which sometimes requires a repository to change how it deals with the material it holds in ways that could not have been anticipated at an earlier stage. The repository should periodically review its preservation plans and the technology environment and, if necessary, makes changes to those plans to ensure their continued effectiveness. Another possible response to information gathered by monitoring is for the repository to update and create additional Representation Information and/or PDI.

4.3.3.1 The repository shall have mechanisms for creating, identifying or gathering any extra Representation Information required.

Supporting Text

This is necessary in order to ensure that the preserved information remains understandable and usable by the Designated Community.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Subscription to a format registry service; subscription to a technology watch service; preservation plans.

Discussion

The repository should have mechanisms in place for monitoring and notification when Representation Information (including formats) approaches obsolescence or is no longer viable, and it should be able to show that it has mechanisms to address such notifications.

4.3.4 The repository shall provide evidence of the effectiveness of its preservation activities.

Supporting Text

This is necessary in order to assure the Designated Community that the repository will be able to make the information available and usable over the mid-to-long-term.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Collection of appropriate preservation metadata; proof of usability of randomly selected digital objects held within the system; demonstrable track record for retaining usable digital objects over time; Designated Community polls.

Discussion

The repository should be able to demonstrate the continued preservation, including understandability, of its holdings. This could be evaluated at a number of degrees and depends on the specificity of the Designated Community. If a Designated Community is fairly broad, an auditor could represent the test subject in the evaluation. More specific Designated Communities could require significant efforts.

4.4 AIP PRESERVATION

4.4.1 The repository shall have specifications for how the AIPs are stored down to the bit level.

Supporting Text

This is necessary in order to ensure that the information can be extracted from the AIP over the long-term.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Documentation of the format of AIPs; EAST and DEDSL descriptions of the data components (see references [B6] and [B7]).

Discussion

The repository should specify the Representation information down to the bit level of each AIP component and must specify how the separate components are packaged together. The Representation Information must be available for each AIP and must be appropriately linked to the AIP. Often, repositories are tempted to describe AIP content only down to a level where a program will then be used to convert the information to a form understandable to their Designated Communities. However, if those programs ever fail to operate, then the information would be lost in all the AIPs that relied on that program.

4.4.1.1 The repository shall preserve the Content Information of AIPs.

Supporting Text

This is necessary because it is the fundamental mission of a repository to preserve the Content Information for its Designated Communities.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Preservation workflow procedure documentation; workflow procedure documentation; Preservation Policy documents specifying treatment of AIPs and under what circumstances they may ever be deleted; ability to demonstrate the sequence of conversions for an AIP for any particular digital object or group of objects ingested; documentation linking ingested objects and the current AIPs.

Discussion

The repository should be able to demonstrate that the AIPs faithfully reflect the information that was captured during ingest and that any subsequent or future planned transformations will continue to preserve all the required Information Properties of the Content Information. One approach to this requirement assumes that the repository has a policy specifying that AIPs cannot be deleted at any time. This particularly simple and robust implementation preserves links between what was originally ingested, as well as new versions that have been transformed or changed in any way. Depending upon implementation, these newer objects may be completely new AIPs or merely updated AIPs. Either way, persistent links between the ingested object and the resulting AIP should be maintained.

4.4.1.2 The repository shall actively monitor the integrity of AIPs.

Supporting Text

This is necessary in order to protect the integrity of the archival objects over time.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Fixity information (e.g., checksums) for each ingested digital object/AIP; logs of fixity checks; documentation of how AIPs and Fixity information are kept separate; documentation of how AIPs and accession registers are kept separate.

Discussion

A repository should have logs that show actions taken to check the integrity of archival objects in order to assure funders, producers, and users—and to allow them to audit/validate—that the repository is taking the necessary steps to ensure the long-term integrity of the digital objects. The repository should also document that integrity checks are carried out on a regular basis, in order to catch any changes in AIPs as soon as possible so that corrective action can be taken as soon as possible. The repository should allow interested parties to verify that this is the case.

At present, most repositories deal with this at the level of individual information objects by using a checksum of some form, such as MD5. In this case, the repository should be able, and may want to demonstrate that, the Fixity Information (checksums, and the information that ties them to AIPs) are stored separately or protected separately from the AIPs themselves, so that accidental alteration of the AIP would not also damage the Fixity Information. Also, someone who can maliciously alter an AIP would not likely be able as easily to alter the Fixity Information as well.

4.4.2 The repository shall have contemporaneous records of actions and administration processes that are relevant to storage and preservation of the AIPs.

Supporting Text

This is necessary in order to ensure documentation is not omitted or erroneous or of questionable authenticity.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Written documentation of decisions and/or action taken; preservation metadata logged, stored, and linked to pertinent digital objects.

Discussion

The records may be automated or may be written by individuals, depending on the nature of the actions described. Where community or international standards are used, the repository must demonstrate that all relevant actions are appropriately performed.

4.4.2.1 The repository shall have procedures for all actions taken on AIPs.

Supporting Text

This is necessary in order to ensure that any actions performed against an AIP do not alter the AIP information in a manner unacceptable to its Designated Communities.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Written documentation describing all actions that can be performed against an AIP.

Discussion

This documentation is normally created during design of the repository. It should detail the normal handling of AIPs, all actions that can be performed against the AIPs, including success and failure conditions and details of how these processes can be monitored.

4.4.2.2 The repository shall be able to demonstrate that any actions taken on AIPs were compliant with the specification of those actions.

Supporting Text

This is necessary in order to ensure that any actions performed against an AIP do not alter the AIP information in a manner unacceptable to its Designated Communities.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Preservation metadata logged, stored, and linked to pertinent digital objects and documentation of that action; procedural audits of the repository showing that all actions conform to the documented processes.

Discussion

Successful preservation of information in the archive is strongly linked to following established and documented procedures to complete any actions that affect the repository data. The more often ‘special handling’ of repository data occurs and the more often this ‘special handling’ is not overseen in a consistent manner, the more likely that the data held by the repository will be compromised. When procedures are regularly followed, any deviation from procedures that would be likely to cause an alteration in the data will more likely be noticed or, if not noticed, may more likely be able to be corrected, or the timing and likely change could be identified in the future.

4.5 INFORMATION MANAGEMENT

4.5.1 The repository shall specify minimum information requirements to enable the Designated Community to discover and identify material of interest.

Supporting Text

This is necessary in order to enable discovery of the repository's holdings.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Retrieval and descriptive information, discovery metadata, such as Dublin Core, and other documentation describing the object.

Discussion

The repository should be able to deal with the types of requests that will come from a typical user from the Designated Community. A repository does not necessarily have to satisfy every possible request. Retrieval metadata is distinct from descriptive information that describes what has been found.

4.5.2 The repository shall capture or create minimum descriptive information and ensure that it is associated with the AIP.

Supporting Text

This is required in order to ensure that descriptive information is associated with the AIP.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Descriptive metadata; internal or external persistent, unique identifier or locator that is associated with the AIP (see also 4.2.4 about persistent, unique identifier); system documentation and technical architecture; depositor agreements; metadata policy documentation, incorporating details of metadata requirements and a statement describing where responsibility for its procurement falls; process workflow documentation.

Discussion

The repository should show that it associates with each AIP, minimum descriptive information that was either received from the producer or created by the repository. Associating the descriptive information with the object is important, although it does not require one-to-one correspondence, and may not necessarily be stored with the AIP. Hierarchical schemes of description can allow some descriptive elements to be associated with many items.

4.5.3 The repository shall maintain bi-directional linkage between each AIP and its descriptive information.

Supporting Text

This is necessary to ensure that all AIPs can be located and retrieved.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Descriptive metadata; unique, persistent identifier or locator associated with the AIP; documented relationship between the AIP and its metadata; system documentation and technical architecture; process workflow documentation.

Discussion

Repositories must implement procedures to establish and maintain relationships to associate descriptive information for each AIP, and should ensure that every AIP has some descriptive information associated with it and that all descriptive information must point to at least one AIP.

4.5.3.1 The repository shall maintain the associations between its AIPs and their descriptive information over time.

Supporting Text

This is necessary to ensure that all AIPs can continue to be located and retrieved.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Log detailing ongoing maintenance or checking of the integrity of the data and its relationships to the associated descriptive information, especially following repair or modification of the AIP; legacy descriptive information; persistence of identifier or locator; documented relationship between AIP and its descriptive information; system documentation and technical architecture; process workflow documentation.

Discussion

Repositories must implement procedures that let them know when the relationship between the data and the associated descriptive information is temporarily broken to ensure that it can be restored.

4.6 ACCESS MANAGEMENT

The term ‘access’ has a number of different senses, including access by users to the repository system, for example, physical security and user authentication, and the different stages of accessing records (making a request, verifying the rights of the requester, and

preparing and sending a Dissemination Information Package [DIP]). This subsection is concerned with all of these. It is divided into two main requirements, one concerned with the existence and implementation of access policies, and one with the capacity of the repository to provide demonstrably authentic objects as DIPs. Thus the first requirement relates to requests initiated by a user and how the repository handles them to ensure that rights and agreements are respected, that security is monitored, that requests are fulfilled, etc. The second requirement relates to what is delivered to the Consumer and the trust that can be placed in it.

It must be understood that the capabilities and sophistication of the access system will vary depending on the repository's Designated Community and the access mandates of the repository. Because of the variety of repositories and access mandates, these criteria may be subject to questions about applicability and interpretation at a local level.

4.6.1 The repository shall comply with Access Policies.

Supporting Text

This is necessary in order to ensure the repository has fully addressed all aspects of usage which might affect the trustworthiness of the repository, particularly with reference to support of the user community.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Statements of policies that are available to the user communities; information about user capabilities (authentication matrices); logs and audit trails of access requests; explicit tests of some types of access.

Discussion

Depending on the nature of the repository, the Access Policies may cover:

- statements of what is accessible to which community, and on what conditions;
- requirements for authentication and authorization of accessors;
- enforcement of agreements applicable to access conditions;
- recording of access actions.

Access may be managed partly by computers and partly by humans, checking passports, for instance, before issuing a user ID and password may be an appropriate part of access management for some institutions.

4.6.1.1 The repository shall log and review all access management failures and anomalies.

Supporting Text

This is necessary in order to identify security threats and access management system failures.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Access logs, capability of the system to use automated analysis/monitoring tools and generate problem/error messages; notes of reviews undertaken or action taken as a result of reviews.

Discussion

A repository should have some automated mechanism to note anomalous or unusual denials and use them to identify either security threats or failures in the access management system, such as valid users' being denied access. This does not mean looking at every denied access.

4.6.2 The repository shall follow policies and procedures that enable the dissemination of digital objects that are traceable to the originals, with evidence supporting their authenticity.

Supporting Text

This is necessary to establish an auditable chain of authenticity from the AIP to disseminated digital objects.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

System design documents; work instructions (if DIPs involve manual processing); process walkthroughs; production of a sample copy with evidence of authenticity; documentation of community requirements for evidence of authenticity.

Discussion

Authenticity is not an 'all or nothing' concept, but is a matter of degree, judged on the basis of evidence. Thus the adequacy of the evidence is of key importance in assessing this requirement.

This requirement ensures that ingest, preservation, and transformation actions do not lose information that would support an auditable trail of authenticity between the original deposited object and the eventual disseminated object.

A repository should record the processes to construct the DIPs from the relevant AIPs. This is a key part of establishing that DIPs reflect the content of AIPs, and hence of original material, in a trustworthy and consistent fashion. DIPs may simply be a copy of AIPs, or may

result from a simple format transformation of an AIP. But in other cases, they may be derived in complex ways. A user may request a DIP consisting of the title pages from all e-books published in a given period, for instance, which will require these to be extracted from many different AIPs. Or a repository may disseminate automatically generated transcripts of voice recordings. A repository that allows requests for such complex DIPs will need to put more effort into demonstrating how it meets this requirement than a repository that only allows requests for DIPs that correspond to an entire AIP.

This requirement is concerned only with the relation between DIPs and the AIPs from which they are derived; elsewhere the link between the originals SIPs and the AIPs is considered.

4.6.2.1 The repository shall record and act upon problem reports about errors in data or responses from users.

Supporting Text

This is necessary in order for the users to consider the repository to be a trustworthy source of information.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

System design documents; work instructions (if DIPs involve manual processing); process walkthroughs; logs of orders and DIP production; documentation of error reports and the actions taken.

Discussion

The objective of access management is to ensure that a user receives a usable and correct version of the digital object(s) (i.e., DIP) that he or she requested. A repository should show that any problems that do occur and are brought to its attention are investigated and acted on. Such responsiveness is essential for the repository to be considered trustworthy.

5 INFRASTRUCTURE AND SECURITY RISK MANAGEMENT

5.1 TECHNICAL INFRASTRUCTURE RISK MANAGEMENT

5.1.1 The repository shall identify and manage the risks to its preservation operations and goals associated with system infrastructure.

Supporting Text

This is necessary to ensure a secure and trustworthy infrastructure.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Infrastructure inventory of system components; periodic technology assessments; estimates of system component lifetime; export of authentic records to an independent system; use of strongly community supported software e.g., Apache, iRODS, Fedora); re-creation of archives from backups.

Discussion

The repository should conduct or contract assessments of the risks related to hardware and software infrastructure, and operational procedures. The repository should provide mechanisms that minimize risk from dependencies on proprietary or obsolete system infrastructure and from operational error. The degree of support required relates to the criticality of the subsystem(s) involved in long-term preservation. The repository should maintain a system that is scalable (e.g., able to handle anticipated future volumes of both bytes and files) without a major disruption of the system. The repository should maintain a system that is evolvable. That is, the system should be designed in such a way that major components of the system can be replaced with newer technologies without major disruption of the system as a whole. The repository system should be extensible. That is, the system should be designed to accommodate future formats (media and files) without major disruption of the system as a whole. The repository should be able to export its holdings to a future custodian. The repository should be able to re-create the archives after an operational error that overwrites or deletes digital holdings.

5.1.1.1 The repository shall employ technology watches or other technology monitoring notification systems.

Supporting Text

This is necessary to track when hardware or software components will become obsolete and migration is needed to new infrastructure.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Management of periodic technology assessment reports. Comparison of existing technology to each new assessment.

Discussion

The objective is to understand when any subsystem poses a risk of obsolescence, and enable planning migration to new technology before interoperability mechanisms are no longer available. This can be driven by proprietary software dependencies (the vendor no longer supports the subsystem component), and by emergence of new protocols (the mechanism for accessing the system has become obsolete and is no longer supported).

5.1.1.1.1 The repository shall have hardware technologies appropriate to the services it provides to its designated communities.

Supporting Text

This is necessary to provide expected, contracted, secure, and persistent levels of service including: ease of ingest and dissemination through appropriate depositor and user interfaces and technologies such as upload mechanisms; on-going digital object management; preservation approaches and solutions, such as migration; and system security.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Maintenance of up-to-date Designated Community technology, expectations, and use profiles; provision of bandwidth adequate to support ingest and use demands; systematic elicitation of feedback regarding hardware and service adequacy; maintenance of a current hardware inventory.

Discussion

The repository should be aware of the types of storage, file management, preservation and access services expected by its Designated Community, including where applicable, the types of media to be delivered, and needs to make sure its hardware capabilities can support these services. The objective is to track when changes in service requirements by the designated communities require a corresponding change in the hardware technology, when changes in ingestion policies require expanded capabilities, and when changes in preservation policies require new preservation capabilities. This can be driven by changes in capacity requirements (the time needed to read all media is longer than the media lifetime), by changes in delivery mechanisms (new clients for displaying authentic records), and changes in the number and size of archived records.

5.1.1.1.2 The repository shall have procedures in place to monitor and receive notifications when hardware technology changes are needed.

Supporting Text

This is necessary to ensure expected, contracted, secure, and persistent levels of service.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Audits of capacity versus actual usage; audits of observed error rates; audits of performance bottlenecks that limit ability to meet user community access requirements; documentation of technology watch assessments; documentation of technology updates from vendors.

Discussion

The repository should conduct or contract frequent environmental scans regarding hardware status, sources of failure, and interoperability among hardware components. The repository should also be in contact with its hardware vendors regarding technology updates, points of likely failure, and how new components may affect system integration and performance. The objective is to track when changes in service requirements by the designated communities require a corresponding change in the hardware technology, when changes in ingestion policies require expanded capabilities, and when changes in preservation policies require new preservation capabilities. This can be driven by changes in capacity requirements (the time needed to read all media is longer than the media lifetime), by changes in delivery mechanisms (new clients for displaying authentic records), and changes in the number and size of archived records.

5.1.1.1.3 The repository shall have procedures in place to evaluate when changes are needed to current hardware.

Supporting Text

This is necessary to ensure that the repository has the capacity to make informed and timely decisions when information indicates the need for new hardware.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Evaluation procedures in place; documented staff expertise in each technology subsystem.

Discussion

Given information from technology watches or other technology monitoring notification systems, the repository should have procedures and expertise to evaluate this data and make sound decisions regarding the need for new hardware. The objective is to track when technology providers have developed subsystems that minimize risk, or that minimize cost, or that improve performance. This is necessary to track emerging technologies and plan for upgrades before capacity limits occur. The evaluation should identify when the risk of using new technology outweighs the expected benefit, and when the new technology is sufficiently mature to minimize risk.

5.1.1.1.4 The repository shall have procedures, commitment and funding to replace hardware when evaluation indicates the need to do so.

Supporting Text

This is necessary to ensure hardware replacement in a timely fashion so as to avert system failure or performance inadequacy. Without such a commitment, and more importantly, without escrowed financial resources or a secure funding stream, technology watches and notifications are of little value. The repository must have mechanisms for evaluating the efficacy of the new systems before implementation in the production system.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Statement of commitment to provide expected and contracted levels of service; evidence of ongoing financial assets set aside for hardware procurement; demonstration of cost savings through amortized cost of new system.

Discussion

The objective is to demonstrate that the repository has the ability to incorporate new technology, both financially through funding commitments or cost reduction, and operationally through verification of the capabilities of the new systems.

5.1.1.1.5 The repository shall have software technologies appropriate to the services it provides to its designated communities.

Supporting Text

This is necessary to provide expected, contracted, secure, and persistent levels of service including: ease of ingest and dissemination through appropriate depositor and user interfaces and technologies such as upload mechanisms; on-going digital object management; preservation approaches and solutions, such as migration; and system security.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Maintenance of up-to-date Designated Community technology, expectations, and use profiles; provision of software systems adequate to support ingest and use demands; systematic elicitation of feedback regarding software and service adequacy; maintenance of a current software inventory.

Discussion

The objective is to track when changes in service requirements by the designated communities require a corresponding change in the software components, when changes in ingestion policies require support for new data formats and when changes in software technology require new format migration capabilities. This can be driven by changes in access requirements (new clients that require new data formats become preferred), by

changes in delivery mechanisms (new data transfer mechanisms), and changes in the number and size of archived records that require more scalable software.

5.1.1.1.6 The repository shall have procedures in place to monitor and receive notifications when software changes are needed.

Supporting Text

This is necessary to ensure expected, contracted, secure, and persistent levels of service.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Audits of capacity versus actual usage; audits of observed error rates; audits of performance bottlenecks that limit ability to meet user community access requirements; documentation of technology watch assessments; documentation of software updates from vendors.

Discussion

The objective is to track when changes in service requirements by the designated communities require a corresponding change in the hardware technology, when changes in ingestion policies require expanded capabilities, and when changes in preservation policies require new preservation capabilities. This can be driven by security updates (vendor supplied corrections to newly identified vulnerabilities), by changes in delivery mechanisms (new software clients for displaying authentic records), and changes in the number and size of archived records (expanded database requirements). The repository should conduct or contract frequent environmental scans regarding software evolution, likely points of failure, and interoperability among the software and hardware components. The repository should also be in contact with its software vendors regarding technology updates, points of likely failure, and how new programs may affect system integration and performance.

5.1.1.1.7 The repository shall have procedures in place to evaluate when changes are needed to current software.

Supporting Text

This is necessary to ensure that the repository has the capacity to make informed and timely decisions when information indicates the need for new software.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Evaluation procedures in place; documented staff expertise in each software technology subsystem.

Discussion

Given information from technology watches or other technology monitoring notification systems, the repository should have procedures and expertise to evaluate this data and make sound decisions regarding the need for new software. The objective is to track when technology providers have developed software infrastructure that minimizes risk, or that minimizes cost, or that improves performance. This is necessary to track emerging technologies, and plan for upgrades before capacity limits occur. The evaluation should identify when the risk of using new technology outweighs the expected benefit, and when the new technology is sufficiently mature to minimize risk.

5.1.1.1.8 The repository shall have procedures, commitment and funding to replace software when evaluation indicates the need to do so.

Supporting Text

This is necessary to ensure software replacement in a timely fashion so as to avert system failure or performance inadequacy. Without such a commitment, and more importantly, without escrowed financial resources or a secure funding stream, technology watches and notifications are of little value. The repository must have mechanisms for evaluating the efficacy of the new systems before implementation in the production system.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Statement of commitment to provide expected and contracted levels of service; evidence of ongoing financial assets set aside for software procurement; demonstration of cost savings through amortized cost of new system.

Discussion

The objective is to demonstrate that the repository has the ability to incorporate new technology, both financially through funding commitments or cost reduction, and operationally through verification of the capabilities of the new systems.

5.1.1.2 The repository shall have adequate hardware and software support for backup functionality sufficient for preserving the repository content and tracking repository functions.

Supporting Text

This is necessary in order to ensure continued access to and tracking of preservation functions applied to the digital objects in their custody.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Documentation of what is being backed up and how often; audit log/inventory of backups; validation of completed backups; disaster recovery plan, policy and documentation; fire drills; testing of backups; support contracts for hardware and software for backup mechanisms; demonstrated preservation of system metadata such as access controls, location of replicas, audit trails, checksum values.

Discussion

The repository should be able to demonstrate the adequacy of the processes, hardware and software for its backup systems and the full range of ingest, preservation, and dissemination functions required of a repository entrusted with long-term preservation. Simple backup mechanisms must preserve not only the repository main content, but also the system metadata generated by the preservation functions. Repositories need to develop backup plans that ensure their continuity of operations across all failure modes.

5.1.1.3 The repository shall have effective mechanisms to detect bit corruption or loss.

Supporting Text

This is necessary in order to ensure that AIPs and metadata are uncorrupted or any data losses are detected and fall within the tolerances established by repository policy (see 3.3.5).

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Documents that specify bit error detection and correction mechanisms used; risk analysis; error reports; threat analysis; periodic analysis of the integrity of repository holdings.

Discussion

The objective is a comprehensive treatment of the sources of data loss and their real-world complexity. Any data or metadata that is (temporarily) lost should be recoverable from backups. Routine systematic failures must not be allowed to accumulate and cause data loss beyond the tolerances established by the repository policies. Mechanisms such as checksums (MD5 signatures) or digital signatures should be recognized for their effectiveness in detecting bit loss and incorporated into the overall approach of the repository for validating integrity.

5.1.1.3.1 The repository shall record and report to its administration all incidents of data corruption or loss, and steps shall be taken to repair/replace corrupt or lost data.

Supporting Text

This is necessary in order to ensure the repository administration is being kept informed of incidents and recovery actions, and to enable identification of sources of data corruption or loss.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Procedures related to reporting incidents to administrators; preservation metadata (e.g., PDI) records; comparison of error logs to reports to administration; escalation procedures related to data loss; tracking of sources of incidents; remediation actions taken to remove sources of incidents.

Discussion

Having effective mechanisms to detect bit corruption and loss within a repository system is critical but it is only the initial step of a larger process. In addition to recording, reporting and repairing as soon as possible all violations of data integrity, these incidents and the recovery actions and their results must be reported to administrators and made available to all relevant staff. Given identification of the sources of data loss, an assessment of revisions to software and hardware systems, or operational procedures, or management policies is needed to minimize future risk of data loss.

5.1.1.4 The repository shall have a process to record and react to the availability of new security updates based on a risk-benefit assessment.

Supporting Text

This is necessary in order to protect the integrity of the archival objects from unauthorized changes or deletions.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Risk register (list of all patches available and risk documentation analysis); evidence of update processes (e.g., server update manager daemon); documentation related to the update installations.

Discussion

Decisions to apply security updates are likely to be the outcome of a risk-benefit assessment; security patches are frequently responsible for upsetting alternative aspects of system functionality or performance. It may not be necessary for a repository to implement all software patches, and the application of any must be carefully considered. Each security update implemented by the repository must be documented with details about how it is

completed; both automated and manual updates are acceptable. Significant security updates might pertain to software other than core operating systems, such as database applications and Web servers, and these should also be documented. Security updates are not limited to software security updates. Updates to actual hardware or to the hardware system's firmware are included. Over time it is likely that security updates will also be needed for the repository processes and for its physical security. Although security updates can be considered as a part of the change control, they are identified separately here because there are often outside services that compile and circulate information on security issues and updates. At a minimum, repositories should be monitoring these services to ensure that repository-held data is not subject to compromise by identified threats.

5.1.1.5 The repository shall have defined processes for storage media and/or hardware change (e.g., refreshing, migration).

Supporting Text

This is necessary in order to ensure that data is not lost when either the media fail or the supporting hardware can no longer be used to access the data.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Documentation of migration processes; policies related to hardware support, maintenance, and replacement; documentation of hardware manufacturer's expected support life cycles; policies related to migration of records to alternate hardware systems.

Discussion

The repository should have estimates of the access speed and the quantity of information for each type of storage media. Then with estimates of the reliable lifetime of the storage media and information of system loading, etc., the repository can estimate the time required for storage media migration, or refreshing or copying between media without reformatting the bit stream. The repository can then set triggers for initiating the action at an appropriate time so the actions will be completed before data is lost. Copying large quantities of data can take a long time and can affect other system performance metrics. Repositories should also consider the obsolescence of any and all hardware components within the repository system as potential trigger events for migration. Increasingly, long-term, appropriate support for system hardware components is difficult to obtain, exposing repositories to risks and liabilities should they chose to continue to operate the hardware beyond the manufacturer or third-party support warranties. Repositories will likely need to perform media migration off of some types of media onto better supported media based on the estimated lifetime of hardware support rather than on the longer life expected from the media. It is important that the process include a check that the copying has happened correctly.

5.1.1.6 The repository shall have identified and documented critical processes that affect its ability to comply with its mandatory responsibilities.

Supporting Text

This is necessary in order to ensure that the critical processes can be monitored to ensure that they continue to meet the mandatory responsibilities and to ensure that any changes to those processes are examined and tested.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Traceability matrix between processes and mandatory requirements.

Discussion

Examples of critical processes include data management, access, archival storage, ingest, and security processes. Traceability makes it possible to understand which repository processes are required to meet each of the mandatory responsibilities.

5.1.1.6.1 The repository shall have a documented change management process that identifies changes to critical processes that potentially affect the repository's ability to comply with its mandatory responsibilities.

Supporting Text

This is necessary in order to ensure that the repository can specify not only the current processes, but the prior processes that were applied to the repository holdings.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Documentation of change management process; assessment of risk associated with a process change; analysis of the expected impact of a process change; comparison of logs of actual changes to processes versus associated analyses of their impact and criticality.

Discussion

Examples of this would include changes to processes for data management, access, archival storage, ingest, and security. The really important thing is to be able to know what changes were made and when they were made. Traceability makes it possible to understand what was affected by particular changes to the systems. If unintended consequences are later discovered, then having this record may make it possible to reverse the changes or at least to document the changes that were introduced. Change management is a component of the broader topic of configuration management described by ISO 10007:2003 which includes configuration management planning, configuration identification, change control, configuration status accounting and configuration audit. Configuration Management efforts should result in a complete audit trail of decisions and design modifications.

5.1.1.6.2 The repository shall have a process for testing and evaluating the effect of changes to the repository's critical processes.

Supporting Text

This is necessary in order to protect the integrity of the repository's critical processes such that they continue in their ability to meet the repository's mandatory requirements.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Documented testing procedures; documentation of results from prior tests and proof of changes made as a result of tests; analysis of the impact of a process change.

Discussion

Changes to critical systems should be, where possible, pre-tested separately, the expected behaviors documented, and roll-back procedures prepared. After changes, the systems should be monitored for unexpected and unacceptable behavior. If such behavior is discovered the changes and their consequences should be reversed. Whole-system testing or unit testing can address this requirement; complex safety-type tests are not required. Testing can be very expensive, but there should be some recognition of the fact that a completely open regime where no changes are ever evaluated or tested will have problems.

5.1.2 The repository shall manage the number and location of copies of all digital objects.

Supporting Text

This is necessary in order to assert that the repository is providing an authentic copy of a particular digital object.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Random retrieval tests; validation of object existence for each registered location; validation of a registered location for each object on storage systems; provenance and fixity checking information; location register/log of digital objects compared to the expected number and location of copies of particular objects.

Discussion

A repository can have different preservation policies for different classes of objects, depending on factors such as the producer, the information type, or its value. Repositories may require a different number of copies for each class, or manage versions needed to meet access requirements. There may be additional identification requirements if the data integrity mechanisms use alternative copies to replace failed copies. The location of each digital object must be described such that the object can be located precisely, without ambiguity. The location can be an absolute physical location or a logical location within a storage media

or a storage subsystem. Provenance information about copying and moving the data must be maintained/updated, including the identification of those responsible. This is necessary in order to track chain of custody and assert that the repository is providing an authentic copy of a particular digital object. The repository must be able to distinguish between versions of objects or copies and identical copies. This is necessary in order that a repository can assert that it is providing an authentic copy of the correct version of an object.

5.1.2.1 The repository shall have mechanisms in place to ensure any/multiple copies of digital objects are synchronized.

Supporting Text

This is necessary in order to ensure that multiple copies of a digital object remain identical, within a time established as acceptable by the repository, and that a copy can be used to replace a corrupted copy of the object.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Synchronization workflows; system analysis of how long it takes for copies to synchronize; procedures/documentation of synchronization processes.

Discussion

The disaster recovery plan should address what to do should a disaster and an update coincide. For example, if one copy of an object is altered and a disaster occurs while the second is being updated, there needs to be a mechanism to assure that the copy will be updated at the first available opportunity. The mechanisms to synchronize copies of digital objects should be able to detect bit corruption and validate fixity checks before synchronization is attempted.

5.2 SECURITY RISK MANAGEMENT

5.2.1 The repository shall maintain a systematic analysis of security risk factors associated with data, systems, personnel, and physical plant.

Supporting Text

This is necessary to ensure ongoing and uninterrupted service to the Designated Community.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Repository employs the codes of practice found in the ISO 27000 series of standards system control list; risk, threat, or control analysis.

Discussion

The repository should conduct regular risk assessments and maintain adequate security protection in order to provide expected and contracted levels of service, following codes of practice such as ISO 27000.

'System' here refers to more than IT systems, such as hardware, software, communications equipment and facilities, and firewalls. Fire protection and flood detection systems are also significant, as are means to assess personnel, management and administration procedures, resources, as well as operations and service delivery. Loss of income, budget and reputation are significant threats to overall operations as is loss of mandate. On-going internal and external evaluation should be conducted to assess quality of service and relevance to user community served and periodic financial audits should be secured to ascertain ethical and legal practice and maintenance of required operating funds. Intellectual property rights practices should also be reviewed regularly as well as the repository's liability for regulatory non-compliance as applicable. The repository should assess its staff's skills against those required in the evolving digital repository environment and ensure acquisition of new staff or retraining of existing staff as necessary. Regular risk assessment should also address external threats and denial of service attacks and loss of or unacceptable quality of third party services. The repository may conduct overall risk assessments with tools such as DRAMBORA.[2]

5.2.2 The repository shall have implemented controls to adequately address each of the defined security risks.

Supporting Text

This is necessary in order to ensure that controls are in place to meet the security needs of the repository.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Repository employs the codes of practice found in the ISO 27000 series of standards; system control list; risk, threat, or control analyses; and addition of controls based on ongoing risk detection and assessment. Repository maintains ISO 17799 certification.

Discussion

The repository should show how it has dealt with its security requirements. If some types of material are more likely to be attacked, the repository will need to provide more protection, for instance. Repositories that have experienced incidents could record such instances, including the times when systems or content were affected and describe procedures that have been put in place to prevent similar occurrences in the future. Repositories may also conduct

[2] See http://www.repositoryaudit.eu/.

a variety of disaster drills that may involve their parent organization or the community at large. Contingency plans are especially important and need to be tested, updated, and revised on a regular basis.

5.2.3 The repository staff shall have delineated roles, responsibilities, and authorizations related to implementing changes within the system.

Supporting Text

This is necessary in order to ensure that individuals have the authority to implement changes, that adequate resources have been assigned for the effort, and that the responsible individuals will be accountable for implementing such changes.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Repository employs the codes of practice found in the ISO 27000 series of standards; organizational chart; system authorization documentation. Repository maintains ISO 17799 certification.

Discussion

Authorizations are about who can do what: who can add users, who has access to change metadata, who can access audit logs. It is important that authorizations are justified, that staff understand what they are authorized to do, that staff have required skills associated with various roles and authorizations, and that there is a consistent view of this across the organization.

5.2.4 The repository shall have suitable written disaster preparedness and recovery plan(s), including at least one off-site backup of all preserved information together with an offsite copy of the recovery plan(s).

Supporting Text

This is necessary in order to ensure that sufficient backup and recovery capabilities are in place to facilitate continuing preservation of and access to systems and their content with limited disruption of services.

Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement

Repository employs the codes of practice found in the ISO 27000 series of standards; disaster and recovery plans; information about and proof of at least one off-site copy of preserved information; service continuity plan; documentation linking roles with activities; local geological, geographical, or meteorological data or threat assessments. Repository maintains ISO 17799 certification.

Discussion

The level of detail in a disaster plan, and the specific risks addressed need to be appropriate to the repository's location and service expectations. Fire is an almost universal concern, but earthquakes may not require specific planning at all locations. The disaster plan must, however, deal with unspecified situations that would have specific consequences, such as lack of access to a building or widespread illness among critical staff. In the event of a disaster at the repository, the repository may want to contact local and/or national disaster recovery bodies for assistance. Repositories may also conduct a variety of disaster drills that may involve their parent organization or the community at large.

ANNEX A

SECURITY CONSIDERATIONS

(NORMATIVE)

A1 INTRODUCTION

The use of the Audit and Certification Recommended Practice has several potential areas of security concern.

One security concern is the possibility that the repository is fooled into undergoing an audit by someone unqualified or even malicious.

Another concern involves the possible release of confidential information which is collected as evidence by the auditor.

A2 SECURITY CONCERNS WITH RESPECT TO THE CCSDS DOCUMENT

The repository may ask someone to perform an audit using this Recommended Practice. There is a possibility that the person contacted is not in fact the person that the repository believes him or her to be. Alternatively the correct person may be contacted but in fact another, possibly malicious, person may turn up to perform the audit.

In the process of collecting evidence for the various metrics the auditor may collect information which is confidential or sensitive, for example details of security weaknesses.

There is a danger that such information may fall into the wrong hands and expose the repository to increased risk. Alternatively in the process of collecting evidence the repository system may be damaged.

While these are all valid security concerns, they fall outside the purview of this Recommended Practice, which applies only to the metrics which an auditor should use for auditing a repository.

A3 POTENTIAL THREATS AND ATTACK SCENARIOS

Impersonation of an auditor and/or release of confidential information could both result in exposing the repository and its holdings to increased risk and loss of reputation of the repository.

A4 CONSEQUENCES OF NOT APPLYING SECURITY TO THE TECHNOLOGY

While these security issues are of concern, they are out of scope with respect to this document. This document aims to provide the basis for an audit and certification process for assessing the trustworthiness of digital repositories. Providing protection against false auditors must rely on the repository's identification and authorization systems. Protection against loss of confidential information in the possession of the auditor must be provided by the security system of that auditor and the method of transmission of information which is agreed between the repository and auditor. Protection against damage to the repository or its holdings during an audit must rely on the security and safety systems of the repository.

ANNEX B

REFERENCES

(INFORMATIVE)

[B1] D. Waters and J. Garrett. *Preserving Digital Information*. Report of the Task Force on Archiving of Digital Information. Washington, DC: CLIR, May 1996.

[B2] *Trusted Digital Repositories: Attributes and Responsibilities*. An RLG-OCLC Report. Mountain View, CA: RLG, May 2002.

[B3] *Trustworthy Repositories Audit & Certification: Criteria and Checklist*. Version 1.0. Chicago: CRL, February 2007.

[B4] *Producer-Archive Interface Methodology Abstract Standard*. Recommendation for Space Data System Standards, CCSDS 651.0-B-1. Blue Book. Issue 1. Washington, D.C.: CCSDS, May 2004.

[B5] *XML Formatted Data Unit (XFDU) Structure and Construction Rules*. Recommendation for Space Data System Standards, CCSDS 661.0-B-1. Blue Book. Issue 1. Washington, D.C.: CCSDS, September 2008.

[B6] The Data Description Language EAST Specification (CCSD0010). Recommendation for Space Data System Standards, CCSDS 644.0-B-3. Blue Book. Issue 3. Washington, D.C.: CCSDS, July 2009.

[B7] Data Entity Dictionary Specification Language (DEDSL)—Abstract Syntax (CCSD0011). Recommendation for Space Data System Standards, CCSDS 647.1-B-1. Blue Book. Issue 1. Washington, D.C.: CCSDS, June 2001.

[B8] "Digital Preservation Management: Implementing Short-Term Strategies for Long-Term Problems." Digital Preservation Management Resources. Cornell University.

[B9] *Quality Management Systems—Fundamentals and Vocabulary*. International Standard, ISO 9000:2005. 3rd edition. Geneva: ISO, 2005.

[B10] *Information Technology—Security Techniques—Code of Practice for Information Security Management*. International Standard, ISO/IEC 17799:2005. 2nd edition. Geneva: ISO, 2005.

[B11] *Information and Documentation—Records Management—Part 1: General*. International Standard, ISO 15489-1:2001. Geneva: ISO, 2001.

[B12] *Information and Documentation—Records Management—Part 2: Guidelines*. International Standard, ISO/TR 15489-2:2001. Geneva: ISO, 2001.

[B13] *Trustworthy Information Systems Handbook*. Version 4. Saint Paul, Minnesota: Minnesota Historical Society, July 2002.

[B14] Ron Ross, et al. *Guide for Assessing the Security Controls in Federal Information Systems*. National Institute of Standards and Technology Special Publication 800-53A. Gaithersburg, Maryland: NIST, July 2008.

[B15] Susanne Dobratz, Astrid Schoger, and Stefan Strathmann. "The nestor Catalogue of Criteria for Trusted Digital Repository Evaluation and Certification." *Journal of Digital Information* 8, no. 2 (2007).