preservation repository CRL specifications certification criteria RLG Programs OCLC audit digital object management NARA trustworthy metadata preservation repository CRL specifications certification criteria RLG Programs OCLC audit digital object management NARA trustworthy metadata preservation repository CRL specifications certification criteria RLG Programs OCLC audit digital object management NARA trustworthy metadata preservation repository CRL specifications certification criteria RLG Programs OCLC audit digital object management NARA trustworthy metadata preservation repository CRL specifications certification criteria RLG Programs OCLC audit digital object management NARA trustworthy metadata

Trustworthy Repositories Audit & Certification: Criteria and Checklist

Contents:

OCLC

CRL

NATIONAL ARCHIVES AND RECORDS ADMINISTRATION

LITTERA SCRIPTA MANET

1985

Version 1.0

February 2007

CRL, The Center for Research Libraries
6050 South Kenwood Avenue, Chicago, Illinois, 60637-2804 USA

OCLC Online Computer Library Center, Inc.
6565 Kilgour Place, Dublin, Ohio, 43017-3395 USA

Foreword

In 2003, RLG and the National Archives and Records Administration created a joint task force to specifically address digital repository certification. The goal of the RLG-NARA Task Force on Digital Repository Certification has been to develop criteria to identify digital repositories capable of reliably storing, migrating, and providing access to digital collections. The challenge has been to produce certification criteria and delineate a process for certification applicable to a range of digital repositories and archives, from academic institutional preservation repositories to large data archives and from national libraries to third-party digital archiving services.

In the last two years, additional work and development has been accomplished through working groups in Europe, as well as through funded projects in the United States. This document incorporates and leverages the combined advances to fulfill the goal of enabling repository audit, assessment, and certification. The RLG-NARA task force, combined with critical contributions from the Center for Research Libraries Auditing and Certification of Digital Archives project, the nestor project, and the Digital Curation Centre have set the stage for official audit and certification of digital repositories and archives to move forward.

Though designed as a set of criteria to facilitate the certification of digital repositories, this document and the enclosed checklist have a number of uses outside of the carefully prescriptive world of certified repositories. Envisioned uses of this document include repository planning guidance, planning and development of a certified repository, periodic internal assessment of a repository, analysis of services which hold critical digital content on which institutions rely, and objective third-party evaluation of any repository or archiving service. In any use however, it is important to understand that this comprehensive set of criteria has been created to measure digital repositories that have long-term access and preservation responsibilities for the content they hold. Other uses, such as repository software evaluation, are possible but the checklist would need to be adapted in order to produce an adequate set of criteria for that context. The document provides further information and instructions about uses of the checklist.

As co-chairs of the RLG-NARA Task Force on Digital Repository and Certification, we are grateful for the full support of our organizations – RLG (now RLG Programs, a part of the OCLC Office of Programs and Research) and the National Archives and Records Administration – during the multi-year evolution of these criteria and the principles that underlie them. We are also grateful for the significant contributions of the task force members, as well as others we have called to attention in the Acknowledgments. Finally, we’re thankful that the Center for Research Libraries has agreed to take on related audit and assessment activities, including future versions of the criteria and checklist.

Robin L. Dale and Bruce Ambacher

Acknowledgments

This work was completed with the assistance of the following people and organizations:

The RLG- National Archives and Records Administration Digital Repository Certification Task Force

Bruce Ambacher, National Archives and Records Administration (Co-chair)
Kevin Ashley, University of London Computing Centre
John Berry, Internet Archive
Connie Brooks, Stanford University
Robin L. Dale, RLG (Co-chair)
Dale Flecker, Harvard University
David Giaretta, Rutherford Appleton Laboratory, Council for the Central Laboratory of the Research Councils, UK
Babak Hamidzadeh, Library of Congress
Keith Johnson, Stanford University
Maggie Jones, Digital Preservation Coalition, UK
Nancy McGovern, Cornell University
Andrew McHugh, Digital Curation Centre
Don Sawyer, National Aeronautics and Space Administration
Johan Steenbakkers, Koninklijke Bibliotheek

With special appreciation for funding further development and application of the criteria to:

The Andrew W. Mellon Foundation

The Center for Research Libraries

National Archives and Records Administration, Electronic Records Archives Program

And for their suggestions and contributions to the criteria development, we are grateful to:

Sayeed Choudhury, Johns Hopkins University

Tim DiLauro, Johns Hopkins University

Susanne Dobratz, Humboldt-Universität zu Berlin and the *nestor* project

Bernard Reilly, the Center for Research Libraries

Dr. Seamus Ross, University of Glasgow and the Digital Curation Centre

Dr. Astrid Schoger, Bayerische Staatsbibliothek München and the *nestor* project

The staff of the Inter-university Consortium on Political and Social Research

The staff of the Koninklijke Bibliotheek

The staff of the LOCKSS Program

The staff of Portico

Dr. Kenneth Thibodeau, National Archives and Record Administration (US)

Marie Waltz, the Center for Research Libraries

Table of Contents

<This page left intentionally blank.>

To state the facts frankly is not to despair the future nor indict the past. The prudent heir takes careful inventory of his legacies and gives a faithful accounting to those whom he owes an obligation of trust.
– John F. Kennedy, State of the Union Address, 1961

Introduction

A decade ago, the Task Force on Archiving of Digital Information (1996) 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." The task force stopped short of articulating the details of such a certification process. Certainly one obstacle was that though some archives and repositories existed at the time, there was no organized "digital preservation community" with common, consensus-driven practices, let alone standards. Each archive or repository conducted digital preservation in its own manner and to the level that seemed to address funding and user community needs.

Work in articulating responsible digital archiving infrastructure was furthered by the development of the *Open Archival Information System (OAIS) Reference Model* (CCSDS 2002). 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, but there was no established understanding of "OAIS-compliance" beyond meeting the high-level responsibilities defined by the standard. There were certainly no criteria for measuring compliance.

Claims of trustworthiness are easy to make but are thus far difficult to justify or objectively prove. As Clifford Lynch has stated, "Stewardship is easy and inexpensive to claim; it is expensive and difficult to honor, and perhaps it will prove to be all too easy to later abdicate" (Lynch 2003). Establishing more clear criteria detailing what a trustworthy repository *is* and *is not* has become vital.

In 2002, RLG and OCLC jointly published *Trusted Digital Repositories: Attributes and Responsibilities* (TDR), 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.

Establishing Audit & Certification Criteria

In 2003, RLG and the National Archives and Records Administration created a joint task force to specifically address digital repository certification. The goal of the RLG-NARA Task Force on Digital Repository Certification has been to develop criteria to identify digital repositories capable of reliably storing, migrating, and providing access to digital collections. The challenge has been to produce certification criteria and delineate a process for certification applicable to a range of digital repositories and archives, from academic institutional preservation repositories to large data archives and from national libraries to third-party digital archiving services.

Digital preservation infrastructure continues to grow through institutional funding and national initiatives like the US National Digital Information Infrastructure and Digital Preservation Program (NDIIPP) or the European Union Research Policy and Funding Framework Programmes. Research and development projects continue to address the remaining digital preservation challenges. While these challenges have not yet been resolved, the proliferation of experience, research, and infrastructure throughout the cultural heritage community has made trustworthy digital repositories conceptually realistic. Over the last four years, projects, programs, and collaborative work have begun to cultivate a shared view among stakeholders on well-defined infrastructure and processes for achieving certain digital preservation objectives.

The groundwork has been laid for the establishment of the long-awaited certification process for digital repositories. Over the last three years, the RLG-NARA task force has worked to define and articulate criteria or indicators of trustworthiness and reliability for digital repositories. The process has been iterative and this version of the criteria and documentation is the sixth generation of the expert group's work. It was influenced by many organizations and their work, especially:

- The Center for Research Libraries (CRL) *Auditing and Certification of Digital Archives Project*, an Andrew W. Mellon Foundation funded activity, which leveraged the earlier public draft of this document in a series of test audits against subject archives and fed back invaluable information for the further development of this document.
- The certification working group of Germany's nestor (Network of Expertise in Long-Term Storage of Digital Resources) project and their development of the *Kriterienkatalog vertrauenswürdige digitale Langzeitarchive (Catalogue of Criteria for Trusted Digital Repositories).*
- The Digital Curation Centre (DCC) and their active contributions toward the revision of this checklist, as well as their work on the role of evidence and qualitative measurement in establishing trust in digital repositories.
- The Australian Partnership for Sustainable Repositories project and the National Library of Australia, whose workshop *Long-Term Repositories: Taking the Shock out of the Future* dedicated a full day to trusted repository methodology and, in particular, use of the earlier public draft of the checklist. The feedback was timely, cross-disciplinary, and very valuable.
- And, of course, the individuals and institutions that took their valuable time to provide important, constructively critical comments on the public draft of this document.

This new document, version 1.0 of the *Criteria for Measuring Trustworthiness of Digital Repositories & Archives: an Audit & Certification Checklist*, represents best current practice and thought about the organizational and technical infrastructure required to be considered trustworthy and capable of certification. It establishes a baseline definition of a trustworthy digital repository and lays out the components that must be considered and evaluated as a part of that determination. It discusses the envisioned uses of this document, and the principles underlying the application of the criteria. Finally, it documents criteria that trustworthy repositories will be able to meet, providing explanations and examples.

This new audit tool is the work of many experts representing an international range of communities in research, governments, data archives, and cultural heritage organizations. Task force members were chosen because of their experience building and managing digital repositories. Additionally, more than a year of public draft and discourse at conferences allowed us to gain invaluable insight and contributions from the community trying to understand and utilize the audit checklist. To be of true value, the tools for auditing repositories needed to be developed by practitioners. The task force gathered for this task represents over 160 years of collective experience in information technology and systems and more than 130 years of collective experience in the preservation of digital information (data archives, electronic records, digital repositories, etc.). Untold years of experience were added by the collective contribution of implementers, readers, and institutions active in the digital preservation community.

As Ross and McHugh (2005) state, digital repositories must engender, establish, and maintain trusted status in a variety of ways. In this new environment of digital information and digital responsibilities, all repositories will need to establish their trusted status. Use of this audit checklist—beginning with self-assessment—is one mechanism a repository can use to understand its capabilities, where it stands against potential threats, and any other risks inherent in its systems.

A Trusted Digital Repository

At the very basic level, the definition of a trusted 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” (*TDR* 2002). 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.

The nestor working group says a trusted, “long-term digital repository is a complex and interrelated system” (nestor 2006). However, more than just the “digital preservation system” drives the management of the digital materials. In determining trustworthiness, one must look at the entire system in which the digital information is managed, including the organization running the repository: its governance; organizational structure and staffing; policies and procedures; financial fitness and sustainability; the contracts, licenses, and liabilities under which it must operate; and trusted inheritors of data, as applicable. Additionally, the digital object management practices, technological infrastructure, and data security in place must be reasonable and adequate to fulfill the mission and commitments of the repository.

A trusted digital repository will understand threats to and risks within its systems. As articulated by Rosenthal et al. (2005), these potential threats include media failure, hardware failure, software failure, communication errors, failure of network services, media and hardware obsolescence, software obsolescence, operator error, natural disaster, external attack, internal attack, economic failure, and organizational failure. 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(ies), 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 trusted status is not a one-time accomplishment—achieved and forgotten. To retain trusted status, a repository will need to undertake a regular cycle of audit and/or certification.

Toward an International Audit & Certification Process

The development of a certification process was always envisioned to take place in an international environment and with a unified set of criteria, but with regional implementation, e.g., by country, continent, or geographic region. For the last 18 months, three organizations have worked to establish a unified, international process for certification:

- The Center for Research Libraries (CRL) through the Auditing and Certification of Digital Archives project funded by an Andrew W. Mellon Foundation grant.
- The Digital Curation Centre (DCC), both in concert with the CRL project and through independent test audits in the UK.
- The certification working group of the nestor (Network of Expertise in Long-Term Storage of Digital Resources) project in Germany.

These efforts to merge development of a certification process highlighted small but important differences between the criteria in this audit checklist and the nestor *Criteria Catalogue*, for example. For now, a single, standardized set of criteria and applicable rules have proven impractical for geopolitical reasons.

Representatives from the three efforts will be working together to identify a core set of criteria or requirements—those common to all three checklists—and a common understanding of their applicability. As well, the criteria identified in this document will enter the ISO standardization process through ISO Technical Committee 20, Subcommittee 13 (ISO/TC20/SC13). Full ISO standardization may take several years. In the interim, this checklist incorporates existing standards and best practices for trustworthy repositories and related digital object management and is applicable for audit and certification activities.

Upon the release of this checklist, the CRL will take on the US activities related to audit and certification. In the UK, the DCC will execute plans to be the audit and certification managing agency for UK repositories and archives; and in Germany, the second phase of the nestor project, funded by Germany’s Federal Ministry of Education and Research, will move forward with building the audit and certification program for Germany using their *Criteria Catalogue*. Representatives of other countries who have approached the members of this working group about establishing local efforts in their region are formally encouraged to the CRL, the DCC, and nestor’s certification working group to build a formal collaboration—indeed a kind of virtual agency—for digital repository audit and certification.

Future Versions of the Criteria

Just as digital repositories will need auditing and revisions over the years, so too will this checklist. Review will be planned and will be developed in concert with stakeholder organizations.

Using this Checklist for Audit & Certification

The digital preservation community has come to not only recognize but embrace the fact that not all repositories will be "equal." This diversity has been made abundantly clear by the proliferation of repository types (institutional repositories, open-access repositories, digital repositories, digital preservation repositories, digital archives, etc.) on local, regional, national, and international levels. For many of these repositories, preservation is not the primary purpose or explicit priority. With that understanding, it is easy to comprehend why some repositories may not choose to pursue certification, just as it is easy to see why others should feel compelled (or perhaps be compelled) to pursue certification.

Regardless of size or purpose, all repositories should be encouraged to use this checklist as a tool for objective evaluation whether it is accomplished in-house or by an objective, third-party auditor, and regardless of whether it is accomplished for local information gathering and evaluation or as a part of an international or national certification process. Audit is the basis for comparing local capabilities against a set of core criteria for a trusted digital repository. Certification is a further step that some repositories will and/or must take for formal, objective recognition at the international or network level. The result of any audit must be viewed in the context in which it was undertaken.

The checklist is divided into three sections:

A. Organizational infrastructure

B. Digital object management

C. Technologies, technical infrastructure, and security.

Evidence

The Audit & Certification Criteria (beginning on page 9) explains and discusses each point on the checklist, providing examples of evidence that can show how the repository meets the criteria. *These examples are illustrative rather than prescriptive, and the lists of possible evidence are not exhaustive.*

Intended Audience

This document is meant 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 this checklist during a design or redesign process for their digital repository. Some repositories may seek to be certified against its requirements.

The requirements touch on every level of a repository's functions, so the document will be relevant to staff with many different roles within a repository and the organization of which it is part. The organization's senior management and policy makers will need to be aware of at least the requirements of section A of the audit and certification criteria. Systems and network staff, who may be responsible for many parts of infrastructure other than those specific to the repository, will have an interest in section C, which is also of relevance to those responsible for matters such as building security and fire protection. Whether producers of material or consumers of it, those who deal with external users will find information relevant to their work in sections A and B.

The document is expected to be of interest to a wider community, however. Organizations planning repositories, and repositories that do not expect to seek certification but are, for example, themselves part of a chain of preservation, or are unsure of its relevance to them, are still likely to find much that is of interest here. The analysis of a repository's functions, the itemized requirements, and the explanations of how they can be tested can all help a repository plan and review its working practices. This may reassure the repository that it is operating in accordance with recognized best practice, it may help staff and users understand the repository's actions, and it can help an organization focus limited resources where they will best ensure that digital resources will survive.

Both producers of digital material that will be preserved for long periods and users of this material will find much useful information here to help them understand what to expect from the repositories they deal with. It may help some producers streamline their interactions with the repositories that take long-term responsibility for their materials.

Even in the absence of any formal certification process, this document will help organizations considering outsourcing some or all aspects of digital preservation by showing how they can ensure that the organizations they contract with are carrying out the task of digital preservation in a way that deserves trust.

Applicability of the Criteria

These criteria are written to be applicable to any kind of digital repository or archives. Every attempt has been made to provide illustrative examples, though these examples should not be construed as representing exclusive application. These criteria are applicable to libraries, museums, archives, data archives, science data archives, etc., as well as the wildly heterogeneous data produced and collected by these kinds of organizations.

In applying the criteria, users and auditors must take into account the context of the institution, its mission, priorities, and stated commitments. To provide clear direction, we have gratefully adapted work of the nestor working group and the Digital Curation Centre articulating these principles for applying any criteria: documentation (evidence), transparency, adequacy, and measurability.

- **Documentation (evidence):** The objectives, the design, specifications, and implementation of the digital long-term repository should be appropriately documented, and documentation should be reviewed and updated on a regular schedule. The documentation can be used to evaluate the status of development both internally and externally. It will also be necessary to provide the documentation as evidence during the course of an audit. In particular, appropriate documentation of all steps permits auditors to evaluate the digital long-term repository as a whole. All quality and security standards, their applicability to the repository, and potential certification must also be suitably documented.
- **Transparency:** Ultimately, examining a repository for trustworthiness relies on another critical component: transparency, both internal and external. Only a repository that exposes its design, specifications, practices, policies, and procedures for risk analysis can be trusted. (Note: These transparency requirements are mandatory for internal use and as provision of evidence during an audit, but obviously not all documentation requires full public disclosure.) Because digital preservation must and will be undertaken through international collaboration (both formal and informal), digital repositories must be transparent in all practices as they relate to preservation capabilities or assertions made about trusted, long-term management of digital materials. Such transparency will expose any risks; allow informed decisions to be made by information stakeholders, funders, and users; and foster collaboration to accomplish what can only be a collaborative, international digital preservation effort.
- **Adequacy:** The principle of adequacy takes into account that absolute standards do not exist for

all aspects of repository organizational infrastructure, digital object management, and technologies and technical infrastructure. Even if they did, they would not apply to all types of repositories and archives and all situations. In an audit using these criteria, an evaluation is always based on the objectives and commitments of the digital long-term repository in question. At its most basic level an audit should assess whether a repository can meet its stated commitments—is it doing what is says it is doing?—and the criteria have to be seen within the context of the special archiving tasks of the repository. Some criteria may therefore not apply in certain cases. Depending on the objectives and tasks of a digital long-term repository, the required degree of fulfillment for a particular criterion may differ as well.

- **Measurability:** In principle, the goal is to have objective controls (criteria) against which repositories can be evaluated. In some cases, however—especially regarding long-term aspects—objective controls may not be available; that is, how can one evaluate the effectiveness of an institution's preserving planning if no actions have yet been taken? In such cases, evaluation must note indicators of the degree of trustworthiness. The degree of transparency will make the indicators accessible for evaluation.

Certification for digital repositories will involve far more than the documentation of criteria. To be useful, a full certification process must provide tools to allow for planning and self-examination, as well as an external, objective audit. It must recognize standards and best practices relevant to the community of the repository, as well as those of the information management and security industries as a whole. In other words, audit and certification of trusted digital repositories cannot exist in a vacuum.

First and foremost, certification must take place within geopolitical contexts. It *is* important to have an international standard specifying criteria against which repositories will be evaluated, but audit and certification processes will likely be implemented in various ways to meet national need or comply with national laws, as with other standards related to audit and certification. The key will be to understand any slight variations from one country or continent to another, as well as the context for the audit and certification process.

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 defines 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 14721:2002, 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.

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” (OAIS, 2002). 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 long-term preservation responsibilities and functionality.

This document borrows from the OAIS the 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. To be clear that the criteria do not assume a single designated community, this document often uses the construction “designated community(ies).”

Finally, this document names criteria that, combined, evaluate the trustworthiness of digital repositories and archives. While the correct phrase to describe such entities is “trustworthy digital repositories,” the community has long used “trusted digital repositories” to convey that same assessment. While grammatically incorrect, it is never the less the phrase most familiar to and in use within the community. Therefore, this document does refer to trustworthiness and trusted digital repositories.

Audit & Certification Criteria

A. Organizational Infrastructure

Though adequate technical architecture, processes, and capabilities underpin a trusted digital repository, the technical aspects are only one piece of the overarching infrastructure supporting the digital repository functions. Organizational attributes of digital repositories are equally critical.

Organizational attributes are characteristics of the repository organization that affect performance, accountability, and sustainability. *Trusted Digital Repositories: Attributes and Responsibilities* (2002) grouped these types of attributes into four of its "Attributes of a Trusted Digital Repository": administrative responsibility, organizational viability, financial sustainability, and procedural accountability. In their training workshop Digital Preservation Management, Cornell University Library refers to these characteristics as "organizational infrastructure." According to Cornell, "an organization's infrastructure is best embodied in its policies and procedures," and documentation of organizational infrastructure is embodied in three distinct levels: policy framework, policies and procedures, and plans and strategies (Cornell 2004). Organizational attributes are indicators of a digital repository's comprehensive planning, readiness, ability to address its responsibilities, and trustworthiness.

Organizational infrastructure includes but is not restricted to these elements:

- Governance
- Organizational structure
- Mandate or purpose
- Scope
- Roles and responsibilities
- Policy framework
- Funding system
- Financial issues, including assets
- Contracts, licenses, and liabilities
- Transparency

Criteria addressing these elements are organized in these five groups:

- A1. Governance and organizational viability
- A2. Organizational structure and staffing
- A3. Procedural accountability and policy framework
- A4. Financial sustainability
- A5. Contracts, licenses, and liabilities

A1. Governance & organizational viability

Regardless of the size, scope, or nature of the digital preservation program, a trusted repository must demonstrate an explicit, tangible, and long-term commitment to compliance with prevailing standards, policies, and practices.

A1.1 Repository has a mission statement that reflects a commitment to the long-term retention of, management of, and access to digital information.

The mission statement of the repository must be clearly identified and accessible to depositors and other stakeholders and contain an explicit long-term commitment.

Evidence: Mission statement for the repository; mission statement for the organizational context in which the repository sits; legal or legislative mandate; regulatory requirements.

A1.2 Repository has 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.

Part of the repository's perpetual-care promise is a commitment to identify appropriate successors or arrangements should the need arise. Consideration needs to be given to this responsibility while the repository or data is viable—not when a crisis occurs—to avoid irreparable loss. Organizationally, the data in a repository can be at risk regardless of whether the repository is run by a commercial organization or a government entity (national library or archives). At government-managed repositories and archives, a change in government that significantly alters the funding, mission, collecting scope, or staffing of the institution may put the data at risk. These risks are similar to those faced by commercial and research-based repositories and should minimally be addressed by succession plans for significant collections within the greater repository.

A formal succession plan should include the identification of trusted inheritors, if applicable, and the return of digital objects to depositors with adequate prior notification, etc. If a formal succession plan is not in place, the repository should be able to point to indicators that would form the basis of a plan, e.g., partners, commitment statements, likely heirs. Succession plans need not specify handoff of entire repository to a single organization if this is not feasible. Multiple inheritors are possible so long as the data remains accessible.

Evidence: Succession plan(s); escrow 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; formal documents describing exit strategies and contingency plans; depositor agreements.

A2. Organizational structure & staffing

A repository must have designated staff with requisite skills and training and must provide ongoing development. The repository should be able to document efforts to define and maintain requisite skills, roles, job descriptions, and development plans.

A2.1 Repository has identified and established the duties that it needs to perform and has appointed staff with adequate skills and experience to fulfill these duties.

The repository must identify the competencies and skill sets required to operate the repository over time and demonstrate that the staff and consultants have the range of requisite skills—e.g., archival training, technical skills, and legal expertise.

Evidence: A staffing plan; competency definitions; job description; development plans; plus evidence that the repository review and maintains these documents as requirements evolve.

A2.2 Repository has the appropriate number of staff to support all functions and services.

Staffing for the repository must be adequate for the scope and mission of the archiving program. The repository should be able to demonstrate an effort to determine the appropriate number and level of staff that corresponds to requirements and commitments. (These requirements are related to the core functionality covered by a certification process. Of particular interest to repository certification is whether the organization has appropriate staff to support activities related to the long-term *preservation* of the data.) The accumulated commitments of the repository can be identified in deposit agreements, service contracts, licenses, mission statements, work plans, priorities, goals, and objectives. Understaffing or a mismatch between commitments and staffing indicates that the repository cannot fulfill its agreements and requirements.

Evidence: Organizational charts; definitions of roles and responsibilities; comparison of staffing levels to commitments and estimates of required effort.

A2.3 Repository has an active professional development program in place that provides staff with skills and expertise development opportunities.

Technology will continue to change, so the repository must ensure that its staff's skill sets evolve, ideally through a lifelong learning approach to developing and retaining staff. As the requirements and expectations pertaining to each functional area evolve, the repository must demonstrate that staff are prepared to face new challenges.

Evidence: 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.

A3. Procedural accountability & policy framework

A repository must provide clear and explicit documentation of its requirements, decisions, development, and actions to ensure long-term preservation and access to digital content in its care. This documentation assures consumers, management, producers, and certifiers that the repository is meeting its requirements and fully performing its role as a trusted digital repository. Certification, the clearest indicator of a repository's sound and standards-based practice, is facilitated by procedural accountability that results in comprehensive and current policies, procedures, and practice.

A3.1 Repository has defined its designated community(ies) and associated knowledge base(s) and has publicly accessible definitions and policies in place to dictate how its preservation service requirements will be met.

The definition of the designated community(ies) (producer and user community) is arrived at through the planning processes used to create the repository and define its services. The definition will be drawn from various sources ranging from market research to service-level agreements for producers to the mission or scope of the institution within which the repository is embedded.

Meeting the needs of the designated community—the expected understandability of the information, not just access to it—will affect the digital object management, as well as the technical infrastructure of the overall repository. For appropriate long-term planning, the repository or organization must understand and institute policies to support these needs.

For a given submission of information, the repository must make clear the operational definition of understandability that is associated with the corresponding designated community(ies). The designated community(ies) may vary from one submission to another, as may the definition of understandability that establishes the repository's responsibility in this area. This may range from no responsibility, if bits are only to be preserved, to the maintenance of a particular level of use, if understanding by the members of the designated community(ies) is determined outside the repository, to a responsibility for ensuring a given level of designated community(ies) human understanding, requiring appropriate Representation Information.

The documentation of understandability will typically include a definition of the applications the designated community(ies) will use with the information, possibly after transformation by repository services. For example, if a designated community is defined as readers of English with access to widely available document rendering tools, and if this definition is clearly associated with a given set of Content Information and Preservation Description Information, then the requirement is met.

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
- 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.

Evidence: Mission statement; written definitions of the designated community(ies); documented policies; service-level agreements.

A3.2 Repository has procedures and policies in place, and mechanisms for their review, update, and development as the repository grows and as technology and community practice evolve.

The policies and procedures of the repository must be complete, written or available in a tangible form, remain current, and must evolve to reflect changes in requirements and practice. The repository must demonstrate that a policy and procedure audit and maintenance is in place and regularly applied. Policies and procedures should address core areas, including, for example, transfer requirements, submission, quality control, storage management, disaster planning, metadata management, access, rights management, preservation strategies, staffing, and security. High-level documents should make organizational commitments and intents clear. Lower-level documents should make day-to-day practice and procedure clear. Versions of these documents must be well managed by the repository (e.g., outdated versions are clearly identified or maintained offline) and qualified staff and peers must be involved in reviewing, updating, and extending these documents. The repository should be able to document the results of monitoring for relevant developments; responsiveness to prevailing standards and practice, emerging requirements, and standards that are specific to the domain, if appropriate; and similar developments. The repository should be able to demonstrate that it has defined “comprehensive documentation” for the repository. See Appendix 3: Minimum Required Documents for more information.

Evidence: Written documentation in the form of policies, procedures, protocols, rules, manuals, handbooks, and workflows; specification of review cycle for documentation; documentation detailing review, update, and development mechanisms. If documentation is embedded in system logic, functionality should demonstrate the implementation of policies and procedures.

A3.3 Repository maintains written policies that specify the nature of any legal permissions required to preserve digital content over time, and repository can demonstrate that these permissions have been acquired when needed.

Because the right to change or alter digital information is often restricted by law to the creator, it is important that digital repositories address the need to be able to work with and potentially modify digital objects to keep them accessible over time. Repositories should have written policies and agreements with depositors that specify and/or transfer certain rights to the repository enabling appropriate and necessary preservation actions to take place on the digital objects within the repository.

Because legal negotiations can take time, potentially slowing or preventing the ingest of digital objects at risk, a digital repository may take in or accept digital objects even with only minimal preservation rights using an open-ended agreement and address more detailed rights later. A repository’s rights must at least limit the repository’s liability or legal exposure that threatens the repository itself. A repository does not have sufficient control of the information if the repository itself is legally at risk.

Evidence: Deposit agreements; records schedule; digital preservation policies; records legislation and policies; service agreements.

A3.4 Repository is committed to formal, periodic review and assessment to ensure responsiveness to technological developments and evolving requirements.

Long-term preservation is a shared and complex responsibility. A trusted digital repository contributes to and benefits from the breadth and depth of community-based standards and practice. Regular review is a requisite for ongoing and healthy development of the repository. The organizational context of the repository should determine the frequency of, extent of, and process for self-assessment. The repository must also be able to provide a specific set of requirements it has defined, is maintaining, and is striving to meet. (See also A3.9.)

Evidence: A self-assessment schedule, timetables for review and certification; results of self-assessment; evidence of implementation of review outcomes.

A3.5 Repository has policies and procedures to ensure that feedback from producers and users is sought and addressed over time.

The repository should be able to demonstrate that it is meeting explicit requirements, that it systematically and routinely seeks feedback from stakeholders to monitor expectations and results, and that it is responsive to the evolution of requirements.

Evidence: A policy that requires a feedback mechanism; a procedure that addresses how the repository seeks, captures, and documents responses to feedback; documentation of workflow for feedback (i.e., how feedback is used and managed); quality assurance records.

A3.6 Repository has a documented history of the changes to its operations, procedures, software, and hardware that, where appropriate, is linked to relevant preservation strategies and describes potential effects on preserving digital content.

The repository must document the full range of its activities and developments over time, including decisions about the organizational and technological infrastructure. If the repository uses software to document this history, it should be able to demonstrate this tracking.

Evidence: Policies, procedures, and results of changes that affect all levels of the repository: objects, aggregations of objects; object-level preservation metadata; repository's records retention strategy document.

A3.7 Repository commits to transparency and accountability in all actions supporting the operation and management of the repository, especially those that affect the preservation of digital content over time.

Transparency is the best assurance that the repository operates in accordance with accepted standards and practice. Transparency is essential to accountability, and both are achieved through active, ongoing documentation. The repository should be able to document its efforts to make information about its development, implementation, evolution, and performance available and accessible to relevant stakeholders. The usual means of communication an organization uses to provide significant news and updates to stakeholders should suffice for meeting this requirement.

Evidence: Comprehensive documentation that is readily accessible to stakeholders; unhindered access to content and associated information within repository.

A3.8 Repository commits to defining, collecting, tracking, and providing, on demand, its information integrity measurements.

The repository must develop or adapt appropriate measures for ensuring the integrity of its holdings. The mechanisms to measure integrity will evolve as technology evolves, but currently include examples such as the use of checksums at ingest and throughout the preservation process. The chain of custody for all of its digital content from the point of deposit forward must be explicit, complete, correct, and current. The repository must demonstrate that the content it has matches the content it received, e.g., with an implemented registry function that documents content from submission onward. Losses associated with migration and other preservation actions should also be documented and made available to relevant stakeholders. (See C1.5 and C1.6.)

If protocols, rules, and mechanisms are embedded in the repository software, there should be some way to demonstrate the implementation of integrity measurements.

Evidence: An implemented registry system; a definition of the repository's integrity measurements; documentation of the procedures and mechanisms for integrity measurements; an audit system for collecting, tracking, and presenting integrity measurements; procedures for responding to results of integrity measurements that indicate digital content is at risk; policy and workflow documentation.

A3.9 Repository commits to a regular schedule of self-assessment and certification and, if certified, commits to notifying certifying bodies of operational changes that will change or nullify its certification status.

A repository cannot self-certify because an objective, external measurement using a consistent and repeatable certification process is needed to ensure and demonstrate that the repository meets and will likely continue to meet preservation requirements. Therefore, certification is the best indicator that the repository meets its requirements, fulfills its role, and adheres to appropriate standards. The repository must demonstrate that it integrates certification preparation and response into its operations and planning. (See also A3.4.)

Evidence: Completed, dated audit checklists from self-assessment or objective audit; certificates awarded for certification; presence in a certification register (when available); timetable or budget allocation for future certification.

A4. Financial sustainability

A trusted digital repository should be able to prove its financial sustainability. Overall, a trusted repository adheres to all good business practices and should have a sustainable business plan—a general set of documents that reflect the past, present, and future of the repository and its activities. A business plan incorporates management plans and financial implications related to development and normal production activities, and may note the strategies and/or risks that would affect operations.

Normal business and financial fitness should be reviewed at least annually. Standard accounting procedures should be used. Both short- and long-term financial planning cycles should demonstrate an ongoing balance of risk, benefit, investment, and expenditure. Operating budgets and reserves should be adequate.

A4.1 Repository has short- and long-term business planning processes in place to sustain the repository over time.

The repository must demonstrate that it has formal, cyclical, proactive business planning processes in place. A brief description of the repository's business plan should show how the repository will generate income and assets through services, third-party partnerships, grants, and so forth. As for A1.2 (succession/ contingency/escrow planning), the repository must establish these processes when it is viable to avoid business crises.

These questions may be pertinent to this requirement:

- Under this plan, to what extent is the repository supported, or expected to be supported, by revenue from content-contributing organizations and agencies, such as publishers?
- To what extent is the repository supported, or expected to be supported, by revenue from subscribers or subscribing institutions?
- What measures are in place, if any, to limit access by nonsubscribing stakeholders?
- What financial incentives are offered, if any, to discourage subscribers from postponing their investment in the repository? From discontinuing investing in the repository?
- To what extent is the repository supported, or expected to be supported, by other kinds of parties?
- How will major future costs, such as migrations, capital improvements, enhancements, providing access in the event of publisher failure, etc., be distributed between publishers, subscribers, and other supporting parties?
- What contingency plans are in place to cover the loss of future revenue and/or outside funding?
- In the event of a catastrophic failure, are reserve assets sufficient to ensure the restoration of subscriber access to content reasonably quickly?
- If this is a national or government-sponsored repository, how is it insulated from political events, such as international conflicts or diplomatic crises, that might affect its ability to serve foreign constituencies?

Evidence: Operating plans; financial reports; budgets; financial audit reports; annual financial reports; financial forecasts; business plans; audit procedures and calendars; evidence of comparable institutions; exposure of business plan to scenarios.

A4.2 Repository has in place processes to review and adjust business plans at least annually.

The repository must demonstrate its commitment to proactive business planning by performing cyclical planning processes at least yearly. The repository should be able to demonstrate its responsiveness to audit results, for example.

Evidence: Business plans, audit planning (e.g., scope, schedule, process, and requirements) and results; financial forecasts; recent audits and evidence of impact on repository operating procedures.

A4.3 Repository's financial practices and procedures are transparent, compliant with relevant accounting standards and practices, and audited by third parties in accordance with territorial legal requirements.

The repository cannot just claim transparency, it must 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 as transparent as it needs to be and can be within the scope of its community.

Evidence: Demonstrated dissemination requirements for business planning and practices; citations to and/or examples of accounting and audit requirements, standards, and practice; evidence of financial audits already taking place.

A4.4 Repository has ongoing commitment to analyze and report on risk, benefit, investment, and expenditure (including assets, licenses, and liabilities).

The repository must commit to at least these categories of analysis and reporting, and maintain an appropriate balance between them. The repository should be able to demonstrate that it 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.

Evidence: 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.

A4.5 Repository commits to monitoring for and bridging gaps in funding.

The repository must recognize the possibility of gaps between funding and the costs of meeting the repository's commitments to its stakeholders. It commits to bridging these gaps by securing funding and resource commitments specifically for that purpose; these commitments can come either from the repository itself or parent organizations, as applicable. Even with effective business planning procedures in place, any repository with long-term commitments will likely face some kind of resource gap in the future. The repository must provide essentially an insurance buffer as a first—and hopefully effective—line of defense, obviating the need to invoke a succession plan except in extreme situations, such as the repository ceasing operations permanently.

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

A5. Contracts, licenses, & liabilities

A repository's contracts, licenses, and liabilities should be explicit. They should define clear and measurable terms; delineate roles, responsibilities, timeframes, and conditions; and be either readily accessible or available to stakeholders on demand. Contracts include those between the repository and content owners (depositors, publishers, etc) and those between the repository and its own service providers (system service/maintenance contracts), with system developers, etc. Regardless of the relationship, these contracts and licenses must be available for audits so that liabilities and risks can be evaluated.

A5.1 If repository manages, preserves, and/or provides access to digital materials on behalf of another organization, it has and maintains appropriate contracts or deposit agreements.

Repositories, especially those with third-party deposit arrangements, should guarantee that relevant contracts, licenses, or deposit agreements express rights, responsibilities, and expectations of each party. Contracts and formal deposit agreements should be countersigned and current.

When the relationship between depositor and repository is less formal (i.e., 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 archiving may find this requirement difficult because of how Web-based information is harvested/captured for long-term preservation. This kind of data is rarely acquired with contracts or deposit agreements. By its very nature, digital information on the Web is perceived to belong to "everyone and no one." 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. Regardless of process, repositories harvesting and ingesting Web-based materials must articulate their rights issues within publicly accessible policies, and have mechanisms to respond to content owners if the repository's rights to collect and preserve certain information are challenged.

Ideally, these agreements will be tracked, linked, managed, and made accessible in a contracts database.

Evidence: Deposit agreements; policies on third-party deposit arrangements; contracts; definitions of service levels; Web archiving policies; procedure for reviewing and maintaining agreements, contracts, and licenses.

A5.2 Repository contracts or deposit agreements must specify and transfer all necessary preservation rights, and those rights transferred must be documented.

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 certain rights to the repository enabling appropriate and necessary preservation actions for the digital objects within the repository. (This requirement is linked to A3.3.)

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. A repository's rights must at least limit the repository's liability or legal exposure that threatens the repository itself.

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

A5.3 Repository has specified all appropriate aspects of acquisition, maintenance, access, and withdrawal in written agreements with depositors and other relevant parties.

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—for instance, because 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.

Evidence: Submission agreements/deposit agreements/deeds of gift; written standard operating procedures.

A5.4 Repository tracks and manages intellectual property rights and restrictions on use of repository content as required by deposit agreement, contract, or license.

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.

Evidence: A 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; demonstrable way to monitor intellectual property; results from monitoring.

A5.5 If repository ingests digital content with unclear ownership/rights, policies are in place to address liability and challenges to those rights.

The repository's policies and mechanisms must 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.

Evidence: A definition of rights; citations for relevant laws and requirements; policy on responding to challenges; documented track record for responding to challenges in ways that do not inhibit preservation; examples of legal advice sought and received.

<This page left intentionally blank.>

B. Digital Object Management

The digital object management responsibilities of a repository include both some “organizational” and technical aspects related to these responsibilities, such as repository functions, processes, and procedures needed to ingest, manage, and provide access to digital objects for the long term. Requirements for these functions are categorized into six groups based on archive functionality, allowing grouping under the well-known OAIS functional entities:

- B1: The initial phase of ingest that addresses acquisition of digital content.
- B2: The final phase of ingest that places the acquired digital content into the forms, often referred to as Archival Information Packages (AIPs), used by the repository for long-term preservation.
- B3: Current, sound, and documented preservation strategies along with mechanisms to keep them up to date in the face of changing technical environments.
- B4: Minimal conditions for performing long-term preservation of AIPs.
- B5: Minimal-level metadata to allow digital objects to be located and managed within the system.
- B6: The repository’s ability to produce and disseminate accurate, authentic versions of the digital objects.

Requirements here assume familiarity with OAIS and/or with detailed repository practices. For more information, see Appendix 4: A Perspective on Ingest; Appendix 5: Preservation Planning & Strategies; and Appendix 6: Understanding Digital Repositories & Access Functionality.

B1. Ingest: acquisition of content

Acquisition involves a crucial interaction between repository and depositor. Success in this phase of ingest indicates the repository’s ability to gain sufficient control over the content.

Repositories are likely to differ the most in this area of ingest processes, depending on the type of material they collect and their relationships with its producers. For any repository, it can be stated with some confidence that ingest finishes when an Archival Information Package (AIP) and its associated metadata are secure in the repository, including the creation of any security copies. It is more difficult to make a general statement about when ingest begins. Some repositories will have content submitted to them by producers, perhaps unexpectedly. Others will actively go out and seek content and request it from producers. Some producer-repository relationships will be more collaborative, making it less clear-cut who initiates a particular transaction.

Relationships between producers and repositories that affect ingest can differ greatly in their formality and the extent to which obligations are placed on different parties. National archives, deposit (copyright) libraries, and institutional repositories may be able to compel their producers (government agencies and publishers) to provide content, but may have little or no control over its form. Other repositories may not be able to compel producers to offer content, but might be able to select the form of acceptable content, whether that applies to file formats or minimal metadata standards, for instance. Some repositories (Web archives, for example) may have little or no relationship with the producers of the content they preserve.

Given these differences, some of the requirements here are very general and require judgments about what is appropriate for a repository given its stated mission and the needs of its designated community(ies). But the result that all repositories are trying to achieve is the same: to preserve content that is understandable and usable in the long term. For more detailed information and thorough discussion of ingest and applicability, see *Appendix 4: A Perspective on Ingest*.

B1.1 Repository identifies properties it will preserve for digital objects.

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 normalize the data during the ingest process.

Evidence: Mission statement; submission agreements/deposit agreements/deeds of gift; workflow and 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.

B1.2 Repository clearly specifies the information that needs to be associated with digital material at the time of its deposit (i.e., SIP).

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.

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 being captured.

Evidence: Transfer requirements; producer-archive agreements.

B1.3 Repository has mechanisms to authenticate the source of all materials.

The repository's written standard operating procedures and actual practices must ensure the digital objects are obtained from the expected source, that the appropriate provenance has been maintained, and that the objects are the expected objects. Confirmation can use various means including, but not limited to, digital processing and data verification and validation, and through exchange of appropriate instrument of ownership (e.g., submission agreements/deposit agreement/deed of gift).

Evidence: Submission agreements/deposit agreements/deeds of gift; workflow documents; evidence of appropriate technological measures; logs from procedures and authentications.

B1.4 Repository's ingest process verifies each submitted object (i.e., SIP) for completeness and correctness as specified in B1.2.

Information collected during the ingest process must be compared with information from some other source—the producer or the repository's own expectations—to verify the correctness of the data transfer and ingest process. 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.

Repositories should have established procedures for handling incomplete SIPs. These can range from rejecting the transfer, to suspending processing until the missing information is received, to 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. But for some activities such as Web harvesting, "complete" may simply mean "whatever we could capture in the harvest session." Whatever checks are carried out must be consistent with the repository's own documented definition and understanding of completeness and correctness.

Evidence: Appropriate policy documents and system log files from system performing ingest procedure; formal or informal "acquisitions register" of files received during the transfer and ingest process; workflow, documentation of standard operating procedures, detailed procedures; definition of completeness and correctness, probably incorporated in policy documents.

B1.5 Repository obtains sufficient physical control over the digital objects to preserve them.

The repository must obtain complete control of the bits of the digital objects conveyed with each SIP. For example, some SIPs may only reference digital objects and in such cases the repository must get the referenced digital objects if they constitute part of the object that the repository has committed to conserve. This will not always be the case: scholarly papers in a repository may contain references to other papers that are held in a different repository, or not held anywhere at all, and harvested Web sites may contain references to material in the same site or different sites that the repository has chosen not to capture or was unable to capture.

Evidence: Submission agreements/deposit agreements/deeds of gift; workflow documents; system log files from the system performing ingest procedures; logs of files captured during Web harvesting.

B1.6 Repository provides producer/depositor with appropriate responses at predefined points during the ingest processes.

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 specific, predetermined points throughout the ingest process. Responses can include initial ingest receipts, or receipts that confirm that the AIP has been created and stored. 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. Depositors can request further information on an ad hoc basis when the previously agreed upon reports are insufficient.

Evidence: Submission agreements/deposit agreements/deeds of gift; workflow documentation; standard operating procedures; evidence of "reporting back."

B1.7 Repository can demonstrate when preservation responsibility is formally accepted for the contents of the submitted data objects (i.e., SIPs).

A key component of a repository's responsibility to gain sufficient control of digital objects is the point when the repository manages the bitstream. For some repositories this will occur when it first receives the SIP transformation, for others it may not occur until the ingested SIP is transformed into an AIP. At this point, the repository formally accepts preservation responsibility of digital objects from the depositor.

Repositories that report back to their depositors generally will mark this acceptance with some form of notification 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.

Evidence: Submission agreements/deposit agreements/deeds of gift; confirmation receipt sent back to producer.

B1.8 Repository has contemporaneous records of actions and administration processes that are relevant to preservation (Ingest: content acquisition).

These records must be created on or about the time of the actions they refer to and are related to actions taken during the Ingest: content acquisition process. 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, such as PREMIS (2005), the repository must demonstrate that all relevant actions are carried through.

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

B2. Ingest: creation of the archivable package

Digital repositories must take actions to preserve the ingested information, and the things they disseminate to end users must be strongly linked to the original objects that were deposited. To paraphrase the OAIS, these requirements are meant to ensure that information (digital objects and all appropriate metadata) received and verified from each producer is put into the archival form (AIP) and is stored in archival storage for long-term preservation. More specifically, the repository must actually complete the ingest process, creating some appropriate form—identifiable as archival storage—in which to store the information. This includes addressing the linkage of appropriate metadata to meet the levels of understanding expected, the association of unique identifiers to be able to reference the digital content, the mapping from the submitted content to the AIP storage forms, and auditable provenance information ensuring no loss or corruption of content in developing the AIPs.

B2.1 Repository has an identifiable, written definition for each AIP or class of information preserved by the repository.

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 bound together and/or related in such a way that they can always be found and managed within the archive.

It is merely necessary that definitions 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.

While this requirement is primarily concerned with issues of identifying and binding key components of the AIP, B2.2 places more stringent conditions on the content of the key components to ensure that they are fit for the intended purpose. Separating the two criteria is important, particularly if a repository does not satisfy one of them. It is important to know whether some or all AIPs are not defined, or that the definitions exist but are not adequate.

Evidence: Documentation identifying each class of AIP and describing how each is implemented within the repository. Implementations may, for example, involve some combination of files, databases, and/or documents.

B2.2 Repository has a definition of each AIP (or class) that is adequate to fit long-term preservation needs.

In many cases, if the definitions required by B2.1 exist, this requirement is also satisfied, but 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, say two repositories both only preserve 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.

Evidence: Documentation that relates 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.

B2.3 Repository has a description of how AIPs are constructed from SIPs.

The repository must be able to show how the preserved object is constructed from the object initially submitted for preservation. In some cases, the AIP and SIP will be almost identical apart from packaging and location, and the repository need only state this. More commonly, complex transformations (e.g., data normalization) may be applied to objects during the ingest process, and a precise description of these actions (i.e., preservation metadata) may be necessary to ensure that the preserved object represents the information in the submitted object. The AIP construction description should include documentation that gives the provenance 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, in which case diaries or logs of actions taken to produce each AIP will be needed. In these cases, documentation needs to be mapped between to individual AIPs, and the mapping needs to be available for examination. Other repositories that can run a more production-line approach may have a description for how each class of incoming object 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.

Evidence: 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 outcome is different from SIP.

B2.4 Repository can demonstrate that all submitted objects (i.e., SIPs) are either accepted as whole or part of an eventual archival object (i.e., AIP), or otherwise disposed of in a recorded fashion.

The timescale of this process will vary between repositories from seconds to many months, but SIPs must not remain in a 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.

Evidence: System processing files; disposal records; donor or depositor agreements/deeds of gift; provenance tracking system; system log files.

B2.5 Repository has and uses a naming convention that generates visible, persistent, unique identifiers for all archived objects (i.e., AIPs).

A repository needs to ensure that an accepted, standard naming convention is in place 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.

Equally important is a system of reliable linking/resolution services in order to find the uniquely named object, no matter its physical location. This is so that actions relating to AIPs can be traced over time, over system changes, and over storage changes. Ideally, the unique ID lives as long as the AIP; if it does not, there must be traceability. The ID system must be seen to fit the repository's current and foreseeable future requirements for things like numbers of objects. It must be possible to demonstrate that the identifiers are unique. Note that B2.1 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 show how the persistent identifiers of the AIP and its components are assigned and maintained so as to be unique within the context of the repository. The documentation must also describe any processes used for changes to such identifiers. It must be possible to obtain a complete list of all such identifiers and do spot checks for duplications.

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

B2.6 If unique identifiers are associated with SIPs before ingest, the repository preserves the identifiers in a way that maintains a persistent association with the resultant archived object (e.g., AIP).

SIPs will not always contain unique identifiers when the repository receives them. But where they do, and particularly where those identifiers were widely known before the objects were ingested, it is important that they are either retained as is, or that some mechanism allows the original identifier to be transformed into one used by the repository.

For example, consider an archival repository whose SIPs consist of file collections from electronic document management systems (EDMS). Each incoming SIP will contain a unique identifier for each file within the EDMS, which may just be the pathname to the file. The repository cannot use these as they stand, since two different collections may contain files with the same pathname. The repository may generate unique identifiers by qualifying the original identifier in some way (e.g., prefixing the pathname with a unique ID assigned to the SIP of which it was a part). Or it may simply generate new unique numeric identifiers for every file in each SIP. If it qualifies the original identifier, it must explain the scheme it uses. If it generates entirely new identifiers, it will probably need to maintain a mapping between original IDs and generated IDs, perhaps using object-level metadata.

Documentation must show the policy on handling the unique identification of SIP components as the objects to be preserved are ingested, preserved, and disseminated. Where special handling is required, this must be documented for each SIP as a part of the provenance information capture (see B2.3).

Evidence: Workflow documents and evidence of traceability (e.g., SIP identifier embedded in AIP, mapping table of SIP IDs to AIPs).

B2.7 Repository demonstrates that it has access to necessary tools and resources to establish authoritative semantic or technical context of the digital objects it contains (i.e., access to appropriate international Representation Information and format registries).

The Global Digital Format Registry (GDFR), the UK National Archives' file format registry PRONOM, and the UK Digital Curation Centre's Representation Information Registry are three emerging examples of potential international standards a repository might adopt. Whenever possible, the repository should 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.

Most repositories will maintain format information locally to maintain their independent ability to verify formats or other technical or semantic details associated with each archival object. In these cases, the use of international format registries is not meant to replace local format registries but instead serve as a resource to verify or obtain independent, authoritative information about any and all file formats.

Evidence: Subscription or access to such registries; association of unique identifiers to format registries with digital objects.

B2.8 Repository records/registers Representation Information (including formats) ingested.

When international standards for the associated Representation Information are not available, the repository needs to capture such information and register it so that it is readily findable and reusable. Some of it may be incorporated into software. The Representation Information is critical to the ability to turn bits into usable information and must be permanently associated with the Content Information.

Evidence: Viewable records in local format registry (with persistent links to digital objects); local metadata registry(ies); database records that include Representation Information and a persistent link to relevant digital objects.

B2.9 Repository acquires preservation metadata (i.e., PDI) for its associated Content Information.

Preservation metadata (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.

Evidence: Viewable records in local format registry (with persistent links to digital objects); local metadata registry(ies); database records that include Representation Information and a persistent link to relevant digital objects.

B2.10 Repository has a documented process for testing understandability of the information content and bringing the information content up to the agreed level of understandability.

If Content Information or Preservation Description Information (PDI) is not directly usable by the current application tools of the designated community(ies), the repository needs to have a defined process for giving it usable form or for making additional Representation Information available (see B3.2).

Repositories that share the burden of ensuring that adequate metadata or documentation is captured or generated to meet a required degree of understandability can implement any number of procedures to address this requirement. Such repositories typically have a narrowly defined designated community, such as a particular science discipline.

Evidence: Retention of individuals with the discipline expertise; periodic assembly of designated or outside community members to evaluate and identify additional required metadata.

B2.11 Repository verifies each AIP for completeness and correctness at the point it is generated.

If the repository has a standard process to verify SIPs for either or 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. 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 such as calculating the brightness histogram of an original and preserved image), but they might be complex or contain some subjective elements.

Documentation should describe how completeness and correctness of SIPs and AIPs are ensured, starting with ensuring 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 SIP/AIP instance or class.

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

B2.12 Repository provides an independent mechanism for audit of the integrity of the repository collection/content.

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 e-mail. 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 was the e-mail lost before the SIP was constructed? This case could be resolved by the repository 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.

Evidence: Documentation provided for B2.1 through B2.6; documented agreements negotiated between the producer and the repository (see B 1.1-B1.9); logs of material received and associated action (receipt, action, etc.) dates; logs of periodic checks.

B2.13 Repository has contemporaneous records of actions and administration processes that are relevant to preservation (AIP creation).

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, such as PREMIS (2005), the repository must demonstrate that all relevant actions are carried through.

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

B3. Preservation planning

A repository or archiving system must have current, sound, and documented preservation strategies in place and demonstrably implemented. It is not enough simply to preserve information. A repository must do so in accordance with predefined, documented, preservation policies and procedures, and it must have identified mechanisms to update those policies and procedures in response to changing technologies. Without such documentation, a repository cannot pass an audit even if its work is otherwise exemplary.

Documentation need not be particularly complex. It also does not need to prescribe in detail how a repository will deal with the unknown. For instance, a repository cannot be required to document how it will preserve a file format that has not yet been invented. But it may be expected to describe what it will do when first presented with an object in a format that it has not encountered before. It may not have strategies for every single kind of file within the repository (especially important for institutions acquiring and ingesting the product of Web harvesting/archiving activities), but it needs to be able to articulate organizational awareness of the diversity of information within the repository as well as plans or assertions about the preservation strategies that will or will not be employed against certain files. Organizational policy may be to reject the object or to investigate the feasibility of dealing with it, or the decision may depend on other factors, such as who offered the object or what information it contains.

A trusted digital repository cannot simply say what it will do; it must demonstrate its policies, practices, and procedures. This documentation should be explicit, comprehensive, current, and available. For a detailed discussion of preservation planning, as well as examples of demonstrable policies, procedures, and practices required, see Appendix 5: Preservation Planning & Strategies.

B3.1 Repository has documented preservation strategies.

A repository or archiving system must have current, sound, and documented preservation strategies. These will typically address the degradation of storage media, the obsolescence of media drives, and the obsolescence of Representation Information (including formats), 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 policy on what triggers a migration and what types of migration are expected for the solution of each preservation issue identified.

Evidence: Documentation identifying each preservation issue and the strategy for dealing with that issue.

B3.2 Repository 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 (including formats) 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.

In all cases, 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(ies). This requirement ensures that the preserved information remains understandable and usable by the designated community(ies). If the mechanism depends on an external registry, the repository must demonstrate how it uses the information from that registry.

Evidence: Subscription to a format registry service; subscription to a technology watch service; percentage of at least one staff member dedicated to monitoring technological obsolescence issues.

B3.3 Repository has mechanisms to change its preservation plans as a result of its monitoring activities.

The repository must 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 unexpected ways. Plans as simple as migrating from format X to format Y when the registries show that format X is no longer supported are not sufficiently flexible—other events may have made format Y a bad choice. The repository must be prepared for changes in the external environment that may make its current plan (to migrate from X to Y in 10 years) a bad choice as the time to implement draws near. The repository should be able to show that it can revise long-range plans in light of changing circumstances.

Another possible response to information gathered by monitoring is for the repository to create additional Representation Information and/or PDI.

Evidence: Preservation planning policies tied to formal or information technology watch(es); preservation planning or processes that are timed to shorter intervals (e.g., not more than five years); proof of frequent preservation planning/policy updates.

B3.4 Repository can provide evidence of the effectiveness of its preservation planning.

The repository should be able to demonstrate the continued preservation, including understandability, of its holdings over a number of years, given the age of the repository and its holdings.

This could be evaluated at a number of degrees and depends on the specificity of the designated community(ies). 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. If judgment must be exercised as to whether adequate efforts have been made, it must be justified in detail.

Evidence: 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.

B4. Archival storage & preservation/maintenance of AIPs

There is a minimal set of conditions for performing long-term preservation of AIPs. The system infrastructure (discussed in C1) must provide suitable services to allow higher-level repository (object management) functions operating on AIPs to perform their tasks reliably. But if the higher-level functions do not use these services, or do not use them properly, then preservation is not assured. The preservation of AIPs must follow the documented preservation strategies, typically including such topics as the use of migration, transformations, checksums, multiple copies, distributed storage, and tracking of processing history that might affect preservation confidence.

B4.1 Repository employs documented preservation strategies.

Documented preservation strategies include evidence of planning for strategies not yet employed against the repository's digital objects. A repository is likely to employ multiple strategies. Different strategies may be employed by class (type) of digital object, and/or multiple strategies may be employed on a single object class. This will depend upon local repository policies and practices, though any such strategy decisions should be documented and should be based on sound community practice.

Minimally, documentation of preservation strategies must be included in repository policies and practices. Good repository practice also requires that preservation strategies employed against digital objects are recorded in the object's preservation metadata. (See also B3.3.)

Evidence: Documentation of strategies and their appropriateness to repository objects; evidence of application (e.g., in preservation metadata); see B3.3.

B4.2 Repository implements/responds to strategies for archival object (i.e., AIP) storage and migration.

At least two aspects of the strategy must be acted upon: that which pertains to how AIPs are currently stored (including physical requirements, media requirements, location of copies, formats and metadata) and that which may require AIP migration of any form. For example, AIP migrations that result in transformations of content need to be tracked to allow subsequent users to understand the repository's processing implications.

If a repository has not yet needed to carry out any sort of preservation strategy on AIP(s), it must demonstrate that its policy has not required it yet.

Evidence: Institutional technology and standards watch; demonstration of objects on which a preservation strategy has been performed; demonstration of appropriate preservation metadata for digital objects.

B4.3 Repository preserves the Content Information of archival objects (i.e., AIPs).

The repository must be able to demonstrate that the AIPs faithfully reflect what was captured during ingest and that any subsequent or future planned transformations will continue to preserve that aspect of the repository's holdings.

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 AIP should be maintained.

Evidence: Policy documents specifying treatment of AIPs and whether they may ever be deleted; ability to demonstrate the chain of AIPs for any particular digital object or group of objects ingested; workflow procedure documentation.

B4.4 Repository actively monitors integrity of archival objects (i.e., AIPs).

In OAIS terminology, this means that the repository must have Fixity Information for AIPs and must make some use of it. 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 must be able 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 someone who can maliciously alter an AIP would not likely be able to alter the Fixity Information as well. A repository should have logs that show this check being applied and an explanation of how the two classes of information are kept separate.

AIP integrity also needs to be monitored at a higher level, ensuring that all AIPs that should exist actually do exist, and that the repository does not possess AIPs it is not meant to. Checksum information alone will not be able to demonstrate this.

Evidence: Logs of fixity checks (e.g., checksums); documentation of how AIPs and Fixity information are kept separate.

B4.5 Repository has contemporaneous records of actions and administration processes that are relevant to preservation (Archival Storage).

These records must be created on or about the time of the actions they refer to and are related to actions associated with archival storage. 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, such as PREMIS (2005), the repository must demonstrate that all relevant actions are carried through.

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

B5. Information management

A critical component of any repository is its information management functionality. Regardless of technical composition and regardless of whether it is considered a “light” repository or a “dark” one—holding material for access by future generations—the system needs to be able to store, track and use metadata which supports the core functionality of the digital repository. The OAIS (2002) describes this functionality within Data Management, but, practically, this information is critical to and is generated within other digital repository functions such as ingest, archival storage, preservation planning, and access. For that reason, this section, Information Management, addresses the remaining needs associated with descriptive metadata.

Regardless of system, descriptive information (metadata) will be acquired and maintained for access and retrieval. If people cannot find what they want, the repository is not serving the needs of its users. The minimum metadata requirements for data management may be very basic. In most cases, the minimum requirement for discovery may be nothing more than an identifier a designated community uses to request a deposited object, such as a catalog number or an archival reference. People also need to know whether they are permitted to get a usable copy of it and how.

A repository’s minimum descriptive metadata requirements must match the minimum needs of the repository’s designated community(ies). This does not mean the repository needs to be able to respond to every one of its users’ requests for additional catalog information. Rather, it must assess what it can provide to a representative member of its designated community(ies), based on utility and cost. If the repository serves multiple communities, each interested in different segments of its holdings, then the minimum requirements may vary from AIP to AIP. If a repository holds both digital films and digital music, the minimum descriptive elements for film and music will differ.

Descriptive information can include much more than the narrative description that might be familiar to the user of a traditional library or archive catalog. It may also include any information that the potential user may find helpful in assessing the appropriateness and ease of use of an object, including indications of types of tools needed for use. If a repository’s holdings vary greatly in size and the larger objects are not suitable for downloading over a network connection, for instance, information about size enables a user to choose an optimum delivery method, such as a tape to be delivered by mail. Or a repository’s holdings may require special software to be available to the user to allow an object to be interpreted. Users must be able to determine this in advance, rather than possibly paying to acquire material only to discover that they do not have the tools to use it.

A repository can address this need in the more general information it makes available to its users, rather than placing specific information in the descriptive information for each AIP. For instance, a repository that holds only PDF files can:

- State in the information for each AIP that it is a PDF file.
- Have general information on how to use the repository that states that you will need a PDF reader to use its holdings.
- Define its designated community(ies) as people with access to a PDF reader.

It is the repository’s job to ensure that each and every stored object has descriptive information associated with it. This audit checklist does not specify how the repository does this, only that it must be clear how it is done. The repository may shift the burden entirely to the producers of information by requiring that say that material offered to the repository must contain a minimum amount of metadata to enable storage of descriptive information. The repository may take on the task of producing the information itself. Or it may fill in the gaps in what producers provide—using their metadata when it is sufficient, and adding metadata when it is not. Whichever the repository does, it must set out in advance the minimum metadata requirements that enable material to be discovered and identified again.

B5.1 Repository articulates minimum metadata requirements to enable the designated community(ies) to discover and identify material of interest.

Retrieval metadata is distinct from metadata that describes what has been found. For example, in a library we might say that a book's title is mandatory, but its publisher is not, because people generally search on the title.

A repository does not necessarily have to satisfy every possible request, but must be able to deal with the types of request that will come from a typical user from the designated community(ies). The minimum requirements must be articulated. The minimum may be nothing more than an identifier the designated community(ies) would know and use to request a deposited object.

Evidence: Descriptive metadata.

B5.2 Repository captures or creates minimum descriptive metadata and ensures that it is associated with the archived object (i.e., AIP).

The repository has to show how it gets its required metadata. Does it require the producers to provide it (refusing a deposit that lacks it) or does it supply some metadata itself during ingest?

Associating the metadata with the object is important, though it does not require a one-to-one correspondence, and metadata need not necessarily be stored with the AIP. Hierarchical schemes of description allow some descriptive elements to be associated with many items. The association should be unbreakable—it must never be lost even if other associations are created.

Evidence: Descriptive metadata; persistent identifier/locator associated with AIP; 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.

B5.3 Repository can demonstrate that referential integrity is created between all archived objects (i.e., AIPs) and associated descriptive information.

Every AIP must have some descriptive information and all descriptive information must point to at least one AIP, such that the integrity can be validated. This should be an easy requirement to satisfy and is a prerequisite for the next one.

Evidence: Descriptive metadata; persistent identifier/locator associated with AIP; documented relationship between AIP and metadata; system documentation and technical architecture; process workflow documentation.

B5.4 Repository can demonstrate that referential integrity is maintained between all archived objects (i.e., AIPs) and associated descriptive information.

Particular attention must be paid to operations that affect AIPs and their identifiers and how integrity is maintained during these operations. There may be times, depending on system design, when the repository cannot demonstrate referential integrity because some system component is out of action. However, repositories, must implement procedures that let them know when referential integrity is temporarily broken and ensure that it can be restored.

Evidence: Log detailing ongoing monitoring/checking of referential integrity, especially following repair/modification of AIP; legacy descriptive metadata; persistence of identifier/locator; documented relationship between AIP and metadata; system documentation and technical architecture; process workflow documentation.

B6. Access management

It must be understood that the capabilities and sophistication of the access system will vary depending on the repository's designated community(ies) and the access mandates of the repository. Because of the variety of repositories, archives, and access mandates, these criteria may be subject to questions about applicability and interpretation at a local level. For in-depth discussion of access management issues, see Appendix 6: Understanding Digital Repositories & Access Functionality.

Repositories with a mandate to provide current access must be able to produce Dissemination Information Packages (DIPs) that meet the needs of their users or are appropriate to the levels of access they offer. "Dark" archives or national archives that may have mandates restricting access for a certain number of years will produce most DIPs for internal requirements, such as performing migrations, rather than for access. In any case, any repository must be able to produce a DIP, however primitive and whatever its purpose.

These requirements ensure that access is implemented according to the repository's stated policies:

- B6.1 to B6.4 are primarily concerned with access conditions and actions related to the designated community(ies);
- B6.5 and B6.6 are primarily concerned with access security, with a focus on internal (staff) access;
- B6.7 to B6.9 ensure that the access function is implemented correctly. Access should always deliver what is required, or else make clear that it is not possible for whatever reason. Timeliness may be measured in seconds or weeks, since access may be an online function or a postal function or may be mediated through some other mechanism or a combination of them.
- B6.10 adds a specific requirement over and above the need to simply provide access to a repository's holdings. For the repository to be trusted, it must be able to provide a copy of material that can be traced back to originals.

B6.1 Repository documents and communicates to its designated community(ies) what access and delivery options are available.

Repository policies should document the various aspects of access to and delivery of the preserved information. Generally, the designated community(ies) should know the policies or at least the consequences of them. The users should know what they can ask for, when, and how, and what it costs, among other things. See Appendix 6: Understanding Digital Repositories & Access Functionality for an in-depth review of digital repository access requirements.

Repositories might have to deal with a single, homogeneous community or with multiple or disparate communities. Different policies might be needed for different communities as well as for different collection types.

Evidence: Public versions of access policies; delivery policies; fee policies.

B6.2 Repository has implemented a policy for recording all access actions (includes requests, orders etc.) that meet the requirements of the repository and information producers/depositors.

A repository need only record the actions that meet the requirements of the repository and its information producers/depositors. This may mean that little or no information is recorded about access. That is acceptable if the repository can demonstrate that it does not need to do more. Some repositories may want information about what is being accessed, but not about the users. Others may need much more detailed information about access. A policy should be established and implemented that relates to demonstrable needs. Are these figures being monitored? Are statistics produced and made available?

Evidence: Access policies; use statements.

B6.3 Repository ensures that agreements applicable to access conditions are adhered to.

The repository must be able to show what producer/depositor agreements apply to which AIPs and must validate user identities in order to ensure that the agreements are satisfied. Although it is easy to focus on denying access when considering conditions of this kind (that is, preventing unauthorized people from seeing material), it is just as important to show that access is granted when the conditions say it should be.

Access conditions are often just about who is allowed to see things, but they can be more complex. They may involve limits on quantities—all members of a certain community are permitted to access 10 items a year without charge, for instance. Or they may involve limits on usage or type of access—some items may be viewed but not saved for later reuse, or items may only be used for private research but not commercial gain, for instance.

Various scenarios may help illustrate what is required:

If a repository’s material is all open access, the repository can simply demonstrate that access is truly available to everyone.

If all material in the repository is available to a single, closed community, the repository must demonstrate that it validates that users are members of this community, perhaps by requesting some proof of identity before registering them, or just by restricting access by network addresses if the community can identified in that manner. It should also demonstrate that all members of the community can indeed gain access if they wish.

If different access conditions apply to different AIPs, the repository must demonstrate how these are realized.

If access conditions require users to make some declaration before receiving DIPs, the repository must show that the declarations have been made. These might be signed forms, or evidence that a statement has been viewed online and a button clicked to signify agreement. The declarations might involve nondisclosure or agreement to no commercial use, for instance.

Evidence: Access policies; logs of user access and user denials; access system mechanisms that prevent unauthorized actions (such as save, print, etc.); user compliance agreements.

B6.4 Repository has documented and implemented access policies (authorization rules, authentication requirements) consistent with deposit agreements for stored objects.

User credentials are only likely to be relevant for repositories that serve specific communities or that have access restrictions on some of their holdings. A user credential may be as simple as the IP address from which a request originates, or may be a username and password, or may be some more complex and secure mechanism. Thus, while this requirement may not apply to some repositories, it may require very formal validation for others. The key thing is that the access and delivery policies are reflected in practice and that the level of validation is appropriate to the risks of getting validation wrong. Some of the requirements may emerge from agreements with producers/depositors and some from legal requirements.

Repository staff will also need to access stored objects occasionally, whether to complete ingest functions, perform maintenance functions such as verification and migration, or produce DIPs. The repository must have policies and mechanisms to protect stored objects against deliberate or accidental damage by staff (see C3.3).

Evidence: Access validation mechanisms within system; documentation of authentication and validation procedures.

B6.5 Repository access management system fully implements access policy.

The repository must demonstrate that all access policies are implemented. 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.

Evidence: Logs and audit trails of access requests; information about user capabilities (authentication matrices); explicit tests of some types of access.

B6.6 Repository logs all access management failures, and staff review inappropriate "access denial" incidents.

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. This requirement does not apply to repositories with unrestricted access.

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

B6.7 Repository can demonstrate that the process that generates the requested digital object(s) (i.e., DIP) is completed in relation to the request.

If a user expects a set, the user should get the whole set. If the user expects a file, the user should get the whole file. If the user's request cannot be satisfied, the user should be told this; for instance, resource shortages may mean a valid request cannot be satisfied. Acceptable scenarios include:

The user receives the complete DIP asked for and it is clear to the user that this has happened.

The user is told that the request cannot be satisfied.

Part of the request cannot be satisfied, the user receives a DIP containing the elements that can be provided, and the system makes clear that the request is only partially satisfied.

Unacceptable scenarios include:

The request can only be partially satisfied and a partial DIP is generated, but it is not clear to the user that it is partial.

The request is delayed indefinitely because something it requires, such as access to a particular AIP, is not available, but the user is not notified nor is there any indication as to when the conflict will be resolved.

The user is told the request cannot be satisfied, implying nothing can be delivered, but actually receives a DIP, and is left unsure of its validity or completeness.

Evidence: System design documents; work instructions (if DIPs involve manual processing); process walkthroughs; logs of orders and DIP production; test accesses to verify delivery of appropriate digital objects.

B6.8 Repository can demonstrate that the process that generates the requested digital

object(s) (i.e., DIP) is correct in relation to the request.

The right material should be delivered and appropriate transformations should be applied, if necessary to generate the DIP. A simple example is that if the repository stores TIFF images but delivers JPEGS, the conversion should be shown to be correct to whatever standards seem appropriate. If the repository offers delivery as JPEG or PNG, the user should receive the format requested. Many repositories may apply more complex transformations to generate DIPs from AIPs.

Evidence: System design documents; work instructions (if DIPs involve manual processing); process walkthroughs; logs of orders and DIP production.

B6.9 Repository demonstrates that all access requests result in a response of acceptance or rejection.

Eventually a request must succeed or fail, and there must be limits on how long it takes for the user to know this. Access logs are the simplest way to demonstrate response time, even if the repository does not retain this information for long. However, a repository can demonstrate compliance if it can show that all failed requests result in an error log of some sort, and that requests are bounded in duration in some way.

Evidence: System design documents; work instructions (if DIPs involve manual processing); process walkthroughs; logs of orders and DIP production.

B6.10 Repository enables the dissemination of authentic copies of the original or objects traceable to originals.

Part of trusted archival management deals with the authenticity of the objects that are disseminated. A repository's users must be confident that they have an authentic copy of the original object, or that it is traceable in some auditable way to the original object. This distinction is made because objects are not always disseminated in the same way, or in the same groupings, as they are deposited. A database may have subsets of its rows, columns, and tables disseminated so that the phrase "authentic copy" has little meaning. Ingest and preservation actions may change the formats of files, or may group and split the original objects deposited.

The distinction between authentic copies and traceable objects can also be important when transformation processes are applied. For instance, a repository that stores digital audio from radio broadcasts may disseminate derived text that is constructed by automated voice recognition from the digital audio stream. Derived text may be imperfect but useful to many users, though these texts are not authentic copies of the original audio. Producing an authentic copy means either handing out the original audio stream or getting a human to verify and correct the transcript against the stored audio.

This requirement ensures that ingest, preservation, and transformation actions do not lose information that would support an auditable trail between the original deposited object and the eventual disseminated object. For compliance, the chain of authenticity need only reach as far back as ingest, though some communities, such as those dealing with legal records, may require chains of authenticity that reach back further.

A repository should be able to demonstrate the processes to construct the DIP from the relevant AIP(s). 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 from a large set of AIPs. 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. 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.

A repository is not required to show that every DIP it provides can be verified as authentic at a later date; it must show that it can do this when it is required at the time of production of the DIP. The level of authentication is to be determined by the designated community(ies). This requirement is meant to enable high levels of authentication, not to impose it on all copies, since it may be an expensive process.

Evidence: System design documents; work instructions (if DIPs involve manual processing); process walkthroughs; production of a sample authenticated copy; documentation of community requirements for authentication.

C. Technologies, Technical Infrastructure, & Security

These requirements do not prescribe specific hardware and software to ensure AIPs can be preserved for the long term, but describe best practices for data management and security. In total, these criteria measure the adequacy of the repository's technical infrastructure and its ability to meet object management and security demands of the repository and its digital objects.

Criteria here are similar to the good computing practices required in international management standards like ISO 17799. Repositories or organizations that have undergone ISO 17799 certification are very likely to meet many of these criteria. Providing proof of certification to relevant IT management or security standards can serve as the required evidence for some of the criteria within section C.

These requirements are grouped into three layers:

- C1: General system infrastructure requirements.
- C2: Appropriate technologies, building on the system infrastructure requirements, with additional criteria specifying the use technologies and strategies appropriate to the repository's designated community(ies).
- C3: Security–from IT systems, such as servers, firewalls, or routers to fire protection systems and flood detection to systems that involve actions by people.

C1. System infrastructure

Without a secure and trusted infrastructure, the functions carried out on AIPs cannot be trusted—they are built on a house of cards. Actions specified here are general enough to apply to systems other than repositories and archives.

C1.1 Repository functions on well-supported operating systems and other core infrastructural software.

The requirement specifies "well-supported" as opposed to manufacturer-supported or other similar phrases. The level of support for these elements of the infrastructure must be appropriate to their uses; the repository must show that it understands where the risks lie. The degree of support required relates to the criticality of the subsystem involved. A repository may deliberately have an old system using out-of-date software to support some aspects of its ingest function. If this system fails, it may take some time to replace it, if it can be replaced at all. As long as its failure does not affect mission-critical functions, this is acceptable. Systems used for internal development may not be protected or supported to the same level as those for end-user service.

Evidence: Software inventory; system documentation; support contracts; use of strongly community-supported software (i.e., Apache).

C1.2 Repository ensures that it has adequate hardware and software support for backup functionality sufficient for the repository's services and for the data held, e.g., metadata associated with access controls, repository main content.

The repository needs to be able to demonstrate the adequacy of the processes, hardware and software for its backup systems. Some will need much more elaborate backup plans than others.

Evidence: 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; "firedrills"—testing of backups; support contracts for hardware and software for backup mechanisms.

C1.3 Repository manages the number and location of copies of all digital objects.

The repository system must be able to identify the number of copies of all stored digital objects, and the location of each object and their copies. This applies to what are intended to be identical copies, not versions of objects or copies.

The location must be described such that the object can be located precisely, without ambiguity. It can be an absolute physical location or a logical location within a storage media or a storage subsystem. One way to test this would be to look at a particular object and ask how many copies there are, what they are stored on, and where they are.

A repository can have different policies for different classes of objects, depending on factors such as the producer, the information type, or its value. Some repositories may have only one copy (excluding backups) of everything, stored in one place, though this is definitely not recommended. There may be additional identification requirements if the data integrity mechanisms use alternative copies to replace failed copies.

Evidence: random retrieval tests; system test; location register/log of digital objects compared to the expected number and location of copies of particular objects.

C1.4 Repository has mechanisms in place to ensure any/multiple copies of digital objects are synchronized.

If multiple copies exist, there has to be some way to ensure that intentional changes to an object are propagated to all copies of the object. There must be an element of timeliness to this. It must be possible to know when the synchronization has completed, and ideally to have some estimate beforehand as to how long it will take. Depending whether it is automated or requires manual action (such as the retrieval of copies from off-site storage), the time involved may be seconds or weeks. The duration itself is immaterial—what is important is that there is understanding of how long it will take.

There must also be something that addresses what happens while the synchronization is in progress. This has an impact on disaster recovery: what happens if a disaster and an update coincide? If one copy of an object is altered and a disaster occurs while other copies are being updated, it is essential to be able to ensure later that the update is successfully propagated.

Evidence: Workflows; system analysis of how long it takes for copies to synchronize; procedures/documentation of operating procedures related to updates and copy synchronization; procedures/documentation related to whether changes lead to the creation of new copies and how those copies are propagated and/or linked to previous versions.

C1.5 Repository has effective mechanisms to detect bit corruption or loss.

The repository must detect data loss accurately to ensure that any losses fall within the tolerances established by policy (see A3.6). Data losses must be detected and detectable regardless of the source of the loss. This applies to all forms and scope of data corruption, including missing objects and corrupt or incorrect or imposter objects, corruption within an object, and copying errors during data migration or synchronization of copies. Ideally, the repository will demonstrate that it has all the AIPs it is supposed to have and no others, and that they and their metadata are uncorrupted.

The approach must be documented and justified and include mechanisms for mitigating such common hazards as hardware failure, human error, and malicious action. Repositories that use well-recognized mechanisms such as MD5 signatures need only recognize their effectiveness and role within the overall approach. But to the extent the repository relies on homegrown schemes, it must provide convincing justification that data loss and corruption are detected within the tolerances established by policy.

Data losses must be detected promptly enough that routine systemic sources of failure, such as hardware failures, are unlikely to accumulate and cause data loss beyond the tolerances established by the repository's policy or specified in any relevant deposit agreement. For example, consider a repository that maintains a collection on identical primary and backup copies with no other data redundancy mechanism. If the media of the two copies have a measured failure rate of 1% per year and failures are independent, then there is a 0.01% chance that both copies will fail in the same year. If a repository's policy limits loss to no more than 0.001% of the collection per year, with a goal of course of losing 0%, then the repository would need to confirm media integrity at least every 72 days to achieve an average time to recover of 36 days, or about one tenth of a year. This simplified example illustrates the kind of issues a repository should consider, but the objective is a comprehensive treatment of the sources of data loss and their real-world complexity. Any data that is (temporarily) lost should be recoverable from backups.

Evidence: Documents that specify bit error detection and correction mechanisms used; risk analysis; error reports; threat analyses.

C1.6 Repository reports to its administration all incidents of data corruption or loss, and steps taken to repair/replace corrupt or lost data.

Having effective mechanisms to detect bit corruption and loss within a repository system is critical, but is only one important part of a larger process. As a whole, the repository must record, report, and repair as possible all violations of data integrity. This means the system should be able to notify system administrators of any logged problems. These incidents, recovery actions, and their results must be reported to administrators and should be available.

For example, the repository should document procedures to take when loss or corruption is detected, including standards for measuring the success of recoveries. Any actions taken to repair objects as part of these procedures must be recorded. The nature of this recording must be documented by the repository, and the information must be retrievable when required. This documentation plays a critical role in the measurement of the authenticity and integrity of the data held by the repository.

Evidence: Preservation metadata (e.g., PDI) records; comparison of error logs to reports to administration; escalation procedures related to data loss.

C1.7 Repository has defined processes for storage media and/or hardware change (e.g., refreshing, migration).

The repository should have triggers for initiating action and understanding of how long it will take for storage media migration, or "refreshing"—copying between media without reformatting the bitstream. Will it finish before the media is dead, for instance? Copying large quantities of data can take a long time and can affect other system performance. It is important that the process includes a check that the copying has happened correctly. (See B4.2.)

Repositories should also consider the obsolescence of any/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.

Evidence: Documentation of processes; policies related to hardware support, maintenance, and replacement; documentation of hardware manufacturers' expected support life cycles.

C1.8 Repository has a documented change management process that identifies changes to critical processes that potentially affect the repository's ability to comply with its mandatory responsibilities.

Examples of this would include changes in processes in 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.

Evidence: Documentation of change management process; comparison of logs of actual system changes to processes versus associated analyses of their impact and criticality.

C1.9 Repository has a process for testing the effect of critical changes to the system.

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.

Evidence: Documented testing procedures; documentation of results from prior tests and proof of changes made as a result of tests.

C1.10 Repository has a process to react to the availability of new software security updates based on a risk-benefit assessment.

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 was 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.

Evidence: 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.

C2. Appropriate technologies

A repository should use strategies and standards relevant to its designated community(ies) and its digital technologies.

C2.1 Repository has hardware technologies appropriate to the services it provides to its designated community(ies) and has procedures in place to receive and monitor notifications, and evaluate when hardware technology changes are needed.

The repository needs to be aware of the types of access services expected by its designated community(ies), including, where applicable, the types of media to be delivered, and needs to make sure its hardware capabilities can support these services. For example, it may need to improve its networking bandwidth over time to meet growing access data volumes and expectations.

Evidence: Technology watch; documentation of procedures; designated community profiles; user needs evaluation; hardware inventory.

C2.2 Repository has software technologies appropriate to the services it provides to its designated community(ies) and has procedures in place to receive and monitor notifications, and evaluate when software technology changes are needed.

The repository needs to be aware of the types of access services expected by its designated community(ies), and to make sure its software capabilities can support these services. For example, it may need to add format translations to meet the needs of currently widely used application tools, or it may need to add a data subsetting service for very large data objects.

Evidence: Technology watch; documentation of procedures; designated community profiles; user needs evaluation; software inventory.

C3. Security

“System” here refers to more than IT systems, such as servers, firewalls, or routers. Fire protection and flood detection systems are also significant, as are systems that involve actions by people. The first two requirements here are general and the third addresses internal security, while the remainder address disaster recovery.

C3.1 Repository maintains a systematic analysis of such factors as data, systems, personnel, physical plant, and security needs.

Regular risk assessment should address external threats and denial of service attacks. These analyses are likely to be documented in several different places, and need not be comprehensively contained in a single document

Evidence: ISO 17799 certification; documentation describing analysis and risk assessments undertaken and their outputs; logs from environmental recorders; confirmation of successful staff vetting.

C3.2 Repository has implemented controls to adequately address each of the defined security needs.

The repository must 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.

Evidence: ISO 17799 certification; system control list; risk, threat, or control analyses; addition of controls based on ongoing risk detection and assessment.

C3.3 Repository staff have delineated roles, responsibilities, and authorizations related to implementing changes within the system.

Authorizations are about who can do what—who can add users, who has access to change metadata, who can get at audit logs. It is important that authorizations are justified, that staff understand what they are authorized to do, and that there is a consistent view of this across the organization.

Evidence: ISO 17799 certification; organizational chart; system authorization documentation.

C3.4 Repository has suitable written disaster preparedness and recovery plan(s), including at least one off-site backup of all preserved information together with an off-site copy of the recovery plan(s).

The repository must have a written plan with some approval process for what happens in specific types of disaster (fire, flood, system compromise, etc.) and for who has responsibility for actions. 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.

Evidence: ISO 17799 certification; 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.

<This page left intentionally blank.>

Trustworthy Repositories Audit & Certification: Criteria Checklist

<This page left intentionally blank.>

Trustworthy Repositories Audit & Certification: Criteria Checklist

Organization:		**Auditor:**		**Page**	
Section:	**A. Organizational Infrastructure**	**Interviewee(s):**		**Date**	
Aspect:	**A1. Governance & organizational viability**				

Criterion	Evidence (Documents) Examined	Findings and Observations	Result
A1.1. Repository has a mission statement that reflects a commitment to the long-term retention of, management of, and access to digital information.			
A1.2. Repository has 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.			

Trustworthy Repositories Audit & Certification: Criteria Checklist

Organization:		**Auditor:**		**Page**	
Section:	**A. Organizational Infrastructure**	**Interviewee(s)::**		**Date**	
Aspect:	**A2. Organizational structure & staffing**				

Criterion	**Evidence (Documents) Examined**	**Findings and Observations**	**Result**
A2.1. Repository has identified and established the duties that it needs to perform and has appointed staff with adequate skills and experience to fulfill these duties.			
A2.2. Repository has the appropriate number of staff to support all functions and services.			
A2.3. Repository has an active professional development program in place that provides staff with skills and expertise development opportunities.			

Trustworthy Repositories Audit & Certification: Criteria Checklist

Organization:		**Auditor:**		**Page**	
Section:	**A. Organizational Infrastructure**	**Interviewee(s):**		**Date**	
Aspect:	**A3. Procedural accountability & policy framework**				

Criterion	Evidence (Documents) Examined	Findings and Observations	Result
A3.1. Repository has defined its designated community(ies) and associated knowledge base(s) and has publicly accessible definitions and policies in place to dictate how its preservation service requirements will be met.			
A3.2. Repository has procedures and policies in place, and mechanisms for their review, update, and development as the repository grows and as technology and community practice evolve.			
A3.3. Repository maintains written policies that specify the nature of any legal permissions required to preserve digital content over time, and repository can demonstrate that these permissions have been acquired when needed.			
A3.4. Repository is committed to formal, periodic review and assessment to ensure responsiveness to technological developments and evolving requirements.			
A3.5. Repository has policies and procedures to ensure that feedback from producers and users is sought and addressed over time.			

Trustworthy Repositories Audit & Certification: Criteria Checklist

Organization:		**Auditor:**		**Page**	
Section:	**A. Organizational Infrastructure**	**Interviewee(s):**		**Date**	
Aspect:	**A3. Procedural accountability & policy framework**				

Criterion	**Evidence (Documents) Examined**	**Findings and Observations**	**Result**
A3.6. Repository has a documented history of the changes to its operations, procedures, software, and hardware that, where appropriate, is linked to relevant preservation strategies and describes potential effects on preserving digital content.			
A3.7. Repository commits to transparency and accountability in all actions supporting the operation and management of the repository, especially those that affect the preservation of digital content over time.			
A3.8 Repository commits to defining, collecting, tracking, and providing, on demand, its information integrity measurements.			
A3.9 Repository commits to a regular schedule of self-assessment and certification and, if certified, commits to notifying certifying bodies of operational changes that will change or nullify its certification status.			

Trustworthy Repositories Audit & Certification: Criteria Checklist

Organization:		**Auditor:**		**Page**	
Section:	**A. Organizational Infrastructure**	**Interviewee(s):**		**Date**	
Aspect:	**A4. Financial sustainability**				

Criterion	**Evidence (Documents) Examined**	**Findings and Observations**	**Result**
A4.1. Repository has short- and long-term business planning processes in place to sustain the repository over time.			
A4.2. Repository has in place processes to review and adjust business plans at least annually.			
A4.3. Repository's financial practices and procedures are transparent, compliant with relevant accounting standards and practices, and audited by third parties in accordance with territorial legal requirements.			
A4.4. Repository has ongoing commitment to analyze and report on risk, benefit, investment, and expenditure (including assets, licenses, and liabilities).			
A4.5. Repository commits to monitoring for and bridging gaps in funding.			

Trustworthy Repositories Audit & Certification: Criteria Checklist

Organization:		**Auditor:**		**Page**	
Section:	**A. Organizational Infrastructure**	**Interviewee(s):**		**Date**	
Aspect:	**A5. Contracts, Licenses and Liabilities**				

Criterion	**Evidence (Documents) Examined**	**Findings and Observations**	**Result**
A5.1 If repository manages, preserves, and/or provides access to digital materials on behalf of another organization, it has and maintains appropriate contracts or deposit agreements.			
A5.2 Repository contracts or deposit agreements must specify and transfer all necessary preservation rights, and those rights transferred must be documented.			
A5.3 Repository has specified all appropriate aspects of acquisition, maintenance, access, and withdrawal in written agreements with depositors and other relevant parties.			
A5.4 Repository tracks and manages intellectual property rights and restrictions on use of repository content as required by deposit agreement, contract, or license.			
A5.5 If repository ingests digital content with unclear ownership/rights, policies are in place to address liability and challenges to those rights.			

Trustworthy Repositories Audit & Certification: Criteria Checklist

Organization:		**Auditor:**		**Page**	
Section:	**B. Digital Object Management**	**Interviewee(s):**		**Date**	
Aspect:	**B.1 Ingest: acquisition of content**				

Criterion	**Evidence (Documents) Examined**	**Findings and Observations**	**Result**
B1.1. Repository identifies properties it will preserve for digital objects.			
B1.2. Repository clearly specifies the information that needs to be associated with digital material at the time of its deposit (i.e., SIP).			
B1.3. Repository has mechanisms to authenticate the source of all materials.			
B1.4. Repository's ingest process verifies each submitted object (i.e., SIP) for completeness and correctness as specified in B1.2.			
B1.5. Repository obtains sufficient physical control over the digital objects to preserve them (Ingest: content acquisition).			

Trustworthy Repositories Audit & Certification: Criteria Checklist

Organization:		**Auditor:**		**Page**	
Section:	**B. Digital Object Management**	**Interviewee(s):**		**Date**	
Aspect:	**B.1 Ingest: acquisition of content**				

Criterion	**Evidence (Documents) Examined**	**Findings and Observations**	**Result**
B1.6. Repository provides producer/depositor with appropriate responses at predefined points during the ingest processes.			
B1.7. Repository can demonstrate when preservation responsibility is formally accepted for the contents of the submitted data objects (i.e., SIPs).			
B1.8. Repository has contemporaneous records of actions and administration processes that are relevant to preservation.			

Trustworthy Repositories Audit & Certification: Criteria Checklist

Organization:		**Auditor:**		**Page**	
Section:	**B. Digital Object Management**	**Interviewee(s):**		**Date**	
Aspect:	**B.2 Ingest: creation of the archivable package**				

Criterion	**Evidence (Documents) Examined**	**Findings and Observations**	**Result**
B2.1. Repository has an identifiable, written definition for each AIP or class of information preserved by the repository.			
B2.2. Repository has a definition of each AIP (or class) that is adequate to fit long-term preservation needs.			
B2.3. Repository has a description of how AIPs are constructed from SIPs			
B2.4. Repository can demonstrate that all submitted objects (i.e., SIPs) are either accepted as whole or part of an eventual archival object (i.e., AIP), or otherwise disposed of in a recorded fashion.			
B2.5. Repository has and uses a naming convention that generates visible, persistent, unique identifiers for all archived objects (i.e., AIPs).			

Trustworthy Repositories Audit & Certification: Criteria Checklist

Organization:		**Auditor:**		**Page**	
Section:	**B. Digital Object Management**	**Interviewee(s):**		**Date**	
Aspect:	**B.2 Ingest: creation of the archivable package**				

Criterion	Evidence (Documents) Examined	Findings and Observations	Result
B2.6. If unique identifiers are associated with SIP**s** before ingest, the repository preserves the identifiers in a way that maintains a persistent association with the resultant archived object (e.g., AIP).			
B2.7. Repository demonstrates that it has access to necessary tools and resources to establish authoritative semantic or technical context of the digital objects it contains (i.e., access to appropriate international Representation Information and format registries).			
B2.8 Repository records/registers Representation Information (including formats) ingested.			
B2.9 Repository acquires preservation metadata (i.e., PDI) for its associated Content Information.			
B2.10 Repository has a documented process for testing understandability of the information content and bringing the information content up to the agreed level of understandability.			

Trustworthy Repositories Audit & Certification: Criteria Checklist

Organization:		**Auditor:**		**Page**	
Section:	**B. Digital Object Management**	**Interviewee(s):**		**Date**	
Aspect:	**B.2 Ingest: creation of the archivable package**				

Criterion	Evidence (Documents) Examined	Findings and Observations	Result
B2.11 Repository verifies each AIP for completeness and correctness at the point it is generated.			
B2.12 Repository provides an independent mechanism for audit of the integrity of the repository collection/content.			
B2.13 Repository has contemporaneous records of actions and administration processes that are relevant to preservation (AIP creation).			

Trustworthy Repositories Audit & Certification: Criteria Checklist

Organization:		**Auditor:**		**Page**	
Section:	**B. Digital Object Management**	**Interviewee(s):**		**Date**	
Aspect:	**B.3 Preservation Planning**				

Criterion	**Evidence (Documents) Examined**	**Findings and Observations**	**Result**
B3.1. Repository has documented preservation strategies.			
B3.2. Repository has mechanisms in place for monitoring and notification when Representation Information (including formats) approaches obsolescence or is no longer viable.			
B3.3 Repository has mechanisms to change its preservation plans as a result of its monitoring activities.			
B3.4. Repository can provide evidence of the effectiveness of its preservation planning.			

Trustworthy Repositories Audit & Certification: Criteria Checklist

Organization:		**Auditor:**		**Page**	
Section:	**B. Digital Object Management**	**Interviewee(s):**		**Date**	
Aspect:	**B.4 Archival storage & preservation/ maintenance of AIPs**				

Criterion	Evidence (Documents) Examined	Findings and Observations	Result
B4.1. Repository employs documented preservation strategies.			
B4.2. Repository implements/responds to strategies for archival object (i.e., AIP) storage and migration.			
B4.3 Repository preserves the Content Information of archival objects (i.e., AIPs).			
B4.4 Repository actively monitors integrity of archival objects (i.e., AIPs).			
B4.5 Repository has contemporaneous records of actions and administration processes that are relevant to preservation (Archival Storage).			

Trustworthy Repositories Audit & Certification: Criteria Checklist

Organization:		**Auditor:**		**Page**	
Section:	**B. Digital Object Management**	**Interviewee(s):**		**Date**	
Aspect:	**B.5 Information Management**				

Criterion	**Evidence (Documents) Examined**	**Findings and Observations**	**Result**
B5.1 Repository articulates minimum metadata requirements to enable the designated community to discover and identify material of interest.			
B5.2 Repository captures or creates minimum descriptive metadata and ensures that it is associated with the archived object (i.e., AIP).			
B5.3 Repository can demonstrate that referential integrity is created between all archived objects (i.e., AIPs) and associated descriptive information.			
B5.4 Repository can demonstrate that referential integrity is maintained between all archived objects (i.e., AIPs) and associated descriptive information.			

Trustworthy Repositories Audit & Certification: Criteria Checklist

Organization:		**Auditor:**		**Page**	
Section:	**B. Digital Object Management**	**Interviewee(s):**		**Date**	
Aspect:	**B.6 Access Management**				

Criterion	**Evidence (Documents) Examined**	**Findings and Observations**	**Result**
B6.1 Repository documents and communicates to its designated community what access and delivery options are available.			
B6.2 Repository has implemented a policy for recording all access actions (includes requests, orders etc.) that meet the requirements of the repository and information producers/depositors.			
B6.3 Repository ensures that agreements applicable to access conditions are adhered to.			
B6.4 Repository has documented and implemented access policies (authorization rules, authentication requirements) consistent with deposit agreements for stored objects.			
B6.5 Repository access management system fully implements access policy..			

Trustworthy Repositories Audit & Certification: Criteria Checklist

Organization:		**Auditor:**		**Page**	
Section:	**B. Digital Object Management**	**Interviewee(s):**		**Date**	
Aspect:	**B.6 Access Management**				

Criterion	**Evidence (Documents) Examined**	**Findings and Observations**	**Result**
B6.6 Repository logs all access management failures, and staff review inappropriate “access denial” incidents.			
B6.7 Repository can demonstrate that the process that generates the requested digital object(s) (i.e., DIP) is completed in relation to the request.			
B6.8 Repository can demonstrate that the process that generates the requested digital object(s) (i.e., DIP) is correct in relation to the request.			
B6.9 Repository demonstrates that all access requests result in a response of acceptance or rejection.			
B6.10 Repository enables the dissemination of authentic copies of the original or objects traceable to originals.			

Trustworthy Repositories Audit & Certification: Criteria Checklist

Organization:		**Auditor:**		**Page**	
Section:	**C. Technologies, Technical Infrastructure & Security**	**Interviewee(s):**		**Date**	
Aspect:	**C1. System Infrastructure**				

Criterion	Evidence (Documents) Examined	Findings and Observations	Result
C1.1 Repository functions on well-supported operating systems and other core infrastructural software.			
C1.2 Repository ensures that it has adequate hardware and software support for backup functionality sufficient for the repository’s services and for the data held, e.g., metadata associated with access controls, repository main content.			
C1.3 Repository manages the number and location of copies of all digital objects.			
C1.4 Repository has mechanisms in place to ensure any/multiple copies of digital objects are synchronized.			
C1.5 Repository has effective mechanisms to detect bit corruption or loss.			

Trustworthy Repositories Audit & Certification: Criteria Checklist

Organization:		**Auditor:**		**Page**	
Section:	**C. Technologies, Technical Infrastructure & Security**	**Interviewee(s):**		**Date**	
Aspect:	**C1. System Infrastructure**				

Criterion	**Evidence (Documents) Examined**	**Findings and Observations**	**Result**
C1.6 Repository reports to its administration all incidents of data corruption or loss, and steps taken to repair/replace corrupt or lost data.			
C1.7 Repository has defined processes for storage media and/or hardware change (e.g., refreshing, migration).			
C1.8 Repository has a documented change management process that identifies changes to critical processes that potentially affect the repository's ability to comply with its mandatory responsibilities..			
C1.9 Repository has a process for testing the effect of critical changes to the system.			
C1.10 Repository has a process to react to the availability of new software security updates based on a risk-benefit assessment.			

Trustworthy Repositories Audit & Certification: Criteria Checklist

Organization:		**Auditor:**		**Page**	
Section:	**C. Technologies, Technical Infrastructure & Security**	**Interviewee(s):**		**Date**	
Aspect:	**C.2 Appropriate technologies**				

Criterion	**Evidence (Documents) Examined**	**Findings and Observations**	**Result**
C2.1 Repository has hardware technologies appropriate to the services it provides to its designated communities and has procedures in place to receive and monitor notifications, and evaluate when hardware technology changes are needed.			
C2.2 Repository has software technologies appropriate to the services it provides to its designated community(ies) and has procedures in place to receive and monitor notifications, and evaluate when software technology changes are needed.			

Trustworthy Repositories Audit & Certification: Criteria Checklist

	C. Technologies, Technical Infrastructure & Security	**Auditor:**		**Page**	
		Interviewee(s):		**Date**	
	C.3 Security				

	Evidence (Documents) Examined	Findings and Observations	Result
C3.1 Repository maintains a systematic analysis of such factors as data, systems, personnel, physical plant, and security needs.			
C3.2 Repository has implemented controls to adequately address each of the defined security needs.			
C3.3 Repository staff have delineated roles, responsibilities, and authorizations related to implementing changes within the system.			
C3.4 Repository has suitable written disaster preparedness and recovery plan(s), including at least one off-site backup of all preserved information together with an off-site copy of the recovery plan(s).			

References

Consultative Committee for Space Data Systems (CCSDS). 2002. *Reference Model for an Open Archival Information System*. (ISO Standard 14721).
www.ccsds.org/publications/archive/650x0b1.pdf

———. 2003. *Producer-Archive Interface Methodology Abstract Standard*. (ISO Standard 20652).
www.ccsds.org/publications/archive//651x0b1.pdf

———. May 15, 2006. *XML Formatted Data Unit (XFDU) Structure and Construction Rules*.
sindbad.gsfc.nasa.gov/xfdu/pdfdocs/iprwbv2a.pdf

Cornell University Libraries. *Digital Preservation Management: Implementing Short-term Strategies for Long-term Problems.* 2004.
www.library.cornell.edu/iris/tutorial/dpm/index.html

ISO 9000:2000 Quality management systems—Fundamentals and vocabulary. Geneva, Switzerland: International Organization for Standardization.

ISO/IEC 17799:2005 Information technology—Security techniques—Code of practice for information security management. Geneva, Switzerland: International Organization for Standardization.

Lynch, Clifford A. February 2003. "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age." *ARL BiMonthly Report* 226.
www.arl.org/newsltr/226/ir.html

Metadata Encoding and Transmission Standard (METS) version 1.4. 2005.Washington, DC: Digital Library Federation.
www.loc.gov/standards/mets

Minnesota Historical Society, State Archives Department. 2002. *Trustworthy Information Systems Handbook.*
www.mnhs.org/preserve/records/tis/tis.html

National Institute of Standards and Technology. 2001. *Security Self-Assessment Guide for Information Technology Systems (NIST Special Publication 800-26)*. Washington, DC: NIST.
csrc.nist.gov/publications/nistpubs/800-26/sp800-26.pdf

National Institute of Standards and Technology. April 2005. *Revised NIST SP 800-26 System Questionnaire with NIST SP 800-53 References and Associated Security Control Mappings*. Washington, DC: NIST.
csrc.nist.gov/publications/drafts/Draft-sp800-26Rev1.pdf

Nestor Working Group on Trusted Repositories Certification. June 2006. *Catalogue of Criteria for Trusted Digital Repositories*. Version 1 (draft for public comment). English translation December 2006.urn:nbn:de:0008-2006060703.
edoc.hu-berlin.de/series/nestor-materialien/8en/PDF/8en.pdf

PREMIS. May 2005. *Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group*. Dublin, Ohio and Mountain View, CA: OCLC and RLG.
www.oclc.org/research/projects/pmwg/premis-final.pdf

Rosenthal, David, et al. November 2005. "Requirements for Digital Preservation Systems: A Bottom-Up Approach." *D-Lib Magazine* 11:11. www.dlib.org/dlib/november05/rosenthal/11rosenthal.html

Ross, Seamus and Andrew McHugh. 15 October 2005. "Audit and Certification of Digital Repositories: Creating a Mandate for the Digital Curation Centre (DCC)." *RLG DigiNews* 9:5. www.rlg.org/en/page.php?Page_ID=20793#article1

Task Force on Archiving of Digital Information. 1996. *Preserving Digital Information*. Washington, DC, and Mountain View, CA: Commission on Preservation and Access and the Research Libraries Group. www.rlg.org/legacy/ftpd/pub/archtf/final-report.pdf

Trusted Digital Repositories: Attributes and Responsibilities. May 2002. Mountain View, CA: RLG. www.rlg.org/en/pdfs/repositories.pdf

Appendix 1: Glossary

Many of these terms are taken from the glossary of OAIS (2002).

Archival Information Package (AIP): An Information Package, consisting of the Content Information and the associated Preservation Description Information (PDI), that is preserved within an OAIS.

Backup: The periodic capture of information to guard against system or component failure or against accidental or deliberate corruption of the system or system metadata. It is separate from the actions that most repositories will take of holding multiple copies of AIPs. Backups should ensure that lost or corrupted metadata can be restored, or that a failed system can be rebuilt and reintegrated into the repository with minimum loss of information. Backups are not expected to prevent all information loss. They are intended to restore a system or a component to a known state in a manner consistent with other system components, where this is applicable.

Content Information: The set of information that is the original target of preservation. It is composed of the digital object and its Representation Information.

Copies: Different logical or physical instances of the same object. Usually this will mean bit-wise identical copies stored on different file systems, on different media, and/or in different locations. Most, but not all, repositories will have more than one copy of each AIP to guard against media failure or system failure. Some may choose to protect against certain software failures by using two different mechanisms to store the same object—for example, by using both a TAR and a ZIP file containing the same collection of files. In this case, the bitstreams are different because the encapsulation format is different, but there is no question that they represent the same digital object. “Copies” can also be taken to refer to different forms of the same entity that a repository may choose to hold for operational reasons. One trivial example might be the storage of TIFF and JPEG versions of an image to speed the production of DIPs in JPEG format. Here one form is clearly derived from the other, but it is important that changes in one form are propagated to the other in a predictable way.

Descriptive Information: The set of information, consisting primarily of Package Descriptions, that is provided to Data Management to support the finding, ordering, and retrieval of OAIS information holdings by users.

Designated community: An identified group of potential Consumers who should be able to understand a particular set of information. The designated community may be composed of multiple user communities.

Digital repository / Digital archive: These two terms are often used interchangeably. OAIS uses *archive* when referring to an organization that intends to preserve information for access and use by a designated community(ies). *Trusted Digital Repositories: Attributes and Responsibilities* prefers the term digital repository. Digital archives and digital repositories should not be confused with either *digital libraries*, which collect and provide access to digital information, but may not commit to its long-term preservation, or *data archives*, which do commit to long-term preservation but limit their collections to statistical datasets.

Disaster: Any event that threatens or interrupts the operation of the repository and that, without corrective action, threatens the long-term preservation of its holdings. Disasters can include things that threaten the physical environment such as fire, flood, and explosion. They can also include the loss of facilities such as protracted network outages or the inability to gain access to a building for prolonged periods due to severe weather or other contingencies.

Dissemination Information Package (DIP): The Information Package, derived from one or more AIPs, received by the Consumer in response to a request to the OAIS.

Open Archival Information System (OAIS) Reference Model: Developed by the Consultative Committee on Space Data, a conceptual framework and reference tool for defining a digital repository. It provides a model of the environment, functions, and data types for implementing a digital repository. The OAIS is an official ISO standard (14721).

Preservation Description Information (PDI): The information that is necessary for adequate preservation of the Content Information; it can be categorized as Provenance, Reference, Fixity, and Context Information.

Representation Information: The information that maps a Data Object into more meaningful concepts. An example is the ASCII definition that describes how a sequence of bits (i.e., a Data Object) is mapped into a symbol.

Submission Information Package (SIP): An Information Package that is delivered by the Producer to the OAIS for use in the construction of one or more AIPs.

Users: Consumers of archived digital objects. Some digital repositories have human end users and will need to have auditable documentation and access mechanisms to support their needs. For other repositories, “users” may be restricted simply to other machines, and the repository will have associated requirements only to meet those needs.

Versions of an object: A relationship between objects. Some objects can be considered later or alternative forms of other objects, such as the director’s cut of the originally released version of a film compared, or different editions of the same book, or draft and final versions of a given document. A repository will usually choose to identify, through descriptive metadata, this type of relationship, but the relationship does not impinge on the preservation requirements of each object. This phrase will not apply to all repositories; it appears here to avoid possible confusion, as section C does not stipulate how versions of an object are handled.

Appendix 2: Understandability & Use

OAIS states that the “Repository must ensure that the information to be preserved is independently understandable to the Designated Community.” In other words, the community should be able to understand and use the information without needing the assistance of the experts who produced the information.

A repository and a producer must communicate and commit to a shared definition of “understandable” in the context of the repository. The definition may lie somewhere between “reproduce the bitstream as deposited,” in the case where the bitstream by itself is always usable by the designated community, and “ensure the Information Content is rendered or performed, intelligible, and usable to the Designated Community given its current knowledge base, tools, and practices.”

As a part of the process of submitting material to a digital repository, the repository must address the issue of the submission’s information content and the extent to which this content is understandable to its designated community. The repository’s responsibility should be defined in its charter, and it may be further elucidated in the submission agreement negotiated between the repository and the producer. The extent of this responsibility can vary widely. If the repository is only tasked to preserve bits, not information content, for a submission, then this responsibility is not relevant.

More complex cases requiring information preservation can be viewed from two extremes. When a repository has minimal responsibility, the repository may be assured by the producer that the information submitted is understandable to the designated community. The repository must have a clear definition of the designated community that includes the extent to which the repository needs to ensure the information content can be used by the community’s application tools. For example, if a designated community is defined as readers of English with access to widely available document rendering tools, the repository must ensure that the submitted information meets these criteria at the time of submission and that the corresponding information it delivers continues to do so (see B3 on preservation activities).

When a repository takes maximum responsibility, the repository cannot rely solely on the producer’s planned submission and must take additional steps to ensure that the information it receives for preservation can be understood by the designated community and is sufficiently usable. Again, the definition of the designated community should include the extent to which the repository needs to ensure the information content can be used by the community’s application tools. The steps taken may include consulting with outside sources to evaluate the degree to which the information is understandable, and efforts by the repository to gather the additional metadata needed. This enables the repository to perform information preservation as well as bit preservation, and to do so long after the original producers of the information are no longer available. The two major categories of information that must be understandable to the designated community are the Content Information and Preservation Description Information (PDI). This discussion addresses only the Content Information, but it applies to the PDI too as that will have its own Representation Information.

Once the Primary Digital Object, its Representation Information (i.e., Content Information), and a definition of “understandable” have been determined, it is possible to ask whether what the repository disseminates is understandable to the designated community. In other words, it must be possible to apply the Representation Information to the Primary Digital Object and have the result be understandable to typical members of the designated community. This application process could take place within the repository, with only the result presented to the designated community in some new representation, or it could be left for the designated community to accomplish. If the process takes place within the repository, the repository must maintain its ability to perform it. If it is left to the designated community, the repository must also maintain the Representation Information so that it is “understandable” to the designated community, and it will need to periodically verify that most of the designated community can still perform the process, or the designated community must formally commit to this responsibility.

For example, the repository may maintain software that uses, or even partially or fully embodies, the Representation Information to render the Primary Digital Object in an informative visual or auditory manner for human consumption. Alternatively, the repository's software may present all the information through an interface acceptable to designated community applications. Or the repository may provide the Primary Digital Object and Representation Information, including their relationships, directly to the designated community with the understanding that the designated community(ies) can apply the Representation Information to the Primary Digital Object to obtain understandable information. Scientific datasets often fall into this last category.

In short, the repository must maintain whatever is agreed to constitute the Content Information and its understandability requirements for the designated community. For certification, it is important for the repository to make clear its criteria for determining the Content Information and for determining the understandability requirements of its designated community so that a third party can evaluate them in specific cases and with respect to the repository's charter.

Some examples can clarify the relationships among Representation Information, the needs of the designated community, and the repository service that makes the information available to the designated community:

2. **Digital object type: Microsoft Word version 3 binary file from a government agency.**

- Representation Information: Identifier of the format being "Word v3" and being proprietary.
- Content Information: Information from a government agency in a Word document.
- Designated community: General public with access to widely available document rendering tools.
- Definition of "understandable": The Content Information is in a format currently renderable with widely available document rendering tools.
- Repository access service: Provides a binary file in a format currently renderable with widely available document rendering tools along with a unique identifier of the format type and the PDI. Upon request, may send the original binary file with its unique Representation Information identifier, assuming these are different. Note that for this proprietary format, the full Representation Information may only be available in the form of "embedded within the rendering software."

3. **Digital object type: Binary file produced by the PDF application.**

- Representation Information: Identifier of PDF-A format, described in a registry.
- Content Information: Document describing a medical procedure.
- Designated community: English readers having a knowledge base typical of second-year medical students.
- Definition of "understandable": Visually rendered exactly like visual rendering of original submission.
- Repository access service: Makes available the binary file, PDI, and PDF-A rendering application.

4. **Digital object type: Binary file containing observations from an instrument on a satellite.**

- Representation Information: Binary file format definition and the definition of the meaning of the fields in the format (including detailed sensor characteristics of the satellite instrument), all given in an EAST (a formal syntax language) description with associated Data Dictionary.
- Content Information: Data from an instrument on a satellite.
- Designated community: English readers having a third-year graduate school education in the associated

scientific discipline.

- Definition of “understandable”: Original binary file is accompanied with sufficient Representation Information to allow a member of the designated community to understand how to access all the fields in the binary file, to understand what each field means, and to understand the relationships among the fields, and, using the PDI, to understand the context in which the field values were obtained.
- Repository access service: Provides the binary file, the Data Dictionary, EAST description, PDI information, and an identifier referring to the standards document that is the definition of EAST description language.

5. **Digital object type: Software source code to perform simple function “A.”**

- Representation Information: Identification of the language the code is written in, and a pointer to a definition of that language. If available, a natural language description of what function ”A” does. Also, a description of the inputs and the expected outputs, all understandable to the designated community.
- Content Information: Understandable and usable software source code.
- Designated community: Software developers who may have an interest in code for functions like “A.”
- Definition of “understandable”: Fully documented source code is delivered with references (pointers) to primary technology dependencies such as language definition, system call, operating systems dependencies, build system, software environment requirements, relevant data standards, etc. All text is delivered in a currently usable character set. Information is sufficient to allow a member of the designated community to either compile and use the code correctly or to successfully transform the function to another language.
- Repository Access Service: Provides the software source code, Representation Information, and PDI upon request.

6. **Digital object type: Software executable code.**

- Representation Information: Identification of the platform environment in which the software can run, possibly including pointers to full descriptions of that environment and perhaps to an emulation of that environment. Hopefully, a natural-language description of what function the code performs. Also, a description of the inputs and expected outputs, all understandable to the designated community.
- Content Information: Usable software executable.
- Designated community: Software developers who may have an interest in code performing such functions.
- Definition of “understandable”: Binary object, executable in the environment specified in the Representation Information, with Representation Information and PDI that may be read and understood by members of the designated community.
- Repository Access Service: Provides the software executable code, Representation Information, and PDI upon request.

7. **Digital Object Type: Musical score in a nonproprietary binary format.**

- Representation Information: Description of the format in PDF-A, with a pointer to the PDF-A description in a different registry/repository.
- Content Information: Musical score for a synthesizer.

- Designated community: German readers who wish to generate music using a computer and synthesizer from a digital representation of the score.
- Definition of "understandable": Binary bitstream reproduced as delivered, Representation Information and PDI delivered in German.
- Repository access service: Provides the binary file, Representation Information, and PDI upon request.

Appendix 3: Minimum Required Documents

Requirements throughout this checklist refer to documents (policies, procedures, plans, etc.) that a repository should keep current. This list identifies the documents that are stipulated by one or more requirements, so a certified repository should have and keep under review at least these documents. Some repositories will provide evidence of their compliance to certain requirements using documents that do not appear on this list.

Review cycles will vary by institution. Repositories should be prepared to demonstrate that their review cycles are appropriate to their activities and requirements.

Criteria	Documents
A1.2	Contingency plans, succession plans, escrow arrangements (as appropriate)
A3.1	Definition of designated community(ies), and policy relating to service levels
A3.3	Policies relating to legal permissions
A3.5	Policies and procedures relating to feedback
A4.3	Financial procedures
A5.5	Policies/procedures relating to challenges to rights (only if likely to be needed)
B1	Procedures related to ingest
B2.10	Process for testing understandability
B4.1	Preservation strategies
B4.2	Storage/migration strategies
B6.2	Policy for recording access actions
B6.4	Policy for access
C1.7	Processes for media change
C1.8	Change management process
C1.9	Critical change test process
C1.10	Security update process
C2.1	Process to monitor required changes to hardware
C2.2	Process to monitor required changes to software
C3.4	Disaster plans

<This page left intentionally blank.>

Appendix 4: A Perspective on Ingest

“Ingest” is a generic term used by OAIS to describe the processes that take place from the receipt of the digital objects through the final, preserved form of that object in the repository. The object or objects that arrive at the repository are termed Submission Information Packages, or SIPs. By a set of processes that will inevitably be highly repository-specific, one or more SIPs are transformed into preservable digital entities called Archival Information Packages or AIPs. For some repositories there will be a one-to-one relationship between a SIP and an AIP (that is, one SIP results in the creation of one AIP) but for others the relationship will be more complex. For instance, if a repository receives the contents of a database without the schema that allow it to be interpreted, it may choose to class that as a SIP that cannot, on its own, be preserved. If it later receives the schema, that is another SIP that, when combined with the first, allows the repository to create an AIP according to its standards. OAIS uses the terms SIP and AIP primarily as a convenient shorthand to distinguish the digital objects that are received (which are often messy and incomplete) from the things that are preserved, which will generally have some structure. There is no standard format for a SIP, although individual repositories may well have such standards. Similarly, there is no standard format for an AIP, although OAIS is specific about different types of information it expects to appear in an AIP.

The digital objects a repository accepts for preservation should reflect both its mission statement and its designated communities’ spheres of interest. Users should clearly understand the relationship between a repository, its mission statement, and its collections. As a part of the ingest process, the repository must gain both physical *and* intellectual control over the digital objects. Documentation associated with the primary digital objects of a collection must be submitted and should be just as logical. The information to be transferred with specific digital objects (primary and others) will generally be enunciated in the specific transfer agreement for those objects and should be the information and items necessary for Consumers to use the objects without resort to the Producers, to other experts, or, hopefully, to subject matter experts in the repository itself.

In a general statement, it is impossible to enumerate the documentation required for each digital object being preserved by a certified repository. Complete documentation may include various types of metadata such as codes, sample forms, record layouts, codebooks, schema, explanations of the universe, minimum and maximum values, and related studies and results. The documentation is collected both to ensure completeness of the collection and to help the Consumer determine the accuracy or correctness of the data itself. That determination is normally made jointly by the Producer and the repository, with the repository acting for the Consumer.

Fundamentally, the repository is tasked to preserve information, which means digital objects together with their Representation Information. This is the primary information to be preserved and is called the Content Information in OAIS terminology. (This is also applicable to the Preservation Description Information—Provenance, Context, Fixity, and Reference.) A fundamental decision, to be taken by the repository together with the Producer, is the definition of what constitutes the information to be preserved, or Content Information. The OAIS recommendation is to start by deciding what is the Primary Digital Object and then to address the extent of the Representation Information that needs to accompany this digital object. The extent of this Representation Information is not predefined and may vary widely from one submission to another even within a given repository.

To better identify the Representation Information needed, a repository may, depending upon its mission and goals, need to consult with the producer/depositor/rights owner and systems managers, assess the digital object and determine which of its properties are significant for preservation. Other repositories and digital archives may assess needed Representation Information using automated tools that compare the digital objects against expected and/or acceptable formats or other mechanisms that analyze the content systematically as material is deposited into the repository. To ensure long-term preservation, digital repositories need to decide what level of preservation is appropriate for each digital object or class of objects. The significant properties of a digital object (i.e., the acceptable level of functionality) dictate the underlying technical form that needs to be documented and supported to ensure preservation of those properties and the amount of metadata, including detailed technical metadata, that must be stored alongside the bitstream to ensure the object is accessible to the agreed-on level.

Once the necessary Representation Information and other metadata have been submitted to the repository, it must be verified and, as necessary, enhanced to support the object's long-term maintenance as well as continuing access. The creation and maintenance of the detailed metadata associated with the object's significant properties are critical to the repository's preservation function—the detailed descriptions and the technical information necessary for interpreting the bitstream as a meaningful digital object ensure current usability by the contemporaneous designated community and form the basis of long-term preservation. How continuing access is provided over time can and should be kept separate, conceptually, from this basic preservation function.

Digital repositories can store a digital object and its associated metadata as a single bitstream, as multiple separate bitstreams, or as both. For practical reasons, repositories may prefer to store the digital object within the repository and provide only pointers or references to the associated metadata in other systems, such as bibliographic data stored in the library management system. Such "virtual encapsulation" avoids duplicating metadata, but separating a digital object and its metadata may present problems in the future. Some experts think long-term preservation may be best served by storing the digital content and as much as possible of its relevant metadata as a single file.

The Metadata Encoding and Transmission Standard (METS), MPEG-21 Digital Item Declaration Language (MPEG-21 DIDL), and the XFDU standard (the extensible data packaging format) are examples of potential AIP encapsulation structures. The METS standard (2005) was created by the cultural heritage community for encoding descriptive, administrative, and structural metadata. Depending on its use, a METS document could take the role of SIP, AIP, or Dissemination Information Package (DIP). The XFDU standard (2006) is similar to METS and can serve the same "packaging" functions as METS. Other communities may use different, community-generated packaging or encapsulation structures. The only requirement for packaging structures is that they are well documented and, if not openly accessible, the documentation can be produced on demand for auditors.

For useful examples that demonstrate the types and extent of metadata that should be collected for various types of data objects and archival information collections, see:

- Annex A of Consultative Committee for Space Data Systems, Reference Model for an Open Archival Information System (OAIS) (www.ccsds.org/publications/archive/650x0b1.pdf) (ISO 14721).
- US National Archives and Records Administration Electronic and Special Media Records Services Division, Accessioning Procedures Handbook (College Park, MD, loose leaf June 2000).

Cooperative efforts include:

- The Data Documentation Initiative (www.icpsr.umich.edu/DDI/org/index.html), formally established in 2003, which is promoting an XML Document Type Definition that has been widely adopted in some disciplines.
- The Council of European Social Science Data Archives (www.nsd.uib.no/cessda/), which promotes the preservation and exchange of data and technology and the establishment of new organizations to do the same through the use of metadata standards, common thesauri, and standardized rights management, as well as standardized cataloging of data object entries.
- The CCSDS Producer-Archive Interface Methodology Abstract Standard, 2003, (ISO Standard 20652) (www.ccsds.org/publications/archive//651x0b1.pdf), which promotes greater standardization and formalization of the pre-ingest and ingest relationships between the producers and the repositories.

Appendix 5: Preservation Planning & Strategies

A trusted digital repository must have documented preservation strategies. However, a trusted digital repository cannot simply say what it will do; it must demonstrate its policies, practices, and procedures.

The repository must be able to demonstrate:

1. **Relevant decisions about acceptable formats:** For example, standalone or portions of policies that restrict, define, or stipulate formats that may be accepted by the repository.
2. **Comprehensive automated and/or manual workflow for bringing in appropriate digital objects***:* For example, protocols for transfer, including roles and responsibilities of the producer and the repository; explicit evidence of conversions that occur in AIPs that are generated from SIPs; quality assurance mechanisms and measures for assuring the completeness and correctness of resulting AIPs.
3. **Anticipated and/or applied preservation actions pertaining to individual and classes of AIPs:** For example, preservation plans—planned, tested, and/or applied; preservation action logs; policies that address preservation strategies.
4. **Archival storage policies, procedures, and practices that ensure effective capture, ongoing and reliable archival storage, and responsiveness to inevitable technological change.** For example, storage management investment and planning documents, comprehensive security plans to enable the workflow, measures and monitoring protocols for stored AIPs.
5. **Independent means to verify expected repository content based on a secure trace of digital objects received.** For example, an auditable acquisitions register, an inventory that cannot be altered.

This is a key set of activities for collecting those things that make the information available and usable for future generations. The preservation strategy lays out a plan for carrying this out within an evolving environment (social/technological, etc.). The strategy must provide for:

- A process for monitoring change that might affect preservation.
- An understanding/expertise for interpreting the impact/implications of these changes.
- A planned response to these changes.
- An implementation of this response.

Potential strategies are:

- Transform data upon ingest of the format.
- Keep the original format and wait for others to produce a solution to the support of the software.
- Produce a supportable emulation environment to enable the proprietary software to continue to run.

Strategies may be needed for each class (e.g., format) of digital data held by the repository.

A strategy would also be expected to have special checks on AIPs over and above those performed as part of the normal robust infrastructure. These would include packaging the various components of the AIP—Content Information, Representation Information, Preservation Description Information, Packaging Information, and Package Description—and fixity checks on access to or movement of data, e.g., checksums, digests, error correction encodings, etc., including random sampling of holdings to monitor possible degradation of media. Updates are allowed to AIPs, e.g., to incorporate additional PDI. These must produce a new edition/version of an AIP.

Other transformations may be applied to SIPs and AIPs to generate further AIP versions. For example, the repository may wish to keep a more easily preservable format for a particular type of data—making life easier for the repository and more suitable for the community. It is important that contemporaneous records (e.g., logs of processes, history, etc.) be kept of these transformations as well as at least the receipt of SIPs and creation of AIPs. It is difficult to specify the level of detail of this recording. This logging may be detail enough to allow one to regenerate one version from the next or vice versa in a reversible way. In this case, the repository would be able to generate versions of AIPs as required.

Alternatively, if the logging is not sufficiently detailed for such regeneration, then each AIP version has to be kept or the deletion of intermediate versions recorded. The original AIP should never be deleted unless allowed as part of an approved strategy.

Appendix 6:
Understanding Digital Repositories & Access Functionality

The range of capabilities and degree of sophistication of a repository's access system will vary, depending on the nature of its designated community(ies) and the repository's access mandates. Repositories that are committed to providing external access, repositories must produce Dissemination Information Packages (DIPs) that meet the needs of users or are appropriate to the specified levels of access being offered. In other cases, repositories may operate as "dark archives," holding material safely for access by future generations; some archives (such as national archives) may have mandates that access to information be restricted for a certain number of years. In these latter cases, most DIPs produced by the repository will be for internal use (such as performing migrations). Consequently, all repositories must be able to produce a DIP, whatever its level of sophistication or intended use.

Repositories may also note that certain elements of an access system may be positioned beyond the boundary of what constitutes the trusted repository system. A preservation repository may not need to be able to provide "advanced" access options. Users of a preservation system may be other machines. If a repository provides a rich, complex set of access mechanisms, it may be more straightforward to view a significant portion of the access system as external to the trusted repository; the focus of an audit would be the interface between the repository and its rich access system. This approach is only effective if the internal interface is simpler than that provided to the repository's end users.

It is important to understand that repositories are not required to have rigid service levels to which they must always conform, although this is appropriate in some cases. Guarantees may be expressed to users in terms such as, "We will always ship your order within 48 hours, or we will inform you by e-mail if that is not possible." If such guarantees are provided it is important that the repository can demonstrate that they are adhered to. Where repositories are unable, due to scale or demand, to provide such assurances, it is quite acceptable for them to adopt a less strictly defined approach. For instance, repositories may say that order processing will depend upon one or more stated factors, but that users can check an order's status via the Web, e-mail, or telephone.

Universal access need not be provided in order to meet the audit checklist's access requirements. A repository's users (including the designated community or communities) may be a small or limited set of people, and confidentiality requirements or producer agreements may mean that different members may be entitled to access only highly restricted subsets of the repository's holdings. It is vital that the repository is capable of demonstrating that these restrictions are applied properly.

One dimension of trust is the trust that information producers have in the repository. Where producers place requirements on the repository to permit only specific forms of access or use to specific, identified communities, they must have confidence that the repository will implement these restrictions in a suitably secure manner.

Not all repositories will have restrictions on access. Some repositories may require anyone, including even the producers, to get a court order for access to some information. This would be appropriate for highly confidential information preserved for future access, perhaps many years hence. Some repositories might not even communicate to the public the extent, content, or nature of some or all of their holdings.

All these approaches are acceptable provided the repository makes its policies clear and can demonstrate its adherence.

Repository policies should document the various aspects of access to and delivery of the preserved information. There is an additional expectation that the policies, or at least the consequences of them, should be communicated to the designated community. Among other things, the users should know what they can ask for, when, how, and whether a charge will be levied.

The repository must make clear what users can and cannot do, what access is possible, and the mechanisms a repository provides to meet their reasonable needs. Typical questions that may be addressed include: Can users search a catalog via the Web? Can they visit the repository to speak to someone to help them find information? Can they download copies instantly or must they be ordered for delivery by mail? Can they request subsets of AIPs, or multiple AIPs in a single request? Can they choose different file formats for delivery? What types of searches can they perform?

Repository policies should clearly define access and delivery mechanisms available to its designated communities. There are no objectively compulsory request types that must be supported by every repository; instead, individual repositories must state the specific types of request that they can handle: online, batch, on-site, incidental, programmed or repeated—either by request or automatically, based on characteristics of the material. Similarly, a repository need not support any particular kind of delivery mechanism, but it must describe and communicate the types of delivery it can provide. Among other things, repositories should determine whether types of delivery are defined and announced (digital files, sent, for example, by e-mail attachment, Web, or FTP, or sent by mail on disk, tape, or in print) and whether there are limits on the types or the size of the result sets.

Where there are charges associated with using the digital objects within the repository, these should be clearly defined with respect to the repository’s designated communities within repository policies. Not all repositories will charge; some will only charge for certain services; some may have annual subscription fees with unlimited usage and others may charge per item or even per search. In some repositories, charges will be calculated automatically by an online ordering system whereas in others the charge for delivering an item may not be known until the item is produced. The latter approach may be appropriate where substantial manual work is required to produce a DIP and the work is charged by the hour, for instance. Any and all of these policies are acceptable provided the charging mechanism and the services it applies to are made known to users. The repository should also be able to demonstrate that the charging mechanisms are applied consistently.

While some repositories will deal with a single or homogeneous user community, others will be required to work with multiple or disparate communities. Where appropriate, alternative policies should be conceived to meet the challenges posed by different communities as well as for different collection types.