

Excerpted from *Trusted Digital Repositories: Attributes and Responsibilities*, an RLG-OCLC Report, May 2002 (<http://www.rlg.org/longterm/repositories.pdf>)

1 Trusted Digital Repositories

A Definition

A trusted digital repository is one whose mission is to provide reliable, long-term access to managed digital resources to its designated community, now and in the future. Trusted digital repositories may take different forms: some institutions may choose to build local repositories while others may choose to manage the logical and intellectual aspects of a repository while contracting with a third-party provider for its storage and maintenance. Whatever the overall infrastructure, however, to meet expectations all trusted digital repositories must

- accept responsibility for the long-term maintenance of digital resources on behalf of its depositors and for the benefit of current and future users;
- have an organizational system that supports not only long-term viability of the repository, but also the digital information for which it has responsibility;
- demonstrate fiscal responsibility and sustainability;
- design its system(s) in accordance with commonly accepted conventions and standards to ensure the ongoing management, access, and security of materials deposited within it;
- establish methodologies for system evaluation that meet community expectations of trustworthiness;
- be depended upon to carry out its long-term responsibilities to depositors and users openly and explicitly;
- have policies, practices, and performance that can be audited and measured; and
- meet the responsibilities detailed in Section 3 of this paper.

.....

2 Attributes of a Trusted Digital Repository

The attributes of a trusted, reliable digital repository need to be identified. A framework of attributes must accommodate all different situations and institutional responsibilities while providing a basis for expectations of a trusted repository. The following list reflects the emerging expert community's thinking about such attributes:

- Compliance with the *Reference Model for an Open Archival Information System (OAIS)*
- Administrative responsibility
- Organizational viability
- Financial sustainability
- Technological and procedural suitability

- System security
- Procedural accountability

Compliance with the *Reference Model for an Open Archival Information System (OAIS)*

A trusted digital repository will make sure the overall repository system conforms to the OAIS Reference Model. Effective digital archiving services will rely on a shared understanding across the necessary range of stakeholders of what is to be achieved and how it will be done. The Reference Model supplies a common framework, including terminology and concepts, for describing and comparing architectures and operations of digital archives. As well, the OAIS provides both a **functional model**—the specific tasks performed by the repository such as storage or access—and a corresponding **information model** that includes a model for the creation of metadata to support long-term maintenance and access. Organizations and institutions building digital repositories should commit to understanding these models and make sure all aspects of the overall system conform.

Administrative Responsibility

A trusted digital repository will provide evidence that it has a fundamental commitment to implementing the range of community-agreed standards and best practices that affect its operations—particularly those that directly influence its viability and sustainability. Administrative responsibility extends to meeting appropriate national and/or international standards for the physical environment, backup and recovery procedures, and security systems. The trusted repository will meet or exceed community standards for performance and will collect and share data measurements routinely with depositors. It will involve external community experts in validating and/or certifying its processes and procedures on a regular schedule. Written agreements with depositors will address all appropriate aspects of acquisition, maintenance, access, and withdrawal. Further, ongoing risk management and contingency planning will play a routine part of the organization’s annual strategic planning activities. A reliable repository will commit itself to transparency and accountability in all its actions.

Organizational Viability

Organizations choosing to become trusted digital repositories will establish themselves in ways that demonstrate their viability. Their mission statements will reflect a commitment to the long-term retention, management of, and access to digital cultural assets on behalf of depositors and users. Their legal status and standing will be appropriate to the range of responsibilities they are undertaking. Their business practices will be transparent and forthright. Staffing levels and areas of expertise will be appropriate to the work undertaken; further, staff training and professional development opportunities, including conference attendance and participation, will be given priority to ensure the currency of staff skill sets. The repository will continually review its policies and procedures to ensure that appropriate growth can occur and that new processes and procedures are tested for scalability. A formal succession plan or escrow arrangements will be developed in consultation with community

experts, depositors, and peer organizations that identifies all relevant content and designates trusted inheritors should the repository cease to exist.

Financial Sustainability

A trusted digital repository should be able to prove its financial sustainability over time. Overall, trusted repositories will adhere to all good business practices and should have a sustainable business plan in place. 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 commitment to a balance of risk, benefit, investment, and expenditure. Operating budgets and reserves should be adequate.

Technological and Procedural Suitability

Community experts currently advocate a range of preservation strategies. A trusted digital repository will consider all relevant options and will communicate openly about the suitability of various strategies. It will ensure that it has in place all appropriate hardware and software for inventory management functions, including all the forms of acquisition, storage, and access it offers. The repository will also have policies and plans for replacing technology as needed. The repository will comply with all relevant standards and best practices, ensuring that staff have adequate expertise to understand and implement them. It will also undergo regular external audits on its system components and performance.

System Security

All systems used in the operation of a trusted digital repository will be designed to assure the security of the digital assets. Policies and practices will meet community requirements, particularly those pertaining to copying processes, required redundancy of data, authentication systems, firewalls, and backup systems. The repository will have written policies and plans for disaster preparedness, response, and recovery, and staff will be trained appropriately. Special attention will be given to processes that address data integrity to avoid loss of data, detect changes in data, and restore lost or corrupted data. Any detected changes (including loss or corruption and restoration) will be documented and the depositor will be notified both of the changes and any actions taken.

Procedural Accountability

A trusted digital repository is responsible for a range of interrelated tasks and functions (see Section 3, Responsibilities); it will therefore be accountable for all relevant policies and procedures. Repository practices will be documented and made available on request. Monitoring mechanisms that measure and ensure the continued operation of all systems and procedures will be in place. Preservation strategies undertaken (e.g., migration, emulation, etc.) will be recorded and justified in the context of community-wide best practices. Feedback mechanisms will be in place to support the resolution of problems and to negotiate the evolving requirements between the repository, any third-party service providers, and the designated communities.

.....

3 Responsibilities of a Trusted Digital Repository

High-Level Organizational and Curatorial Responsibilities

Research repositories need to understand fully what responsibilities they should assume for the preservation of digital materials. Organizational responsibility must be understood at three basic levels. Organizations must first understand their own local requirements. Second, they need to understand which other organizations might share some of the responsibilities through geography or arrangements such as consortial agreements or shared user communities, disciplines, or format of materials. Third, they need to understand which responsibilities can be shared and how. Assuming that the general model for digital repositories is more or less distributed, its success relies on shared understanding across the federation or network of repositories of their respective duties and roles. Comprehensive coverage within the collections and effective interoperability across repositories will rely on such understandings.

Although a detailed discussion is beyond the scope of this report, a summary of these major factors is useful:¹

- the scope of collections;
- preservation and lifecycle management;
- the wide range of stakeholders;
- ownership of material and other legal issues; and
- cost implications.

The Scope of Collections

Digital materials for libraries and archives range from simple (e.g., text-based) digital files to complex multimedia and database resources. The sheer variety of digital materials and the role that they play in the collection make development and application of collections policies very challenging. The existence or lack of a physical equivalent influences decisions about whether and how the digital resource is preserved. For materials that have a physical counterpart, preservation decisions take into account considerations such as the condition of the original materials and the reason for digitizing (e.g., for increased access to the materials). Materials that are “born digital” can present more challenging problems because their “being digital” is not only a method of access, it represents their value as an information artifact. For many born-digital resources, effective preservation will rely as much on preservation of the object’s digital characteristics or properties as on preservation of its basic intellectual content. More importantly, when a library or archives digitizes its own collections, it can control decisions about standards, formats, quality control, and documentation. The preservation of materials generated outside may not include this degree of control.

Preservation and Lifecycle Management

¹ For a more complete discussion of the roles and responsibilities of different stakeholders in the lifecycle of digital materials, see Neil Beagrie and Daniel Greenstein, *A Strategic Policy Framework for Creating and Preserving Digital Collections*. Version 5.0 (Arts and Humanities Data Service Executive, 1998, updated July 2001) ahds.ac.uk/strategic.pdf.

Preservation decisions for digital items cannot wait until continued use of the materials has proved they are worth keeping. Postponing preservation decisions can and most often will result in preservation actions that are more complex, more labor intensive, and more costly. A resource can even be held hostage by an obsolete piece of software. It is also important to accept the fact that digital information is more transitory and mutable, so it may not survive benign neglect. Preservation requires active management that begins at the creation of the material and depends on a proactive approach by digital repositories and the cooperation of the stakeholders, including data providers.²

The Wide Range of Stakeholders

Content creators, systems developers, custodians, and future users are all potential stakeholders in the preservation of digital materials, and this complicates the determination of responsibilities—who, when, and for how long. Often, those creating digital materials or designing digital content management systems do not take great interest in their long-term preservation. For example, commercial publishers are justifiably interested in the preservation of their materials only as long as they are commercially viable, while libraries and their users are often interested in continued access to materials long after they cease to turn a profit. Similarly, for archives, it is usually when an electronic records management system is designed (well before any records are created) that key decisions are made that affect the long-term preservation of the records themselves. In both cases, decisions about how the materials are handled when created or maintained determine how or whether the repository can preserve them.

Ownership of Material and Other Legal Issues

Responsibility for preservation has traditionally been considered alongside ownership of the materials; that is, the owner of the materials was responsible for determining their life span. However, ownership of digital materials is not often straightforward. While a book can be taken into the collection and set upon the shelf, digital materials are less tangible. For a growing number of digital materials considered “integral” to research collections and archives, access is provided through licensing arrangements—often through a regional or national consortium. Licensing arrangements can apply to either the digital content itself or to software necessary for specific functionality and access to the content. Although the organization may own the right to access material or use the software for a specified period, there is often no guarantee of rights beyond the terms of the license. While commercial publishers are beginning to provide some guarantee of continuing access, most licensing agreements are still perilously vague about how the digital repository will be maintained and how long-term access will be ensured. Reliance solely on creators or producers of digital materials for long-term preservation is potentially risky, not least because digital resources are not generally created or engineered with long-term preservation in mind.

It will be critical in the future for research repositories to work as closely as possible with content creators to ensure that long-term preservation responsibilities are clearly understood and documented in licensing agreements; this is currently being explored by The Andrew W.

² Ibid

Mellon Foundation's e-Journal archiving program.³ It will require increased cooperation and effective communications with publishers, software suppliers, and other producers to ensure that what is deposited is a copy of the data object in the format most suitable for preserving the materials over the long term. Understanding the important difference between long-term preservation and short-term access—particularly while materials are still commercially viable—is critical. Libraries may require different license arrangements for long-term preservation than for end-user access.

Often, rights that relate to the software and systems used to create the material impinge on its preservation. Very little, if any work, has been done with software vendors to raise awareness about the longevity of their materials in the interests of future scholarship and research.

Digital preservation has even wider legal implications. How preservation infringes on copyright remains unclear. For example, the content creator does not usually own the rights to the software and systems used to create the digital file. This raises legal issues when access or changes to those systems are necessary. In such cases, at best, a repository will need to arrange separate rights clearance for long-term maintenance; at worst, preservation will be compromised because rights clearances for access cannot be obtained. Some work has been done on the establishment of repositories for software to help address these concerns, however the research repository community will need to make an appeal to have this conflict taken into consideration in the creation or renewal of national deposit legislation.

³ A program funded by The Andrew W. Mellon Foundation designed to plan the development of e-journal repositories meeting specific requirements developed by the Digital Library Federation (DLF), Council for Library and Information Resources (CLIR), and Coalition for Networked Information (CNI). Seven major libraries have received grants, including the New York Public Library and the university libraries of Cornell, Harvard, MIT, Pennsylvania, Stanford, and Yale; see www.clir.org/diglib/preserve/ejp.htm.