# Minimum criteria
# for an archival repository of digital scholarly journals

**Version 1.2, May 15, 2000**

## Introduction

This document sets out the minimum criteria of a digital archival repository that acts to preserve digital scholarly publications. It is based closely on the Reference Model for an Open Archival Information System and modified to reflect the specific needs of library, publishing, and academic communities. It also indicates some of the key research issues that are likely to emerge for those who establish digital archival repositories that meet these criteria. The research issues are divided into three categories: those associated with the deposit of data, those associated with preservation, and those associated with access.

At the outset, Dan Greenstein and Deanna Marcum extracted the relevant sections of the OAIS Reference Model and presented criteria to a group of fifteen librarians for review and comment. The librarians suggested a number of changes, and the document was modified to reflect their views (Version 1.1 at http://www.diglib.org/ preserve/archreq.htm). On May 1, a group of commercial and non-profit scholarly journal publishers met to review the minimum criteria. They propose the adaptations found in this version of the criteria (Version 1.2)

## Criterion 1. A digital archival repository that acts to preserve digital scholarly publications will be a trusted party that conforms to minimum requirements agreed to by both scholarly publishers and libraries.

Agreed minimum criteria are essential. Libraries need them to assure themselves and their patrons that digital content is being maintained. Publishers need them so they may demonstrate to libraries, but also to their authors, that they are taking all reasonable measures to ensure persistence of their publications. Finally, emerging repositories need them as a blueprint for services, but also as a benchmark against which service can be measured, validated, and above all, trusted by the libraries and publishers that rely upon them.

Trusted parties may include libraries, publishers, or third parties providing archival services.

The key research question entails the definition of those criteria. Initial meetings with librarians and publishers are an essential first step in developing these definitions. Their refinement is expected to be an iterative process, one that takes account of experience in building, maintaining, and using digital archival repositories.

**Criterion 2. A repository will define its mission with regard to the needs of scholarly publishers and research libraries. It will also be explicit about which scholarly publications it is willing to archive and for whom they are being archived.**

This definition will help to focus the repository on the nature and extent of digital information it will acquire and on the requirements of the research library as the primary recipient of any data disseminated by the repository.

**Research issues:**

- Mission statements that document the scope and nature of materials a repository aims to collect, the strategy and methods it adopts for developing its collections (attracting deposits), and the community of libraries (and other users) it seeks to serve. The statement of scope should use a common syntax that is universally accepted.

- The development of registries that document what scholarly publications are archived where (and implicitly those not archived at all) is a further research issue.

**Criterion 3. A repository will negotiate and accept appropriate deposits from scholarly publishers.**

A repository will develop criteria to guide consideration of what publications it is willing to accept. Criteria may include subject matter, information source, degree of uniqueness or originality, and the techniques used to represent the information.

Individual negotiations with publishers may result in deposit agreements between the repository and the data producer. Deposit agreements may identify the detailed characteristics of the data and accompanying metadata that are deposited; the procedures for the deposit; the respective roles, responsibilities, and rights of the repository; the data procedure with regard to those data; references to the procedures and protocols by which a repository will verify the arrival and completeness of the data; etc. The deposit will come with a schedule in which that publisher states what is being deposited and the repository will verify the deposit.

**Research issues:**

**Deposit**

- Selection criteria used by the repository to review potential accessions.

- Guidelines for depositors that identify preferred or required data and metadata formats, transmission methods and media, etc.

- Procedures for verifying the arrival and completeness of deposited data and metadata.

- Adherence by several archives to some common range of data and/or metadata formats.

## Criterion 4. A repository will obtain sufficient control of deposited information to ensure its long-term preservation.

In this respect, a repository will at a minimum require licenses that allow it sufficient control to accession, describe, manage, even transform deposited data and accompanying metadata for the sake of their preservation. Publishers may want to negotiate re-depositing when migration occurs. In any event, publishers must have the right to audit the contents of their deposited data. Where repositories act in association with one another (e.g., to ensure sufficient redundancy in the preservation process), they may also require rights allowing them to mirror or deposit data with other associated archives.

Further, repositories will need to pay attention to whether and how their rights and responsibilities with regard to any particular deposit may change through time. For example, where a depositor ceases to supply its materials to the scholarly community, the repository must be positioned to supply those materials to existing licensees (perhaps at a fee). Similarly, there must be a statement about the rights of the publisher if a repository goes out of business.

**Research issues:**

**Deposit**

- Fuller understanding of how a respository's rights and responsibilities change over time.

**Access**

- Acceptable licenses and licensing principles.

## Criterion 5. A repository will follow documented policies and procedures which ensure that information is preserved against all reasonable contingencies.

Preservation strategies and practices are not right or wrong, but more or less fit for their intended purposes. No general theory of digital preservation or data migration is likely to become available soon. Thus, data in different formats may require different strategies and these may need to be worked out with the data producer (depositor). Documenting how and where different preservation strategies and practices prove cost effective and fit for their intended purposes will be a primary interest of any coordinated approach to developing preservation capacities appropriate to scholarly publishing, research libraries,

and academic communities. Because preservation practices are likely to vary across repositories, and because we have an interest in encouraging the development of different practices, we may wish simply to request that participants in any such coordinated effort agree to document the practices they adopt and disclose them to some community review and evaluation.

**Research Issues:**

**Deposit**

- Preservation metadata

**Preservation**

- Migration strategies (and their application with specific data formats)

- Data validation

- Scaleable infrastructure

## Criterion 6: A repository will make preserved information available to libraries, under conditions negotiated with the publisher.

Although repositories will need to support access at some level, those services should not replace the normal operating services through which digital scholarly publications are typically made accessible to end users. The access rights must be made explicit and must be mutually agreed upon by the publisher and the repository.

**Research issues:**

**Access**

- Resource discovery mechanisms

- Access (data dissemination) strategies supported by archives

- User licenses and how enforced

- Template licensing arrangements

## Criterion 7. Repositories will work as part of a network.

At a minimum, respositories will need to operate as part of a network to achieve a satisfactory degree of redundancy for their holdings. Although an appropriate level of redundancy is difficult to quantify let alone mandate, it will ideally extend for any single data to three archival sites, at least one of which is located off shore.

A network of repositories offers additional advantages to libraries and scholarly publishers. Libraries may benefit from common finding aids, access mechanisms, and registry services that are supported by a network and allow libraries more uniformly to identify trusted repositories. Publishers may benefit from having access to a single repository or group of repositories that specialize in publications of a particular type and from the cost efficiencies that emerge from within a network.

**Research issues:**

**Perceived Value of Deposit**

- Standard methods for data deposit

- Standard deposit licenses and/or user agreements

**Perceived Value of Preservation**

- Standard preservation and other metadata

- Standard migration strategies and implementation procedures

- Standard specifications for physical media

- Standard accreditation of requirement conformant archives

**Perceived Value of Access**

- Standard interfaces among repositories

- Standard methods for data dissemination

- Standard resource discovery practices.