

**IDEALS Digital Preservation  
Current Status and Future Directions  
October 2006  
IDEALS Preservation Working Group**

## **Executive Summary**

A joint initiative between the University Library and CITES with funding from the Provost's Office, the Illinois Digital Environment for Access to Learning and Scholarship (IDEALS) is a set of collections and related services that together constitute the campus institutional repository. Building an institutional repository means that the Library and CITES are making a specific commitment to the campus and the depositors – an extension of the same type of social contract that traditional libraries and archives hold with their constituents. A crucial part of this commitment is the long term preservation and provision of persistent and reliable access to the digital works of faculty, staff, and students.

Ensuring that IDEALS is indeed worthy of the trust of the faculty, students, and staff of the institution requires careful and deliberate planning and ongoing attention to digital preservation. With this in mind, this paper provides a roadmap for the implementation of a comprehensive digital preservation management program for IDEALS. Specifically, this paper outlines the current state of digital preservation for the pilot stage of IDEALS and specifies the minimum requirements for the ongoing production level work. **Most importantly, this paper recommends to CITES and the University Library that IDEALS aims to meet the requirements of a Trusted Digital Repository (TDR) as defined in the RLG/OCLC report and the final report, when issued, of the Audit Checklist for Certification of Trusted Digital Repositories. In addition the long-term goal of IDEALS should be to obtain certification as a trusted digital repository.**

What this means at a high level for CITES and the University Library is:

- There is a specific commitment on the part of both organizations to the long term preservation of the content deposited in IDEALS. This commitment needs to be independent of the people making the commitment, i.e. it needs to be built into the policies and procedures as well as the missions of both organizations.
- Resources in terms of funding and staff are specifically allocated towards the long term preservation of content deposited into IDEALS.
- Processes, policies, and the institutional commitment are documented and as transparent as possible.
- The technical infrastructure is further developed using the TDR checklist and *The Reference Model for an Open Archival Information System (OAIS)* as a framework. This may mean extending DSpace or using an additional system for the archival storage.
- There is a commitment of resources for planning and participation in community standards building in the arena of digital preservation.

In order to make steps towards this goal, this paper sets out a number of specific initial and longer term implementation steps for CITES and the University Library to move towards together.

## Introduction

A joint initiative between the University Library and CITES with funding from the Provost's Office, the Illinois Digital Environment for Access to Learning and Scholarship (IDEALS) is a set of collections and related services that together constitute the campus institutional repository. As part of its mission and charter, the IDEALS initiative commits to the long term preservation and provision of persistent and reliable access to the digital works of faculty, staff, and students. This includes:

- A technological infrastructure that will support long term preservation and ability to migrate content so it maintains essential functionality;
- An easy to use service or set of services that provides reliable and persistent access to the digital research output of faculty, staff, and students at UIUC; and
- A service or set of services that provides reliable storage and long term preservation of the digital research output of faculty, staff, and students at UIUC.

Building an institutional repository means that the Library and CITES are making a specific commitment to the campus and the depositors – an extension of the same type of social contract that traditional libraries and archives hold with their constituents. Clifford Lynch sets forth this commitment:

It's vital that institutions recognize institutional repositories as a serious and long-lasting commitment to the campus community (and to the scholarly world, and the public at large) that should not be made lightly. In establishing institutional repositories, institutions are both accepting risks and making promises; they are creating new expectations. In a budget crunch, the institutional repository may be one of the last things that can be cut, given the way that *digital preservation demands steady and consistent attention and hence funding*. Faculty who choose to rely on institutional repositories to disseminate and preserve their work are placing a great deal of trust in their institution and in the integrity, wisdom, and competence of the people who manage it. We need to ensure that our institutional repositories are worthy of this trust. (*emphasis added*)<sup>1</sup>

Ensuring that IDEALS is indeed worthy of the trust of the faculty, students, and staff of the institution requires careful and deliberate planning and ongoing attention to digital preservation. With this in mind, this white paper provides a roadmap for the implementation of a comprehensive digital preservation management program for IDEALS. Specifically, this paper:

- Presents the current state of digital preservation for the pilot stage (through Spring 2007) of IDEALS;
- Specifies the minimum requirements for the ongoing production-level service (Summer 2007 and forward);
- Recommends that IDEALS pursues certification as a Trusted Digital Repository for endorsement by the IDEALS Working Group and Steering Committee; and
- Identifies the next steps for moving towards compliance as a Trusted Digital Repository.

---

<sup>1</sup>Clifford Lynch. *ARL Bimonthly Report* 226. <<http://www.arl.org/newsltr/226/ir.html>>

## The IDEALS Preservation Working Group

Responsible for developing preservation policies and plans for the IDEALS initiative, the IDEALS Preservation Working Group began meeting in September 2005. The working group's charge includes making recommendations on the preservation strategies as well as the organizational and technical resources and infrastructure needed to meet the stated policy. Members of this working group include: Tom Teper (lead), Tim Donohue, Joanne Kaczmarek, Sue Lewis, and Sarah Shreeves.

Specifically, the IDEALS Preservation Working Group is tasked with:

- An assessment of the current state of digital preservation in the IR environment and what we need to do to meet minimal needs (as defined by the team).
- An assessment of what formats and materials we are committing to support for the long term.
- An interim or 'stop-gap' preservation policy available for the pilot phase of IDEALS.
- A statement of what level of digital preservation the IDEALS initiative is aiming for by the end of the six year initiative in relation to the OAIS Reference Model.
- A plan of what steps are needed to assure adequate resources and infrastructure are in place to meet the goals and key deliverables of IDEALS.
- A preservation policy for Phase II (Production) IDEALS environment.

It is worth stating outright that IDEALS Preservation Working Group (WG) follows the Cornell University Library definition of digital preservation:

Digital Preservation is a management process that encompasses a broad range of activities designed to extend the usable life of machine-readable computer files and protect them from media failure, physical loss, and obsolescence. This management activity involves maintaining the bitstream ... as well as accessibility to the contents. This accessibility is measured through the ability of content to be viable, renderable, and understandable. As these terms imply, it is one thing to preserve a bitstream, but quite another to preserve the content, form, style, appearance, and functionality. (Digital preservation is a) process that requires the use of the best available technology as well as carefully thought out administrative policies and procedures.<sup>2</sup>

The current state of digital preservation in relation to the development of institutional repositories is characterized by a fair amount of uncertainty. Billed as entities designed to preserve deposited content, there is little evidence to suggest that any existing repository software has developed an operating model that will ensure perpetual access without systematic and intensive management. The WG believes that digital preservation requires more than technological solutions; it will require organizational commitment, well defined understood and documented policies and procedures, and dedicated resources.

---

<sup>2</sup> Cornell University Library, Digital Preservation Management: Implementing Short-term Strategies to Long-term Problems. <<http://www.library.cornell.edu/iris/tutorial/dpm/terminology/preservation.html>>

## **Current Digital Preservation Activities for Pilot Phase of IDEALS**

### **Interim Preservation Policy**

As tasked, the IDEALS Preservation Working Group initially focused on developing and implementing an interim or "Stop-Gap" Preservation Policy for the pilot phase of the repository's development. This Stop-Gap Preservation Policy is a minimal level of support for digital preservation that addresses immediate technology and procedural issues. We are currently operating under an implicit organizational and resource commitment as expressed through the funding of the IDEALS initiative.

During the pilot phase, IDEALS' ability to preserve deposited items has focused on safe and frequent backups of data. The interim policy outlines three general areas of concern: Repository Support, Content Preservation, and Preservation Metadata. Each of these areas address issues such as permissible data loss and downtime, planning for data back-up, the development of procedures to ensure data and format integrity, and disaster planning.

While many of the provisions outlined in this document represent in some cases more of an acknowledgement of a problem than a solution, the goal was to provide early adopters with realistic expectations of the capabilities of IDEALS during the pilot phase. Very little is promised in the Interim Preservation Policy, but it alludes to the preparations and preservation planning taking place in anticipation of the ongoing production service.

The full Interim Preservation Policy is available at:  
<<http://ideals.uiuc.edu/about/IDEALSPreservationSupport.html>>.

### **Environment and Back-up Procedures for Pilot Phase**

As of September 30 2006, the technical infrastructure (hardware, software, storage, and programming expertise) is entirely housed in the Grainger Engineering Library and Information Center. CITES provides support for the operating system.

To minimize any loss of content due to server issues, daily backups are performed by CITES, the system administrator at Grainger, and the Library Systems office. The procedures to restore IDEALS from back-up have been documented and tested. IDEALS procedures are available at: <<https://www.ideals.uiuc.edu/wiki/bin/view/IDEALS/SystemDocs>> (IDEALS staff only).

In addition, weekly checksum verification and virus checking upon download of materials has been implemented.

### **Format Support Policy**

The WG was also tasked with an assessment of what formats IDEALS was committed to support for the long term. The WG determined this was less a question of what formats and materials IDEALS would support, than a question of at what level IDEALS would support a variety of formats. The Format Support Policy reflects discussions about IDEALS' ability to provide varying levels of support for various file types and how those levels were to be defined. The working group agreed upon the following categories of file format support:

- Category 1 - Highest Confidence - Full Support
- Category 2 – Moderate Confidence - Intermediate Support
- Category 3 – Low Confidence - Basic Preservation Only

Among other criteria, these categories are largely dependent upon whether the file format:

- Is openly documented (more preservable) or proprietary (less preservable);
- Is supported by a range of software platforms (more preservable) or by only one (less preservable);
- Is widely adopted (more preservable) or has low use (less preservable);
- Is lossless data compression (more preservable) or lossy data compression (less preservable);
- Contains embedded files or embedded programs/scripts (less preservable).

It is our belief that by evaluating files deposited into IDEALS with these criteria, we can make reasonable assumptions about IDEALS' ability to maintain deposited content. It is likely that, while a single format's ranking within the criteria may change, the criteria will remain the same.

The full IDEALS Format Support Policy is available at <https://www.ideals.uiuc.edu/wiki/bin/view/IDEALS/FormatSupport>.

### **IDEALS Format Support Matrix**

Working with the criteria outlined in the IDEALS Format Support Policy, the working group developed a format support matrix. This document provides users with a clear view of evaluated formats for submission of files and may be viewed at:

<https://www.ideals.uiuc.edu/wiki/bin/viewfile/IDEALS/Preservation?rev=1.2&filename=FormatSupportMatrix.xls>.

This document will be reviewed and updated on an ongoing basis.

## **Recommendations for Developing IDEALS' Capacity as a Preservation Environment**

### **Minimum Digital Preservation Requirements for Production**

Procedures referenced in the above policy documents for the pilot phase that have not been fully implemented are (in no particular order):

- Virus-checking upon ingest (*virus checking exists on the IDEALS server itself*)
- Creation of preservation and access copies for Category 1 and 2 formats upon ingest (*currently testing service in October 2006*)
- Procedure for analysis of submitted unknown formats for Category 1, 2, or 3 support
- Disaster plan (*restore procedures from back-up have already been tested*)
- Off-campus backups
- Generation/creation of preservation metadata.

The preservation policies and procedures set out for the pilot phase are not adequate for the production environment. At a minimum, all of the procedures outlined above need to be implemented and documented prior to initiating the production phase of IDEALS.

### **IDEALS as a Trusted Digital Repository**

The IDEALS Preservation Working Group believes that IDEALS must develop itself as a model institutional repository by fully embracing the attributes of a Trusted Digital Repository (TDR), as based on RLG/OCLC's report *Trusted Digital Repositories: Attributes and Responsibilities* <<http://www.rlg.org/legacy/longterm/repositories.pdf>>.

The RLG/OCLC report defines a Trusted Digital Repository (TDR) as a repository 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 have seven attributes:

- Acceptance of responsibility for the long-term maintenance of digital resources on behalf of its depositors and for the benefit of current and future users;
- An organizational system that supports not only long-term viability of the repository, but also the digital information for which it has responsibility;
- Demonstrated fiscal responsibility and sustainability;
- Design of its system(s) in accordance with commonly accepted conventions and standards to ensure the ongoing management, access, and security of materials deposited within it;
- Establishment of methodologies for system evaluation that meet community expectations of trustworthiness;
- Carries out its long-term responsibilities to depositors and users as openly and explicitly as possible; and
- Policies, practices, and performance that can be audited and measured.

In practice, these seven attributes are outlined and codified within the Audit Checklist for Certification of Trusted Digital Repositories. This checklist, issued by RLG and the National Archives and Records Administration (NARA) is designed to help institutions determine their level of readiness for supporting the long-term management of digital resources. A draft report is currently available at <<http://www.rlg.org/en/pdfs/rlgnara-repositorieschecklist.pdf>>.

**The IDEALS Preservation Working Group recommends to CITES and the University Library that IDEALS aims to meet the requirements of a Trusted Digital Repository as defined in the RLG/OCLC report and the final report, when issued, of the Audit Checklist for Certification of Trusted Digital Repositories. In addition the long-term goal of IDEALS should be to obtain certification as a trusted digital repository.**

What this means at a high level for CITES and the University Library is:

1. There is a specific commitment on the part of both organizations to the long term preservation of the content deposited in IDEALS. This commitment will continue even if IDEALS as a service is discontinued or morphed into something else. This commitment needs to be independent of the people making the commitment, i.e. it needs to be built into the policies and procedures as well as the missions of both organizations.

2. Resources in terms of funding and staff are specifically allocated towards the long term preservation of content deposited into IDEALS. For example, the job description for the IDEALS technical lead should explicitly include responsibility for digital preservation rather than an assumption that this is the case.
3. The TDR checklist, in its current form, stresses the importance of transparency and documentation of processes, policies, and commitment.
4. The technical infrastructure is further developed using the TDR checklist and *The Reference Model for an Open Archival Information System (OAIS)* as a framework. This may mean extending DSpace or using an additional system for the archival storage.
5. There is a commitment of resources for planning and participation in community standards building in the arena of digital preservation. Participation in conferences and training sessions allow staff to keep up with changes in this area and to contribute to evolving standards.

In order to make steps towards this goal, the WG has set out a number of high-level objectives that map to the seven attributes of a TDR. Within these, we have identified where work needs to happen within the pilot phase (through Spring 2007) and what can happen throughout the ongoing production phase (Summer 2007 forward).

### **Work to Occur in the Pilot Phase (through Spring 2007)**

#### **0. OAIS compliance**

*The Reference Model for an Open Archival Information System (OAIS) (ISO 14721)* is a conceptual model that provides a common framework, set of terminology, and range of concepts to describe and compare the architectures and operations of archives. Developed to manage access to content held by any type of archive, the OAIS has been readily and actively applied to the architecture required for maintaining access to digital content. The complete Reference Model may be accessed at: <http://nost.gsfc.nasa.gov/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf>.

Technologically, procedurally, and administratively, IDEALS should embrace the OAIS Reference Model as the basis for its organization and operations. As a conceptual model, the OAIS is flexible enough to permit individuals from many units and operations to participate in the development and maintenance of the repository.

As a first phase, the major components (producer, management, consumer, submission information package (SIP), archival information package (AIP), and dissemination information package (DIP)) of the OAIS Reference Model should be identified within the IDEALS environment and the major functions (Ingest, Data Management, Archival Storage, Access, Preservation Management, and Administration) modeled.

Gaps between the current environment and the reference model should be identified, documented and prioritized. Gaps should be prioritized based on the immediate importance to IDEALS and its relationships with its content producers and consumers.



It is important that these gaps also be communicated to our producers and consumers, in order to promote realistic expectations regarding the current state of IDEALS and its ability to preserve and provide access to content under its care.

#### **1. Administrative responsibility**

Arrangements between the University Library and CITES for data management, project development and continued support for IDEALS must be clearly documented and publicly available. The ability to clearly agree upon and publish documents related to areas of administrative responsibility would stand as a clear indication that the unique arrangement between two campus units that underlies this project is viable.

Discussions regarding an administrative group for the production phase are underway. The responsibilities of that group must be documented. This administrative group should be nimble to allow for quick decision-making regarding any issues, upgrades or changes within the IDEALS production level service. This group should be given the authority (both by the Library and by CITES) to make final decisions on behalf of both campus units, and be given the responsibility to report all decisions back to their units.

#### **2. Organizational viability**

IDEALS must have staff necessary to sustain its operations. The roles of these individuals and their relations to IDEALS must be clearly and publicly documented. At a minimum developing a production service must have:

- A project manager;
- A service manager;
- A programmer position dedicated to long term preservation planning and support;
- Preservation metadata production support from the University Library;
- Systems administration support from both CITES and the University Library's Systems Office; and
- Operational support that provides basic operational infrastructure and services.

#### **3. Financial sustainability**

IDEALS must be a financially sustainable operation. While efficiencies should be sought out at all levels, significant consideration should be given to "normalizing" the project budget as soon as possible, i.e., taking the annual funding allocated to the start-up phase and making it the base standing operating budget. Building trust in a project is much more difficult than building trust in a planned and supported service.

IDEALS must have a budget available upon request.

#### **4. Technological and procedural suitability**

All technological procedures and processes must be documented, tested, kept up-to-date, and publicly communicated via the IDEALS website. Documentation deemed inappropriate for



public dissemination will be made available to appropriate persons off-line or within a protected online environment.

#### **5. System security**

IDEALS must pay attention to issues of technological security on all levels. Support for inject-related technological security such as virus scanning, checksums and their validation, as well as a process for identifying authorized depositors must be instituted during the Pilot Phase.

Procedures will be designed to provide network and physical security for the data, including disaster recovery plans. These procedures will be documented in IDEALS System Security Procedures available at:

<<https://www.ideals.uiuc.edu/wiki/bin/view/IDEALS/SystemSecurityProcedures>> (IDEALS Staff Only).

#### **6. Procedural accountability**

All operations in IDEALS shall be documented. If appropriate, this documentation shall be publicly available via the IDEALS website. Otherwise, it shall be maintained behind an authenticated logon.

Items deposited in to IDEALS shall be assigned an audit trail intended to maintain an uninterrupted life cycle for the items deposited into the repository. This shall track changes to the object and its metadata as well as the name of the individual(s) that authorized or carried out the change.

### **Work to Occur in the Production Phase (beginning Summer 2007)**

The details below represent some of the attributes of a TDR that this working group feels should be implemented once IDEALS is a production service, in order to better align IDEALS with potential requirements for TDR certification.

#### **0. OAIS compliance**

As gaps are identified and prioritized in the pilot phase, production activities should be focused on determining how gaps between the IDEALS environment and the OAIS Reference Model should be filled and implementation begun.

Ongoing work will include continual reassessment of OAIS compliance as well as continual communication of this level of compliance to all IDEALS users (both local and remote).

#### **1. Administrative responsibility**

The documented responsibilities of the University Library and CITES within the IDEALS Initiative should be fully vetted and approved through the administrative structures of both units.

The IDEALS administrative group (for the production phase) should begin meeting and performing their duties. This group will need to make decisions regarding the ongoing goals and changes to the IDEALS Production Service, on behalf of both the University Library and CITES.

## **2. Organizational viability**

IDEALS must have dedicated resources necessary to complete the following ongoing activities:

- Ongoing maintenance and support;
- Periodic upgrades to server equipment and software;
- Funding necessary to support the acquisition of software and hardware necessary for preservation planning; and
- Funding necessary to support ongoing research and training for preservation planning.

## **3. Financial sustainability**

IDEALS must be allocated a yearly operating budget. This operating budget should take into account the following re-occurring costs:

- Personnel costs: Staff, management overhead, training, travel, workstation support;
- Technology costs: Equipment, software, contractual service or subscription fees, technology monitoring, server/software maintenance;
- Contingency costs: Unanticipated expenses associated with trigger events or catastrophes; and
- Overhead costs.

IDEALS must have budget information available upon request. As necessary, IDEALS may wish to look at ways to recover or absorb costs to help supplement its budget.

## **4. Technological and procedural suitability**

IDEALS must pursue technological and procedural options that are suitable to the long-term sustainability of the project. In some cases, this may mean not pursuing the most cutting edge technological solutions. But, it should mean that decisions will be made with long-term considerations in mind.

All technological procedures and processes must be documented, kept up-to-date, and publicly communicated (when deemed necessary).

Support for the preservation planning function in the OAIS model must be ongoing. As the repository develops and accumulates content, the ability of the repository to maintain the content entrusted to it will be directly related to this support.

Support for creation of preservation metadata suitable to the long-term maintenance of deposited content shall remain a key aspect of the project. Attention shall be paid to automated generation of metadata, but not at the expense of metadata that is actually suitable and appropriate. This metadata shall be sufficient both to describe the submissions and to encapsulate the objects (originals, derivatives, etc.) so that they can be displayed and maintained together.

## **5. System security**

IDEALS must guard against loss of both content and trust by making content availability a high priority. While preservation is of vital importance, access is a key feature of the repository. As part of the ongoing service, methods to ensure high availability (such as off-site mirroring) should be monitored and implemented as appropriate. All procedures to ensure system security must be implemented prior to the production phase.

The IR will be housed in CITES datacenters. CITES follows and promote the standards and recommendations exemplified in the National Incident Management System and NPFA 1600 for emergency management planning. CITES normally operates their datacenters following TIA 942 (Telecommunications Infrastructure Standards for Data Centers) between a Tier II and Tier III; and CITES system and network management follow nationally recommended best practices.

Archival storage for content shall be developed as soon as possible. This archival storage should be implemented in conjunction with decisions made by the University Library's Digital Preservation Management Program, so as to allow IDEALS archival storage to parallel, and hopefully merge with any archival storage solution implemented within the Digital Preservation Management Program.

## **6. Procedural accountability**

IDEALS shall maintain a level of transparency subject only to the security concerns of the project and associated operations. Documentation shall be maintained online, and public communication shall take the form of regularly updated Frequently Asked Questions (FAQ) and Glossaries.

Responsibility for activities within the program shall be assigned and clearly communicated among program staff and participants.

## **The Relationship between IDEALS and University Library Digital Preservation Management Program**

IDEALS should compliment other developments in Digital Preservation Management within the University Library. Where economies of scale may be realized, efforts should be made to take advantage of them. Such economies are most likely to be felt in areas such as support for applied research, the development of archival storage, and the development of server rooms appropriate to the needs of the equipment. However, the potential to develop standardized archival information packages and uniform metadata schema are among many other areas in which IDEALS and the Library's Digital Preservation Management activities may dovetail. The Library has begun to develop a digital preservation management program to fulfill the needs of its collection as they increasingly shift from being print-based resources to digital content. This time holds great potential for collaborative work.

In addition, IDEALS can work with research efforts undertaken within the University Library, the Graduate School of Library and Information Science (GSLIS), and CITES to leverage findings and to test additional preservation methods. The work undertaken through the National Digital Information Infrastructure and Preservation Program (NDIIPP) grant and the GSLIS data

curation grant from the Institute of Museum and Library Services (IMLS) provide important testing grounds for IDEALS. In addition, IDEALS should work with other campus as well as consortia groups (such as the CIC) to explore additional frameworks for digital preservation.

## **Conclusion**

IDEALS represents both collections and a set of services. The public face of these services is represented by access to the collections. Preservation is the most basic requirement for providing continued access to the resources with which they have been entrusted. In some ways, IDEALS initiative differs little from any museum, library, or archive. Promises that we will preserve content are not going to be enough as potential funding agencies become more aware of the fragility of digital content. Indeed, with efforts such as the Federal Research Public Access Act (FRPAA) of 2006<sup>3</sup>, evidence of a commitment to a digital preservation program is likely to become a metric for the disbursement of grant funds. In order to ensure the long-term success of IDEALS as a repository for locally developed research and scholarship, the IDEALS Preservation Working Group believes the IDEALS initiative should pursue the TDR model as a means of ensuring the preservation of its content and building trust among the faculty and scholars at the University of Illinois at Urbana-Champaign. An established digital preservation management program – particularly one that meets the attributes and certification requirements of a Trusted Digital Repository – will provide evidence of our commitment both to funders and to our faculty and scholars.

---

<sup>3</sup> FRPAA has the support of UIUC Provost Linda Katehi. See <http://www.cic.uiuc.edu/groups/CICMembers/archive/documents/FRPAAletterFinal7-24-06.pdf>.