

The Cornell University Central Depository and Project Harvest

By Nancy Y. McGovern, Digital Preservation Officer (DPO)
March 2002

Background

The Central Depository is a joint initiative between the Digital Preservation Officer (DPO) Instruction, Research, and Information Services (IRIS) and Digital Library and Information Technology (DLIT) unit of Cornell University Library (CUL). The goal of the Central Depository is to provide centralized control over Cornell's growing digital collections, to provide a cohesive and comprehensive organizational and policy framework for the management of digital collections, and to develop systematic modules that will integrate current and future digital collections that are currently stored and accessed in a distributed manner.

The first phase 1 of the Central Depository produced a report entitled: *Establishing a Central Depository for Preserving Digital Image Collections - Part I: Responsibilities of Transferee* (see Relevant Sources section for the full citation). The second phase is producing a report on the responsibilities of the depository. That report includes the status of relevant Cornell initiatives and developments.

Project Harvest has been carried out within the context of the Central Depository. The starting point for the Central Depository was intended to be digital images, but the ultimate objective is to develop a depository that can ingest, preserve, and provide long-term access to all digital formats that are created and/or maintained by Cornell, including electronic journals in XML, PDF or other formats.

Several national and international initiatives have recently produced a series foundational digital archives documents. Cornell has been actively involved in the development of several of these documents. In developing the Central Depository, Cornell has embraced the following framework and standard documents:

Organizational context, responsibilities, and infrastructure

Title: *Attributes of a Trusted Digital Repository*
Source: RLG-OCLC collaboration
Date: Draft, October 2001
CU role: Anne R. Kenney serves on the committee that is developing this document
URL: <http://www.rlg.org/longterm/attributes01.pdf>
Notes: The DPO produced a document that maps the framework proposed in this document to the Central Depository requirements. This document was submitted to RLG. See *Central Depository II: Mapping the RLG/OCLC Attributes of a Trusted Repository to CDI key issues* [draft report]

Open Archival Information System (OAIS) and related initiatives

Title: *OAIS Reference Model*
Source: Consultative Committee for Space Data Systems (CCSDS)
Date: version CCDS 650.0-R-2, July 2001
URL: <http://www.ccsds.org/documents/pdf/CCSDS-650.0-R-2.pdf>
Notes: OAIS is in review by ISO. The DPO produced a document that maps Cornell initiatives and development to the OAIS reference model

Submission Information Package (SIP) Transfer Issues:

Title: *Producer-Archive Interface Methodology Abstract Standard*
Source: CCSDS
Date: January 2002
URL: <http://ssdoo.gsfc.nasa.gov/nost/isoas/CCSDS-651.0-W-1.pdf>
Notes: William Kehoe of DLIT produced a document that maps the proposed methodology to the Central Depository. See *A Draft Of The Central Depository Ingest Process*.

Archival Information Package (AIP) Components:

Title: *Content Information* (draft)
Preservation Description Information (draft)
Source: OCLC/RLG Preservation Metadata Working Group (PMWG)
Date: Final drafts October 2001, March 2002
CU Role: Oya Rieger served on the PMWG
URL: <http://www.oclc.org/research/pmwg/>

Format Issues:

Title: *Data Dictionary: Technical Metadata for Digital Still Images* (draft standard)
Source: National Information Standards Organization (NISO), Technical Metadata for Digital Still Images, Standards Committee AU
Date: February 2001
CU Role: Oya Rieger co-chaired the committee
URL: http://www.niso.org/committees/committee_au.html

Cornell's OAIS Planning and Implementation

The OAIS Reference Model has become the core digital archives planning document. Part of the Central Depository work has been to map current and future projects and developments to the OAIS components and functions. Cornell has been a leader in digital imaging and preservation. CUL has been creating digital collections for over a decade. The Relevant Sources section of this document provides links to some of the most significant digital information management initiatives, including efforts to consolidate existing collections through the Cornell Library Digital Collections Web site and Cornell University Image Collection, a Cornell Institute for Digital Collections (CIDC) collaborative project with Luna Insight.

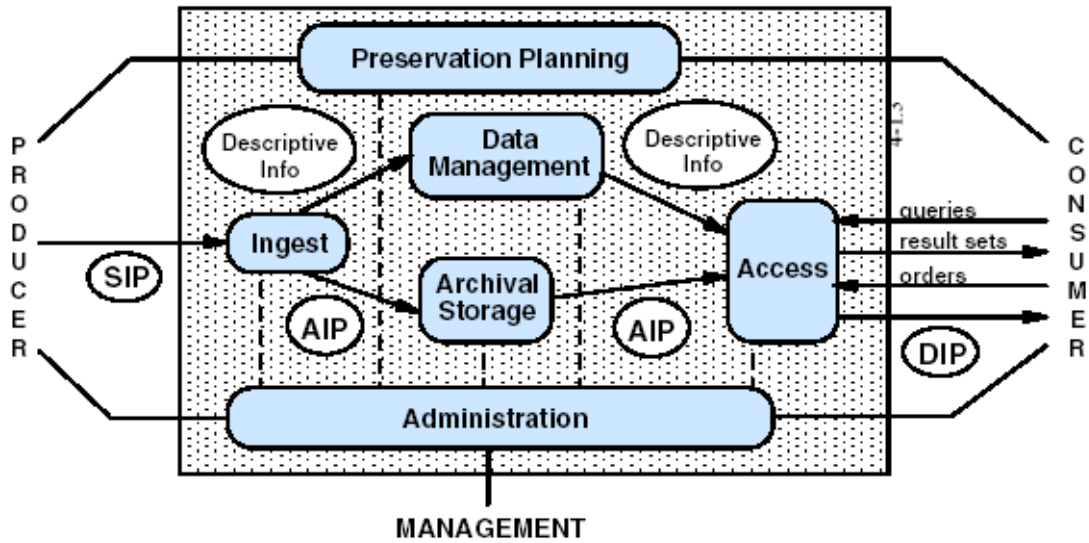


Figure 4-1: OAIS Functional Entities

The above high-level diagram from the OAIS Reference Model document has become very familiar within the digital preservation community. The documents discussed in the previous section are filling in the necessary implementation-level specifications. The diagram below summarizes relevant OAIS developments at Cornell.

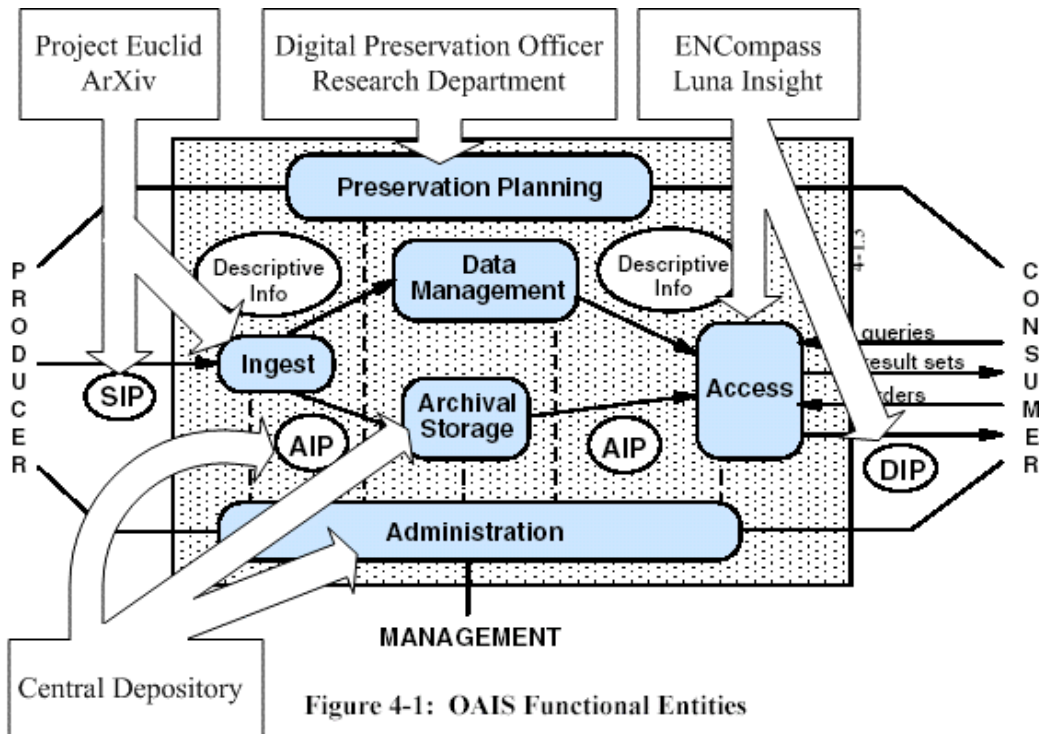
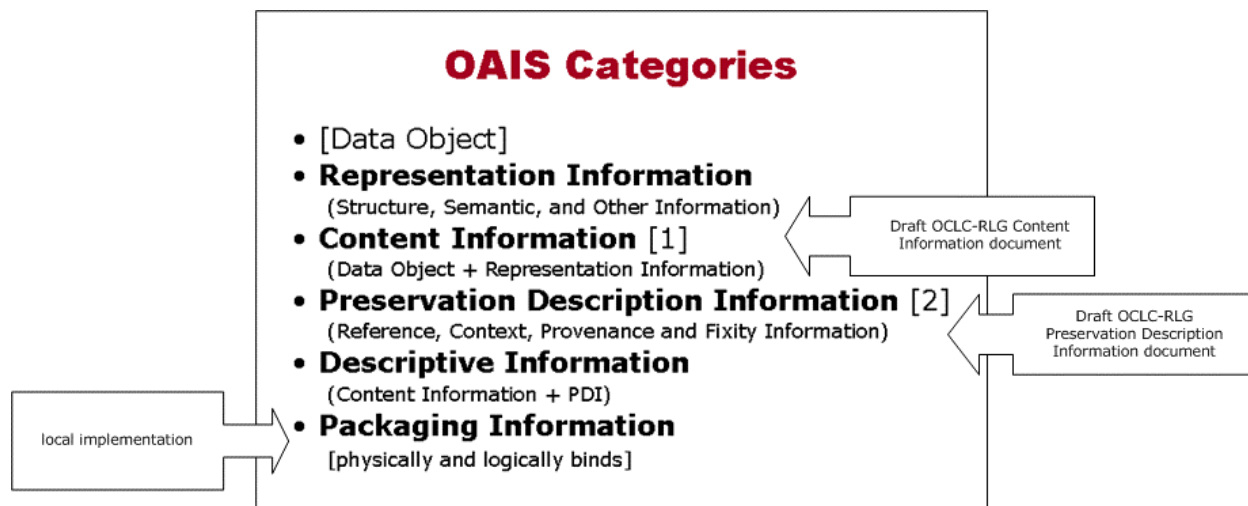


Figure 4-1: OAIS Functional Entities

Function/Entity	Initiative(s) at Cornell
Pre-ingest	<ul style="list-style-type: none"> • Create objects in accordance with standards, policies and established procedures and practices • Work with central depository to transform existing collections/objects to meet requirements
Registration	<ul style="list-style-type: none"> • Establishing registration module/process to track digital objects from project inception/object creation to ingest, storage and use. DLIT/DPO collaborative project April 2002-January 2003 • Potential: automate registration module for remote monitoring of objects (see Project Prism work); integrate module into digital archive implementation
Creation	<ul style="list-style-type: none"> • Images: meet imaging requirements; follow Quality Control recommendations - source: Central Depository I (CDI) report (recommended and minimal requirements) • Other formats: establish benchmarks and QC procedures. Tracking, developing benchmarks for GIS, video, text, etc. • Tracking general and type-specific standards developments • Potential: automate QC process; develop creation toolkits
Selection	<ul style="list-style-type: none"> • Developing integrated selection process for Central Depository. Source: CDI report, collection development policies and procedures (http://www.library.cornell.edu/colldev/collectiondevelopment.html) • Developing selection decision tree for the products of the Digitization Service, existing collections, and future collections • Using Vet School media project as test case for selection decisions
Monitoring/Evaluation	<ul style="list-style-type: none"> • Developing a preservation risk management program (policies, procedures, tools) for monitoring Web-accessible resources as part of Project Prism
Transfer	<ul style="list-style-type: none"> • Establishing transfer requirements that covers formats, storage media, metadata, and documentation (type-specific). Examples: NARA, Australia • Tracking preservation management, standards and technology developments to maintain current and responsive transfer requirements

Ingest	<ul style="list-style-type: none"> • Establishing ingest procedure (manual or automated) that includes QC checking • Evaluating ingest process and Submission Information Package (SIP) of the ArXiv, Project Euclid and other initiatives for the most appropriate foundation upon which to build the ingest module • Potential: use automated scheduling/monitoring to identify collections for ingest using registration module tracking information; develop automatic ingest as last step in creation
Archival Storage	<ul style="list-style-type: none"> • Exploring Archival Storage options to leverage current developments; evaluating Project Euclid as a possible basis for the Archival Storage module; tracking Fedora • Managed storage/access projects: Cornell Library Digital Collections, Cornell University Image Collection (Luna Insight) • See the AIP section below
Access	<ul style="list-style-type: none"> • Developing ENCompass to provide access to digital collection objects; the ENCompass team is developing delivery criteria for text, images, and other object types that will become the basis for the Delivery Information Package (DIP) • Providing access to digital image collections through the Cornell University Image Collection Web site; ongoing development is expanding user options • Evaluating the ArXiv and Project Euclid access interfaces and delivery protocols
Performance measures	<ul style="list-style-type: none"> • Developing policies, procedures, and action plans for integrating performance measurement into existing and future digital collection management based upon work completed to comply with ARL requirements
Preservation Planning	<ul style="list-style-type: none"> • Coordinating preservation planning and providing a policy framework for digital preservation through the work of the DPO with the Central Depository as the focal point for digital preservation initiatives • Continuing to investigate preservation approaches through the work of the Digital Imaging and Preservation Policy Research unit/Research Department • Investigating automated policy enforcement through Project Prism • Potential: develop and integrate automated preservation planning tools; contributions from Project Prism and other research and development projects
Administration	<ul style="list-style-type: none"> • Developing centralized support for administration functions through Central Depository planning • Consolidating developments of administration functions for the ArXiv, Project Euclid, and other projects

Archival Information Package (AIP)



The Central Depository is using the *Content Information* and *Preservation Description Information* draft documents to model the Archival Information Package for the Central Depository. As a result of discussions with the Central Depository project, the ENCompass team prepared a document for submission to Endeavor to determine if ENCompass can accommodate OAIS design elements.

Relevant Sources

ArXiv

<http://arxiv.org/>

Cornell Library Digital Collections

<http://cdl.library.cornell.edu/>

Cornell University Image Collection (Luna Insight)

<http://insight.library.cornell.edu/insightbrowser2/launcher.asp>

Project Euclid

<http://projecteuclid.org/>

Project Harvest

<http://www.library.cornell.edu/harvest/>

Project Prism

<http://www.library.cornell.edu/preservation/prism.html>

USDA Economics and Statistics System

<http://usda.mannlib.cornell.edu/>

“Preservation Risk Management for Web Resources: Virtual Remote Control in Cornell's Project Prism,” by Anne R. Kenney, Nancy Y. McGovern, Peter Botticelli, Richard Entlich, Carl Lagoze, and Sandra Payette, *Dlib Magazine*, January 2002.

<http://www.dlib.org/dlib/january02/kenney/01kenney.html>

Preserving Cornell's Digital Image Collections: Implementing an Archival Strategy, March 2001
Establishing a Central Depository for Preserving Digital Image Collections - Part I: Responsibilities of Transferee

<http://www.library.cornell.edu/imls/index.htm>

Risk Management of Digital Information: A File Format Investigation
<http://www.clir.org/pubs/abstract/pub93abst.html>

Central Depository Phase 1

*Establishing a Central Depository for Preserving Digital Image Collections - Part I:
Responsibilities of Transferee*

Report Outline:

1. Selection and Content Considerations
 - A. Scope
 - B. Content/Functional Criteria
 - C. Priorities for Deposit
2. Legal Considerations
3. Technical Requirements for Conversion
 - A. Source Material for Digitization
 - B. Recommended and Minimal Requirements
 - C. Quality Control
4. Pre-Depository Storage and Maintenance
5. Metadata Requirements
 - A. Descriptive
 - B. Structural Metadata
 - C. Preservation Metadata

Key Issues for Central Depository Part 2

Next Steps