



Library of Congress

Programs & Strategies for Long-Term Usability of Digital Collections

Beth Dulabahn
Library of Congress

Cultural Heritage Online
December 14, 2006

Overview

- National Digital Information Infrastructure & Preservation Program (NDIIPP)
- Activities within the Library of Congress
- Looking ahead – a Network of Partners

NDIIPP Background



- Created by federal legislation in December 2000
- Up to \$100 million
- *Preserving Our Digital Heritage* NDIIPP plan approved December 2002
- Program officially concludes in 2010

NDIIPP Vision



To ensure access over time to a rich body of digital content through the establishment of a national network of committed partners

Three Areas of Focus

- Network of collecting & preserving institutions
- Technical architecture
- Research

Initial Call for Proposals



- Emphasized collection & network building
- 8 awards
- 36 institutions
- Variety of content areas
 - Public television
 - U.S. Southern culture
 - Government & political web sites
 - Social science data
 - Geospatial data
 - Dot.com era business records

Preservation Research Grants

- ✎ Joint Library of Congress / National Science Foundation program
- ✎ Funded 10 grants across different areas, e.g.
 - ✎ Audit trails for documents in file systems
 - ✎ Large-scale data repositories for oceanographic study data and MODIS satellite images
 - ✎ Engineering and design data
 - ✎ Infrastructure and lifecycle issues, including data provenance and versioning
 - ✎ Incentives for data producers to create archive-ready datasets
 - ✎ Video
- ✎ <http://www.digitalpreservation.gov/partners/research.html>

Archive Ingest & Handling Test



- Harvard, Stanford, Johns Hopkins, and Old Dominion universities
- 12-gigabyte test collection
 - 57,000 files
 - 20 file types

Digitalpreservation.gov/technical/aiht.html

Business Models



- Portico
- CLOCKSS

Copyright



- Section 108 Study Group
 - Making recommendations on exceptions for libraries and archives to collect, preserve, and serve digital content
 - Report by mid-2007

New NDIIPP Initiatives



- States
- Technical Architecture
- Preserving Creative America

States Initiative



- Workshops in 2005
- Call for proposals 2006
- Multi-state demonstration projects
 - Partnerships
 - Distributed responsibilities
 - Shared technologies

Technical architecture



- 4 critical areas
 - Storage
 - Content exchange protocols
 - Preservation tools & services
 - Metrics & standards for content preservation

Preserving Creative America



- Planning session in Hollywood April 2006
- Targeted movies, recorded sound, photographs & digital art, video & computer games
- Seeking areas of mutual interest & benefit
- Awards within next few weeks

Activities Within LC



- Copyright e-Deposit
- National Digital Newspaper Program
- Web Harvesting
- Standards

Copyright E-Deposit

- Re-evaluation of “Best Edition” policy
- E-Deposit pilot
 - Starting with E-Journals
 - Developing prototype repository
 - Will move on to other content types

National Digital Newspaper Program (NDNP)

- Cooperative effort with National Endowment for the Humanities
- In pilot phase
- Developed JPEG2000 profile for newspapers
- METS schema

Web Harvesting



- Began in 2000
- Most collecting has been event-based:
 - U.S. elections
 - Sept. 11
 - Iraq War
- Recent pilot to look at other collecting areas
- Charter member of International Internet Preservation Consortium (IIPC)

Standards



- METS profiles
- PREMIS & XML schemas

Building a Network

- Encouraging collaboration
 - Funding incentives
 - Opportunities to meet
- Initial network of 36 institutions now numbers 67
- A glimpse of future possibilities

Network Design Principles



- Flexibility and extensibility, scaling up to serve large institutions and down to serve preserving institutions at the local level
- Collaboration and cooperation of member institutions, with defined roles and responsibilities for all participants
- Transparency and accountability
- Nurturing of experimentation in service of innovation, new knowledge, and improved practices
- Oriented to problem solving in an iterative process
- Respectful of the heterogeneous needs of member institutions and their autonomy in decision-making at the institutional level with respect to mission-critical areas such as collection development, rights management, and economic sustainability.

Layers of a Network

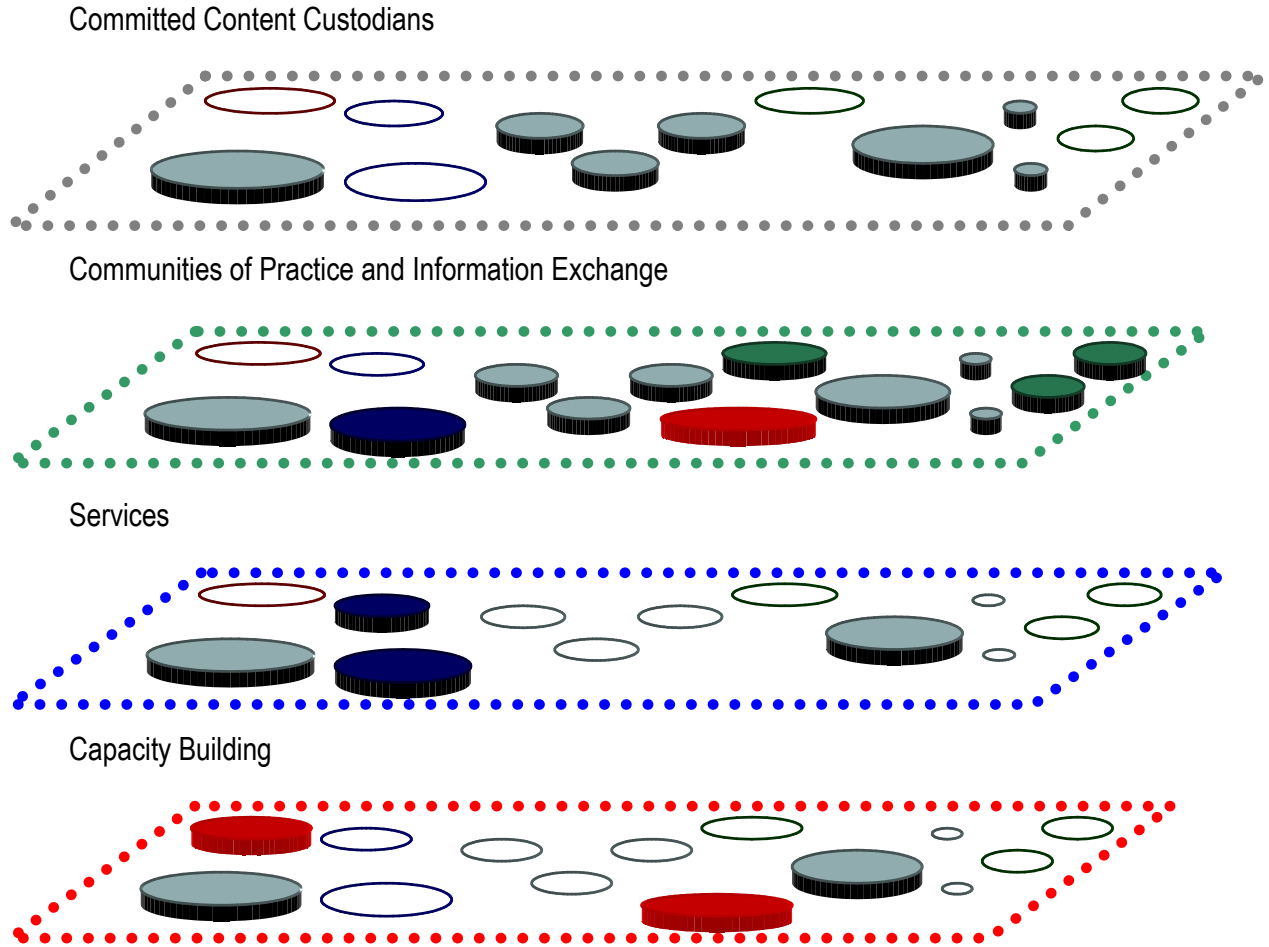


Figure 1: Layers in a (strawman) stewardship network

Building a Future Network



- Key questions
 - Governance
 - Formality
 - Funding
 - Incentives

Network of Networks



- NSF Cyberinfrastructure Program
- Cyberinfrastructure for the Humanities & Social Science
- Other content communities
 - Social science datasets
 - Geospatial
 - U.S. Southern culture

The screenshot shows a web browser window displaying the homepage of the Library of Congress Digital Preservation Program. The browser's address bar shows the URL "Digital Preservation (Library of Congress)". The page header includes the text "The LIBRARY of CONGRESS DIGITAL PRESERVATION" and a search box with a "SEARCH" button and a link to "More Search Options". Below the header, a breadcrumb trail reads "The Library of Congress > Digital Preservation".

The main content area is divided into several sections:

- Digital Preservation Home**: A list of navigation links: [About the Program](#), [Technical Infrastructure](#), [Partnerships](#), [News, Events](#), and [Contact](#).
- Related Resources at the Library**: A list of external links: [Standards at the Library of Congress](#), [Digital Collections & Programs](#), [U.S. Copyright Office](#), and [Webcasts from the Library of Congress](#).
- Digital Preservation**: A section with the subtitle "The National Digital Information Infrastructure and Preservation Program". It contains a paragraph of introductory text, a link to [Learn more about the Digital Preservation Program.](#), and a **Highlights** section. The highlights include: "The inaugural meeting of the Section 108 Study Group was held at the Library on April 14-15, 2005." and a link to [Visit the Section 108 Study Group Web site.](#)
- 2004 OSI Annual Review**: A link to a PDF document titled "2004 OSI Annual Review" (8.2 Mb), accompanied by a PDF icon.
- More News and Events**: A link to further information.

The footer of the page contains links for [Site Map](#), [Contact](#), [Accessibility](#), and [Legal](#).