

The National Digital Information Infrastructure and Preservation Program (NDIIPP)

Digital Preservation Coalition

London, UK

Laura E. Campbell

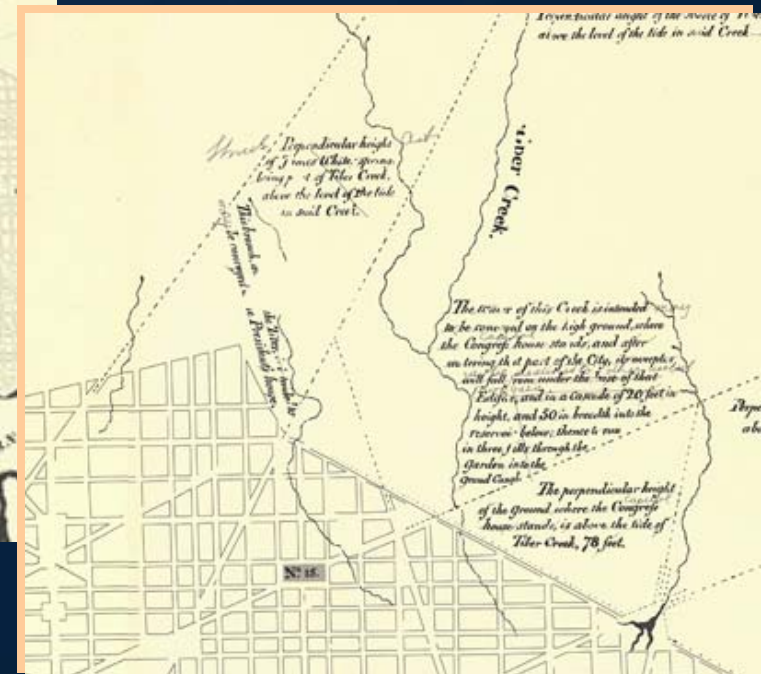
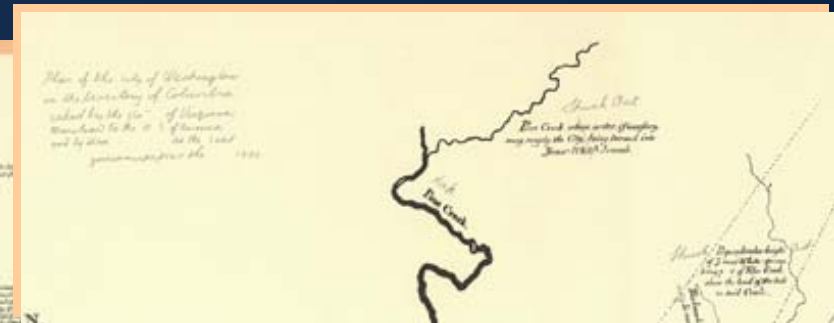
Associate Librarian for Strategic Initiatives

Library of Congress

June 23, 2004



June 2004

[illegible]

EDWARD S. CURTIS'S THE NORTH AMERICAN INDIAN

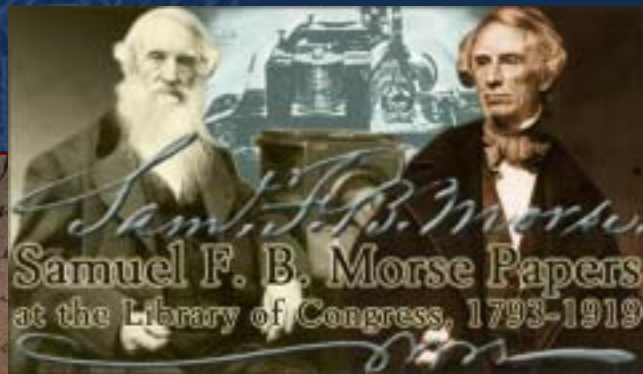
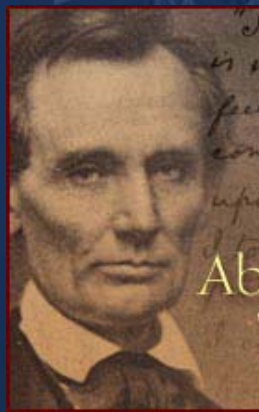


P H O T O G R A P H I C I M A G E S



Our Nation's History and Culture

American Memory



Abraham Lincoln
H A H
at the Library



The Library of Congress

American Memory

Historical Collections for the National Digital Library

Collection Finder
Select collections to search

Search
Search for items across all collections

Learning Page
Teaching and learning with American Memory

Today in History
June 18

[What's New](#)

[FAQs](#)

[How To View](#)

[Copyright & Restrictions](#)

[Technical Information](#)

[Chat with a Librarian](#)
M-F, 2-4 PM, EST

American Memory is a gateway to rich primary source materials relating to the history and culture of the United States. The site offers more than 7 million digital items from more than 100 historical collections.

Featured Collection

America from the Great Depression to World War II

Photographs from the ESA-OWI, 1935-1945

Search example of the day: [steamboats](#)

<http://memory.loc.gov/>

Born Digital "At-Risk" Web Sites



<http://www.loc.gov/minerva/collect/elec2000>



<http://www.loc.gov/minerva/collect/sept11>

NDIIPP Goal

To develop a nationwide collection and preservation strategy for digital materials in cooperation with the information and technology industry, concerned federal agencies, libraries, research institutions and not-for-profit entities.

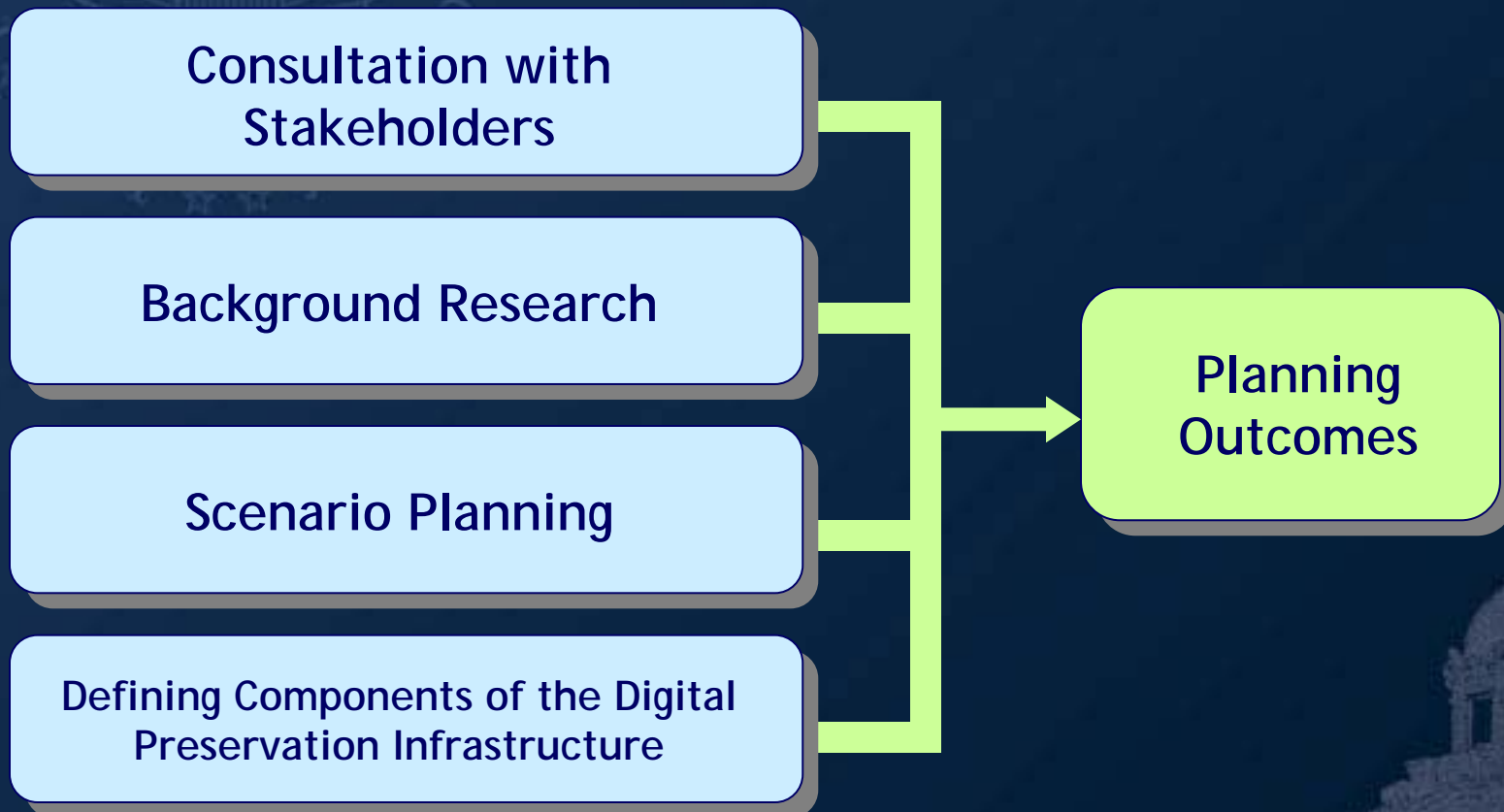
NDIIPP Vision

To ensure the long term access to a rich body of digital content through the establishment of a network of partners, collaborating in a digital preservation architecture with defined roles and responsibilities.

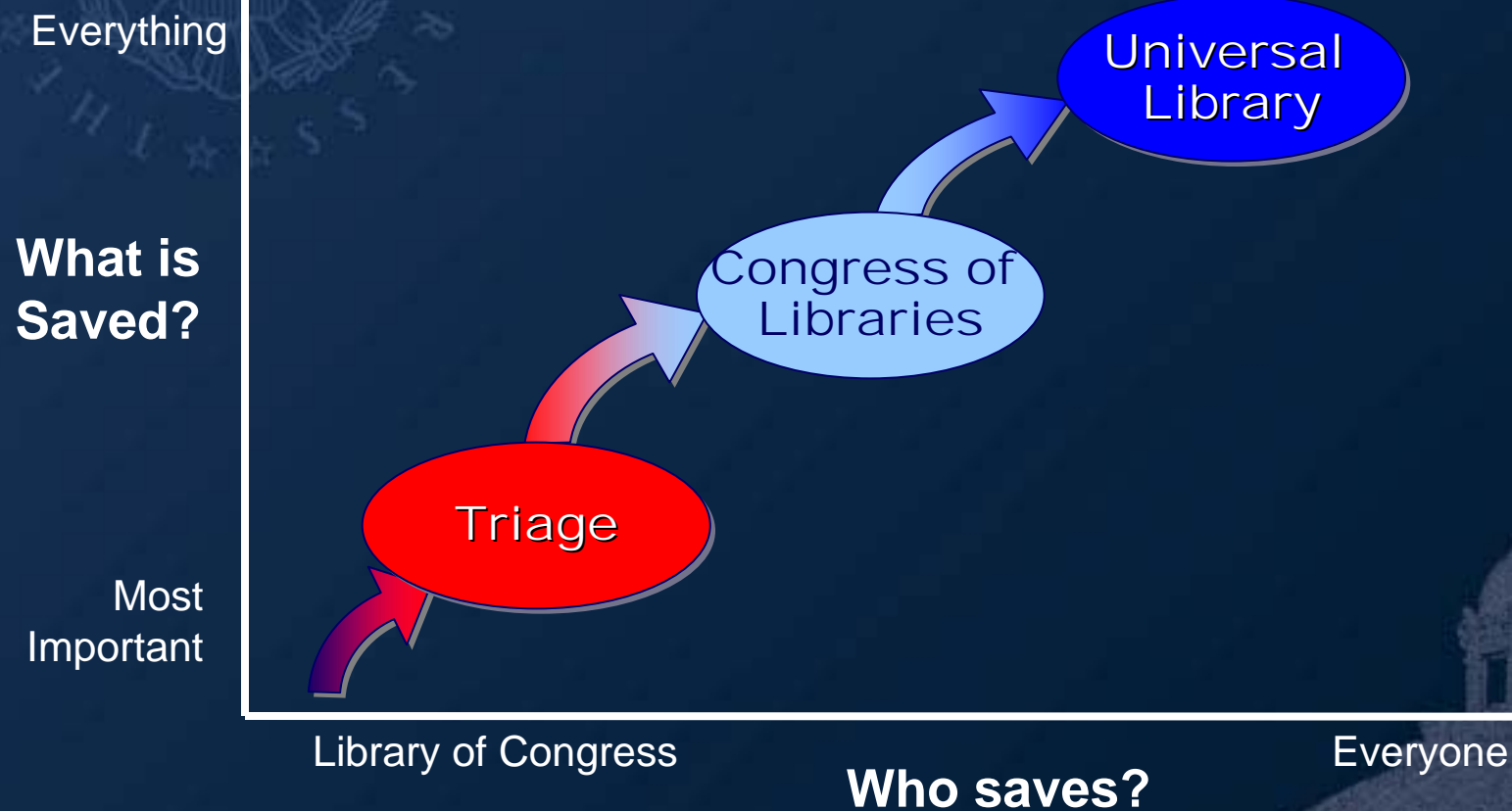
NDIIPP Legislation and Funding

- Created by federal legislation (PL 106-554) in December 2000
- Up to \$175 million potentially available
 - \$5 million (available immediately until expended)
 - \$20 million (subject to advance Congressional approval)
 - \$75 million (subject to \$ for \$ match from non-federal sources)
 - \$75 million (private funds)

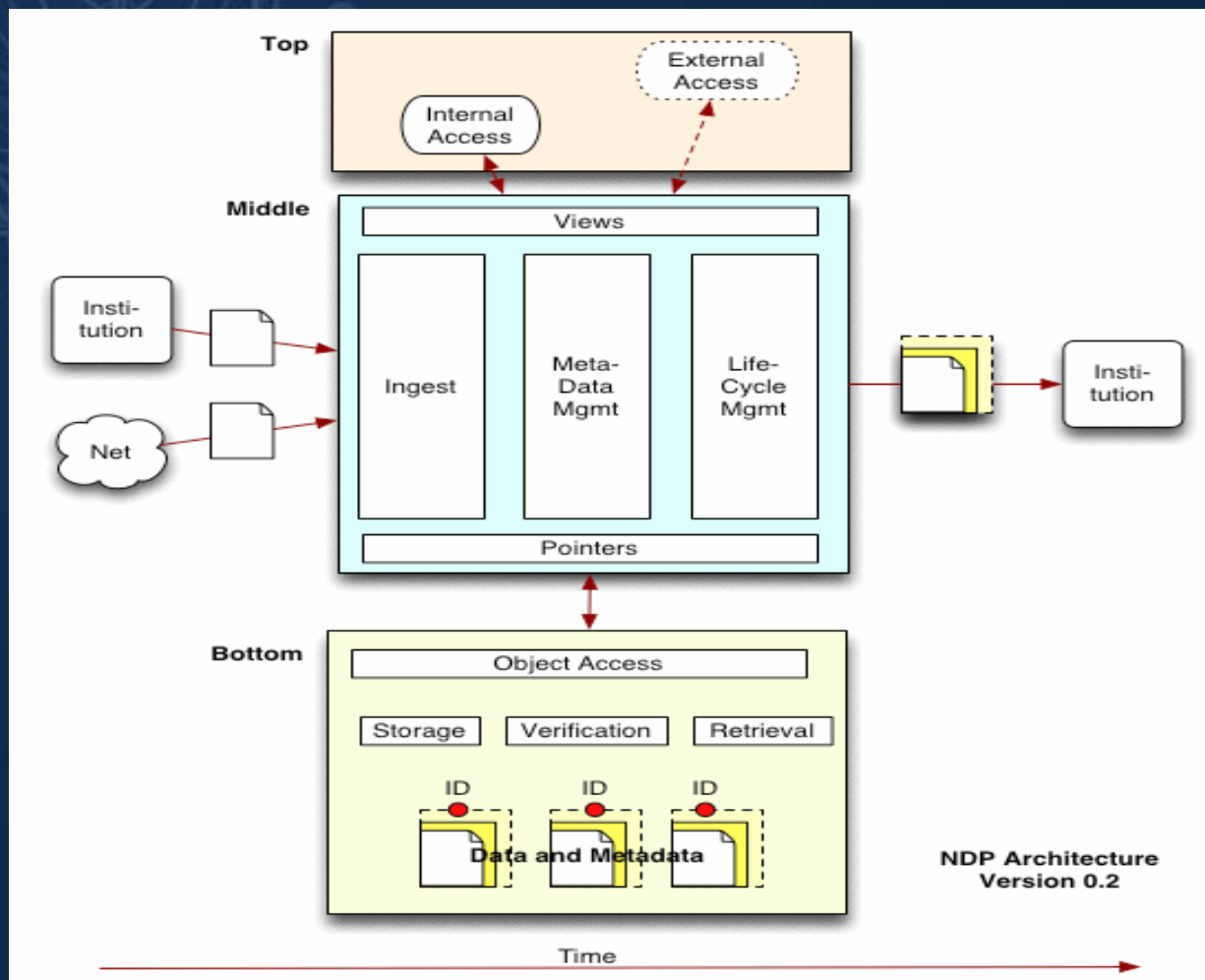
Listening and Learning Planning Phase, 2001-2002



Scenario Planning Plausible Scenarios



NDP Architecture Version 0.2



Facilitated NSF Workshop

- 15 federal agencies, corporate labs and computer scientists
- Areas of Focus
 - Core technical infrastructure
 - Preservation technologies and tools
 - Access to and use of preserved content
 - Basic computer science issues for large-scale storage systems

Major Lessons

- Distributed and decentralized approach
- Need for research into new tools and technologies
- Recognition that technology is only part of the problem
- Great concern for IP issues
- Need to create a balance between preservation and access
- An infrastructure has two major components:
 - A preservation network of partners
 - A technical architecture

3 Areas of Investment

- Network of partners
- Preservation architecture
- Basic research

"At-Risk" Content

- 12 Terabytes
- Selected by curators and subject experts
- Major events
- Specific subject areas
- Material for use by Congress

Acquisitions/Capture Goal

- Web sites
- Web-casts
- e-journals
- TV and radio broadcasts
- Music
- Sound recordings
- Maps

Requires Handling

- Audio
- Text
- Databases
- Video
- Still and moving images
- Graphics
- Geo-spatial materials

Partnerships

- First NDIIPP Solicitation for Partners
- 22 eligible proposals
- Lead Institution with at least 1 other partner
- 2 peer review panels
- 8 recommended

8 Projects with 25 Participating Institutions

- Universities
- State governments
- Private companies
- TV broadcasters
- Law firms
- Non-profit archival organizations

Diverse and Challenging Content

- Social science data sets
- News broadcasting
- Geo-spatial material
- TV programming
- Business records
- Government and political web sites
- Historical multimedial collections

3 Year Period of Performance

- Content selection and collection
- IP issues
- Content transfer and retention
- Costs

International Partnerships

- Digital Preservation Coalition in the UK
- International Internet Preservation Consortium (IIPC)
 - 12 national libraries
 - Technical specifications for an open source web crawler
 - 5 working groups

International Internet Preservation Consortium (IIPC)

- France
- Italy
- Denmark
- Finland
- Iceland
- Canada
- Norway
- Australia
- Sweden
- United Kingdom
- Internet Archive, USA
- Library of Congress, USA

Business Model Partnerships

- Service bureau for archiving foreign news broadcasts
- Subscription and archiving service for e-journals
- Service provider for long-term archiving of large volume multimedial content

Technical Partnerships

- Repository development with NASA/Langley for video materials
- Prototyping development with Los Alamos National Lab for:
 - e-journals
 - Web sites materials, and
 - Multimedial collections

Next Round of Partnerships 2005

- State and local government content
- Corporate technology and computer science laboratories

Preservation Architecture

- In 2003, conducted two workshops with 47 technical experts
- Revised NDIIPP architecture

Architectural Design Principles

- Modular approach
- Minimal requirements at each level
- Implementation-agnostic
- The system should *never* be optimized
- Survive the first migration
- Piggyback on existing work
 - OAIS/W3C/IETF/Digital Libraries

Connection Overview

- Import/Export
- Access/Storage
- These matter more than any instantiation
- Must survive changes of hardware
- Must survive element design changes

Goals of the Archive Ingest and Handling Test (AIHT)

- 4 institutions and 6 existing architectures
- Test of ingest, export, and transfer among architectures
- Identify gaps in functionality
- Focus on protocols and metadata investments

The National Science Foundation (NSF) Call for Research Goals

- Technical Architectures for Archival Repositories
- Attributes of Archival Collections
- Digital Archiving Tools and Technologies
- Organizational, Economic, and Policy Issues

Research Focus Areas

1. Technical Architectures

- Data model driven architecture
- Archives of derived data products
- Archives of temporally changing data
- Archives of evolving data
- Controlled access repositories
- Repurposed archives
- Repository architectures

Research Focus Areas

2. Attributes of Archival Collections

- Preservation of static digital objects
- Preservation of complex digital objects
- Aggregation of items and objectives into collections
- Decision models

Research Focus Areas

3. Digital Archiving Tools and Technologies

- Acquisition and ingest
- Naming and authorization
- Standards and interoperability

Research Focus Areas

4. Organizational, Economic, and Policy Issues

- Metrics
- Incentives for long-term preservation of digital information
- Incentives for deposit of digital content

NSF Call Available

http://www.cise.nsf.gov/funding/pgm_display.cfm?pub_id=13106&div=iis



Due September 14, 2004

5 Year Outcome

- Distributed network of collecting and preserving partners
- Up to 100 terabytes of “at risk” content
- Topics include: American Law, Domestic Social Policy, Foreign Affairs, Defense and Trade, Government and Finance, Science and Technology
- Variety of formats: text, numeric, video, images, audio, and geo-spatial

5 Year Outcome

- Multiple sources: TV and radio broadcasts, electronic journals, private archives, donations and the Web
- Preservation standards for digital content
- Governing body for NDIIPP
- Recommendations to Congress for future funding and legal options for NDIIPP

Digital Universe

