

New Directions in Digital Government Research

NSF Workshop on Long-Term Preservation of
Digital Objects

dg.o 2002

Los Angeles, CA

May 20, 2002

Margaret Hedstrom, School of Information, University of
Michigan

Research Issues in Long-Term Preservation

- Background and Motivation
- Current Environment & Opportunities
- NSF Workshop Summary
- Initial Results
- Next Steps

Digital Archiving Challenges

- Technology obsolescence
- Very large quantities of digital information that may be worth preserving for reuse
- Heterogeneous and complex digital resources
- Lack of scalable methods for preservation
- Economic models

What is at stake?

- Investments in conversion of information to digital form
- Preservation of an increasing share of information that is “born-digital”
- Ability to provide continuing access to information
- Ability to support new analyses or reuse of digital information

Background and Motivation

- 10 + years of concern about research on digital archives and long-term preservation
- Many small projects with a focus on immediate or short-term problems
- Interest in digital archiving as part of Digital Library and ITR Research
- Convergence of concerns around scientific data repositories, e-government and e-records, and preservation of digital cultural heritage

Recent Interest in Digital Archiving

- LC 21 and National Digital Information Infrastructure and Preservation Program
- Open Archival Information System Reference Model (OAIS)
- NARA funding for research at SDSC NSF
- Digital Libraries and National Deposit Libraries
- Many government agencies with preservation concerns

Current Environment -- What is New?

Stakeholder interest

Lots of activity to build on

Lots of resources that are worth preserving

Potential for research funding

Opportunity

Need fresh blood and new ideas

NSF Workshop on Long-Term Preservation of Digital Objects

Sponsors

- NSF Digital Government Program
- NSF Information and Intelligent Systems Division
- Library of Congress, National Digital Information Infrastructure and Preservation Program

Airlie Center, Warrenton, VA

April 12-13, 2002

www.si.umich.edu/digarch/

Goals for the Digital Archiving Workshop

- Bring together stakeholders and researchers
- Conversation across communities (CS, IS, Archives, DL, domain specialists)
- Identify priority research areas for next 3 to 5 years
- Generate a compelling research agenda
- Mobilize funding
- Engage researchers in digital archiving issues

Participants

- University faculty and researchers from computer science, information science, archival science, etc.
- Researchers and developers from industry (IBM, Sun, Microsoft, HP, HD Storage Association, RAND, OCLC, RLG)
- Government agencies (LOC, NLM, NAL, NARA, NASA, DOD, CIA, NCHS, NIST)

Workshop Process

Day 1

- Opening Context Setting
- Breakout Discussions on:
 - Architectures for Repositories
 - Attributes of Archives Collections
 - Policy and Economic Models
 - Tools and Technology
- Synthesis and Summary

Day 2

- Reactions to Day One
- Breakout Discussions
 - What constitutes an infrastructure for long-term preservation?
 - What are the priorities for research?
 - What long-term scenarios may influence research priorities?
 - How can we translate research results into practical applications?
- Summary/Wrap up

Some Early Conclusions

- Driving focus for preservation research is:
 - What is unique about long-term preservation?
 - What is the difference between a digital archive and data warehouse?
 - Between digital archive and digital library?
 - Between a digital archive are material stored or forgotten somewhere?
- Focus on unique challenges
- Identify commonalties and build on them

What is different about digital archiving?

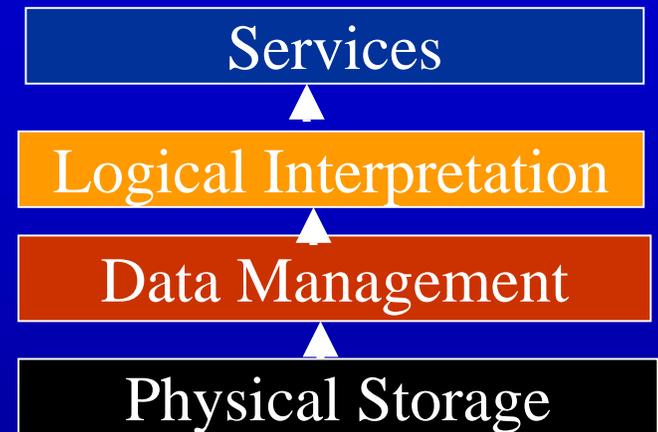
- Long-term
- Threat of interrupted management
- Funding and utility models
 - intergenerational beneficiaries
 - changes in the designated community
- Authenticity/Integrity when records are not maintained in the original state
- Measures of success are long in the future

Research Challenges

- Architectures for Repositories
- Attributes of Archived Collections
- Tools and Technology
- Economic and Policy Models

Architectures for Repositories

- Strong consensus on layered models that separate physical storage from data management, logical interpretation, and services



- Need models for processes and services built on top of bit storage
 - Processes/Business rules for curation
 - Define services that meet the specific requirements of different communities
- Different requirements imply different architectures
 - End-to-end within a single institution
 - Highly distributed

Attributes of Archived Collections

- Archived collections are created through value-adding processes
 - selection
 - organization
 - description
 - quality control
 - stewardship
- Need models for a spectrum of collections and services from storing bits to delivering highly refined products

Tools and Technology

- Human intervention is the expensive part of archiving
- Need tools and technology to automate processes, but we need a clearer definition of what those processes are
- Standards -- how do we predict which standards are Gold Standards

Tools and Technology

- Persistent Naming and Authentication
- Decision Models
- Automated Ingest
- Interoperability/Standards
- Technology for Preservation

Policy and Economic Models

- Political Economy of Public Goods
- Incentives
 - to deposit
 - to preserve
 - play with incentives and constraints and model impacts
- Decision Models
- Costs
- Who does archiving and who pays?

Priority areas for research

- Reference architecture for research that supports:
 - focused research at each layer of the architecture
 - research on how to assemble the stack
- Metrics (costs, value, policy options, outcomes)
- Preservation methods for dynamic objects
- Decision models
- Scalability up and down

Priority areas for research

- Tool development
 - automated ingest
 - metadata capture and management
 - push vs. pull
- Predictive models
 - User requirements
 - Technology evolution

Research Strategies

- Theory-building
- Exploratory
- Simulations
- Experimental
- Observational
- Testbeds

Next Steps

- Draft Report
- Review and Discussion
- Final Report to NSF & LOC -- early summer
- Mobilization of funding
- Input into NSF Calls for Proposals
- Engage researchers in research projects