



MIGRATION AS A PRESERVATION STRATEGY

Fynnette Eaton

Digital Reality II

Monday, June 5, 2000

Outline of talk

- Discuss terminology
- Describe evolution of National Archives procedures for making preservation copies of electronic records
- Briefly outline three current research projects that are focusing on the migration of electronic/digital records over time

Terminology

■ Archives Vs Archiving

- Archives, a noun, which means the records of an organization or the institution responsible for preservation of records
- Archiving, a verb, which describes the process of moving active electronic records to an inactive status, without any judgement as to the value of the information being moved

Terminology

- Refreshing/recopying

- the practice of moving bits to a new physical storage media

- Migration

- the periodic transfer of digital materials from one hardware/software configuration to another, or from one generation of computer technology to a subsequent generation (Margaret Hedstrom and Sheon Montgomery in RLG Study)

Terminology

■ Migration

- Ensuring that past and future users can retrieve, display and otherwise use electronic materials in the face of constantly changing technology (*Preserving Digital Information*)
- Involves records coming from legacy information systems that lack export software functionality, requiring writing special purpose code or programs (Charles Dollar, *Authentic Electronic Records*)

Terminology

- Conversion
 - Automatically exporting or importing records from one software environment to another with little or no loss in structure and no loss in content or context even though the underlying bit stream is altered.

Preserving Electronic Records at the National Archives

- The National Archives has had a program for accessioning, describing, preserving and providing reference service to electronic (machine-readable records) for more than 30 years.

Preserving Electronic Records at the National Archives

- In 1970's all computer processing was performed at service bureaus.
- Original agency tape kept as master tape and copy made by NARS became reference copy.
- In 1980's staff acquired access to NIH Computer Center and wrote computer programs to make copies of electronic files.

Preserving Electronic Records...

- Preservation copies of electronic records limited to what could be produced by the mainframe computer at NIH
- Decision made to reengineer process
- Developed statement of work for Archival Preservation System (APS)

Archival Preservation System

- Four objectives in developing this system
 - Retain control of the media and perform work in-house
 - Handle a wider variety of file formats from Federal agencies and produce standardized output
 - Capture technical attributes automatically
 - Increase types of media acceptable from agencies

Armstrong Vs EOP (PROFS)

- NARA directed to deal with backup tapes from Bush and Reagan White House
- Court required preservation copies of these backup tapes
- APS, not yet fully developed, was modified so that NARA would meet the requirement to produce preservation copies of these backup tapes

AERIC

- Archival Electronic Records Inspection and Control system (AERIC)
 - developed by NARA to preserve the logical and conceptual structure of databases, thus providing a mechanism for migrating files across platforms

Current Research Projects

- InterPARES (International Research on Permanent Authentic Records)
 - goal is to develop theoretical and methodological knowledge essential for permanent preservation of electronic records and to formulate model policies, strategies and standards that will ensure preservation based upon requirements of the communities involved in this work

InterPARES

■ Four Domains of Investigation

- Domain I, Conceptual requirements for preserving authentic records, seeks to identify the elements of electronic records which are necessary to maintain authenticity
- Domain II, Appraisal criteria, is determining whether there are different theoretical criteria for electronic records

InterPARES

- Four domains (continued)
 - Domain III, Preservation, is identifying and developing procedures and resources required for implementation of conceptual requirements and criteria identified in the first two domains
 - Domain IV, Framework for Developing Policies, Strategies and Standards, is targeting the formulation of policies, strategies and standards based upon environment in which each institution is located

InterPARES

■ American Team

- Focusing on electronic records collections in government and universities
- Providing feedback to reinforce or modify theoretical models, templates and typologies
- Generating and testing systems analysis and design tools and strategies for future systems design

InterPARES

- International Team website
 - www.interpares.org
- American Team website
 - <http://is.gseis.ucla.edu/us-interpares/index.html>

San Diego Supercomputer Center

- “Collection-Based Persistent Digital Archives” brings together
 - archival storage technology from supercomputer centers
 - data grid technology from computer science community
 - information models from digital library community
 - preservation models from archival community

Collection-Based Persistent Digital Archives

- Defines the challenge as:
 - maintaining the ability to discover, access and display digital objects that are stored within an archive, while the technology used to manage the archive evolves

Collection-Based Persistent Digital Archives

- Persistent Archive System has four components:
 - support for ingestion
 - archival storage
 - information discovery
 - presentation of the collection

Collection-Based Persistent Digital Archives

- Creation of a one million E-mail collection has two phases:
 - Storage phase: Ingest, transform into an infrastructure independent form using XML, and store the collection
 - Retrieval/ Migration phase: Recreate on new technology, optimize the database and recreate the user interface

Collection-Based Persistent Digital Archives

- For more information about the San Diego Supercomputer Center Project see:
 - www.sdsc.edu/NARA.Publications.html
 - <http://www.dlib.org/dlib/april00/moore/04moore-pt2.html>

Open Archival Information System (OAIS)

- Initiated by NASA's Consultative Committee for Space Data Systems
- Serves as conceptual framework for an archival system dedicated to preserving and maintaining access over time

OAIS

- Minimum Responsibilities of OAIS-type archive include:
 - negotiate and accept appropriate information from information producer
 - Obtain sufficient control of information to ensure long-term preservation
 - Determine the scope of the Designated Community

OAIS

- Minimum Responsibilities of OAIS-type archive include:
 - Ensure that the information is understandable by the Designated Community
 - Follow documented policies and procedures to ensure information is preserved against reasonable contingencies

OAIS

- Workshop at College Park in October 1999
 - Established initial agenda
 - Sought participants for working groups dedicated to pursuing standardization in following areas
 - » Ingest (interaction between archive and data producer)
 - » Identification (establishing system of permanent, unique identifiers for archival digital objects)



OAIS

- » Certification (development of accreditation policies, protocols, etc., to establish authenticity, quality and usefulness of archive's holdings)

OAIS

- For more information about the Open Archival Information System, including the OAIS reference model and proceedings from the AWIICS workshop
- <http://ssdoo.gsfc.nasa.gov/nost/iosas/us/overview.html>