Keys to the Digital Future

Openness, Growth, Evolution, and Closure in Archival Information Systems

Lessons from NARA’s Experience

September 2008

Kenneth Thibodeau, Director
Electronic Records Archives Program
National Archives and Records Administration
IEEE Symposium on Mass Storage Systems & Technologies
Archival Information System

Conceptually: “an archive, consisting of an organization of people and systems, that has accepted the responsibility to preserve information and make it available for a Designated Community.”


Empirically: the National Archives’ Open Archival Information System, the Electronic Records Archives
What is the Electronic Records Archives (ERA)?

ERA is the system the National Archives and Records Administration (NARA) is developing to

- Reengineer and automate the lifecycle management of all types of records of the U.S. Government
- Preserve and provide sustained access to electronic records of the U.S. Government
ERA Base System Development

Focus:
- Federal Records
- National Archives

IOC Functions (2008):
- Creation, review and approval of records schedules
- Requests to transfer records, transfer of physical and legal custody
- Transfer, inspection, and archival storage of electronic records
Initial Users

U.S. Bureau of Labor Statistics

National Nuclear Security Administration

NAVAL OCEANOGRAPHIC OFFICE
John C. Stennis Space Center
We maximize America’s Sea Power by applying relevant oceanographic knowledge across the full spectrum of warfare

UNITED STATES PATENT AND TRADEMARK OFFICE
ERA Search and Access System Development

Initial Focus:
- Electronic records of the Executive Office of the President, G W. Bush
- Presidential Libraries
- ≥100 TB

Functions:
- Rapid ingest & indexing
  - Transformation to more accessible form.
- Archival storage
- Full content search
- Basic case management for special requests
Future Development

- Public Access to
  - Any information about records
    - Ordering of copies of records
  - Electronic records stored in the system
- Long-term preservation of electronic records
  - Ability to use a variety of techniques simultaneously and over time
- Review and redaction of sensitive content
- Support for Federal Records Centers
- Exponential growth in stored data
Keys to the Digital Future

- Openness
- Growth
- Evolution
- Closure

Lessons from the ERA experience
Openness

An Archival Information System needs to be open to

- New types of electronic records
- Rising and changing user expectations
- Creative approaches to meeting the challenges of electronic records and demanding users.
An Archival Information System needs to be open to new types of electronic records, rising and changing user expectations, and creative approaches to meeting the challenges of electronic records and demanding users.
New Types of Records: Geographic Information Systems
New Types of Records: Product Data

Computer Assisted Design

Computer Assisted Engineering

Product Analysis and Testing

Computer Assisted Manufacture
New Types of Records: Critical Infrastructure Data

New Types of Records:
Virtual Reality: Crime Scene Investigation
New Types of Records: Medical Tests and Observations
Openness

An Archival Information System needs to be open to

- New types of electronic records
- Rising and changing user expectations
- Creative approaches to meeting the challenges of electronic records and demanding users.
Rising and Changing User Expectations
An Archival Information System needs to be open to

- New types of electronic records
- Rising and changing user expectations
- Creative approaches to meeting the challenges of electronic records and demanding users.
Creative Approaches

- The conceptual apparatus we bring to bear on
  - The nature of records
  - Requirements for preserving records
  - Requirements for serving users
Creative approaches: Partnerships

National Science Foundation

National Computational Science Alliance

The Library of Congress

Army Research Laboratory

PDES, Inc.

InterPARES Project
Keys to the Digital Future

- Openness
- Growth
- Evolution
- Closure
An Archival Information System needs to be able to grow to

- Process, store and provide access to increasing volumes of electronic records
- Accommodate increasing numbers and frequency of use
Increasing Volumes of Digital Information

• In 2006, the amount of digital information created, captured, and replicated was ...281 exabytes or 281 billion gigabytes. This is about 3 million times the information in all the books ever written.

• By 2011, the digital universe will be 10 times the size it was in 2006.

• Not all information created and transmitted gets stored, but by 2011, almost half of the digital universe will not have a permanent home.

• The number of electronic information “containers”—files, images, packets, tag contents—is growing 50% faster than the number of gigabytes. The information created in 2011 will be contained in more than 20 quadrillion — 20 million billion — of such containers.

Transfers of Digital Files to NARA

- 1970-1988: 6,000
- 1989-1995: 15,000

Transfers of Digital Files to NARA

1970-1988
1989-1995
Reagan/ Bush
Planning for Open-ended Growth

Federation

Base ERA Instance

System Management

Future ERA Instance

SAS ERA Instance

DREN

Internet

Public

Other Agencies

NARANET

Future ERA

Instance

System Management

Public

Other Agencies

DREN

Federation

Internet
Keys to the Digital Future

- Openness
- Growth
- Evolution
- Closure
An Archival Information System needs to be able to evolve in response to

- Changing Information Technology
  - Obsolescence
  - Opportunities

- Changing business requirements
An Archival Information System needs to be able to evolve in response to

- Changing Information Technology
  - Obsolescence
  - Opportunities
- Changing business requirements
Obsolescence of Formats of Electronic Records

• Strategy: Preservation and Access Levels
  – Common:
    • Retain records in original formats
  – Basic Level:
    • Use original or contemporary software for access
  – Enhanced Level
    • Create new version in current format, or
    • Use new software for access to original format
  – Ideal Level
    • Create version in persistent format, or
    • Create persistent software for management and access
Obsolescence of Formats of Electronic Records

• ERA Preservation Framework (Architecture Component)
  – Does not prescribe specific preservation solutions
  – Allows a variety of different software tools to be introduced and used for different formats.
  – Enforces archival requirements
Obsolescence of Formats of Electronic Records

- ERA System Architecture:
  - Does not prescribe specific preservation solutions
  - Allows a variety of different software tools to be introduced and used for different formats.

Example
An Archival Information System needs to be able to evolve in response to

- Changing Information Technology
  - Obsolescence
  - Opportunities
- Changing business requirements
Preservation Options

- General
  - Maintain original technology
  - Emulation
  - Static Translator
  - Dynamic Translator
  - Virtual Machine
  - Re-engineer Software
  - Universal Virtual Computer
- Specific
  - Viewer
  - Version Migration
  - Format Standardization
  - Typed Object Conversion
  - Persistent Object
  - Object Interchange Format
  - Multi-Valent Document
  - Programmable Chips
  - Re-engineer Software
  - Universal Virtual Computer

Preserve Technology

Method

Preserve Objects
An Archival Information System needs to be able to evolve in response to

- Changing Information Technology
  - Obsolescence
  - Opportunities
- Changing business requirements
Changing Information Technology: Service Oriented Architecture
Service Oriented Architecture
AsBuilt
An Archival Information System needs to be able to evolve in response to

- Changing Information Technology
  - Obsolescence
  - Opportunities
- Changing business requirements
Evolution of Business Requirements
# Records Schedule: Current

## Request for Records Disposition Authority

<table>
<thead>
<tr>
<th>Item Number</th>
<th>Description of Item and Proposed Disposition</th>
<th>GRS or Superceded Job Citation</th>
<th>Action Taken (NARA Use Only)</th>
</tr>
</thead>
</table>

### Leave Blank (NARA Use Only)

- **Job Number**
- **Date Received**
- **Notification to Agency**
  - In accordance with the provisions of 44 U.S.C. 3303a, this disposition request, including amendments, is approved except for items that may be marked "disposition not approved" or "withdrawn" in column 10.
- **Date**
- **Archival of the United States**

### Agency Certification

I hereby certify that I am authorized to act for this agency in matters pertaining to the disposition of its records and that the records proposed for disposal on the attached page(s) are no longer needed for the business of this agency or will not be needed after the retention periods specified, and that written concurrence from the General Accounting Office, under the provisions of Title 8 of the GAO Manual for Guidance of Federal Agencies:

- **Is not required**
- **Is attached**
- **Has been requested**

- **Signature of Agency Representative**
- **Title**
- **Date (mm/dd/yyyy)**
Create Schedule Item

Temporary Records

Permanent Records
Keys to the Digital Future

- Openness
- Growth
- Evolution
- Closure
Closure

An Archival Information System needs to be able to provide closure to ensure:

- Preservation and presentation of authentic records
- Comprehensive lifecycle management of electronic records
- Consistency with well-established archival science
ERA: a Set of Nested Systems

- **Outer system**
  lifecycle management of records of all types

- **Inner Electronic Records System**
  Ingest, preservation, disposition, and access to electronic records

- **Search & Preservation Frameworks**
  Support a variety of different approaches to different needs

- **Archival “mini-systems”**
  Specific, systematic management for each series or aggregate of electronic records
Document v. Record

- A **document** is a bounded physical representation of body of information designed with the capacity (and usually intent) to communicate. A document may manifest symbolic, diagrammatic or sensory-representational information. ...

- The information communicated by a document depends on its content and structure.

- A **record** is a document made or received in the course of a practical activity as an instrument or a by-product of such activity, and set aside for action or reference.
  - [http://www.interpares.org/ip2/ip2_terminology_db.cfm](http://www.interpares.org/ip2/ip2_terminology_db.cfm)

- The information communicated by a record depends on its content, structure, and **context**.
What does this document tell us?
What does this document tell us about the U.S. Government?
Archival Aggregate as Directed Graph

Every record has an ‘archival bond,’ the set of relationships established by an actor between that record and other records of the actor’s activity.
Preservation

- Documents can be preserved as individual objects
- Records can only be preserved as ordered sets.

→ An Archival Information System for records must ensure that
  - Submission Information Packages,
  - Archival Information Packages and
  - Dissemination Information Packages
are managed to respect the original order of records.
Electronic Records

May be instantiated as subsets of complex ordered sets
ERA as a Set of Mini-Systems

A Lifecycle Management Plan for a Records Aggregate, such as a series, defines a “Mini-system,” i.e., systematic controls for that aggregate stretching from ingest to dissemination.
Openness, Growth, Evolution, and Closure

- **Openness**
  - New types of electronic records
  - Rising and changing user expectations
  - Creative approaches

- **Growth**
  - Exponential increase in volumes of stored data

- **Evolution**
  - Changing Information Technology
  - Changing business requirements

- **Closure**
  - Preservation of electronic records as members of ordered sets.
Thank you.

For more information:
www.archives.gov/era