

Open

Grow

Close

Evolve

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

Open

Grow

Close

Evolve

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.”
 - **ISO Reference Model for an Open Archival Information System (OAIS). ISO 14721:2003**
- ✦ Empirically: the National Archives’ Open Archival Information System, the Electronic Records Archives

Open

Grow

Close

Evolve

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

Open

Grow

Close

Evolve

ERA Development Timeline

9/05

9/06

9/07

9/08

9/09

9/10

9/11

ERA Base System



6/08



Initial Operating Capability (IOC)

ERA Search & Access System



Enhancement



Enhancement



Enhancement



Enhancement



Full Operating Capability

Operation & Maintenance



Open

Grow

Close

Evolve

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

Open

Grow

Close

Evolve

Initial Users



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

Open

Grow

Close

Evolve

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**

Open

Grow

Close

Evolve

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

Open

Grow

Close

Evolve

Keys to the Digital Future

🔑 **Openness**

🔑 **Growth**

🔑 **Evolution**

🔑 **Closure**

Lessons from the ERA experience

Open

Grow

Close

Evolve

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.

Open

Grow

Close

Evolve

Openness

🔑 An Archival Information System needs to be open to

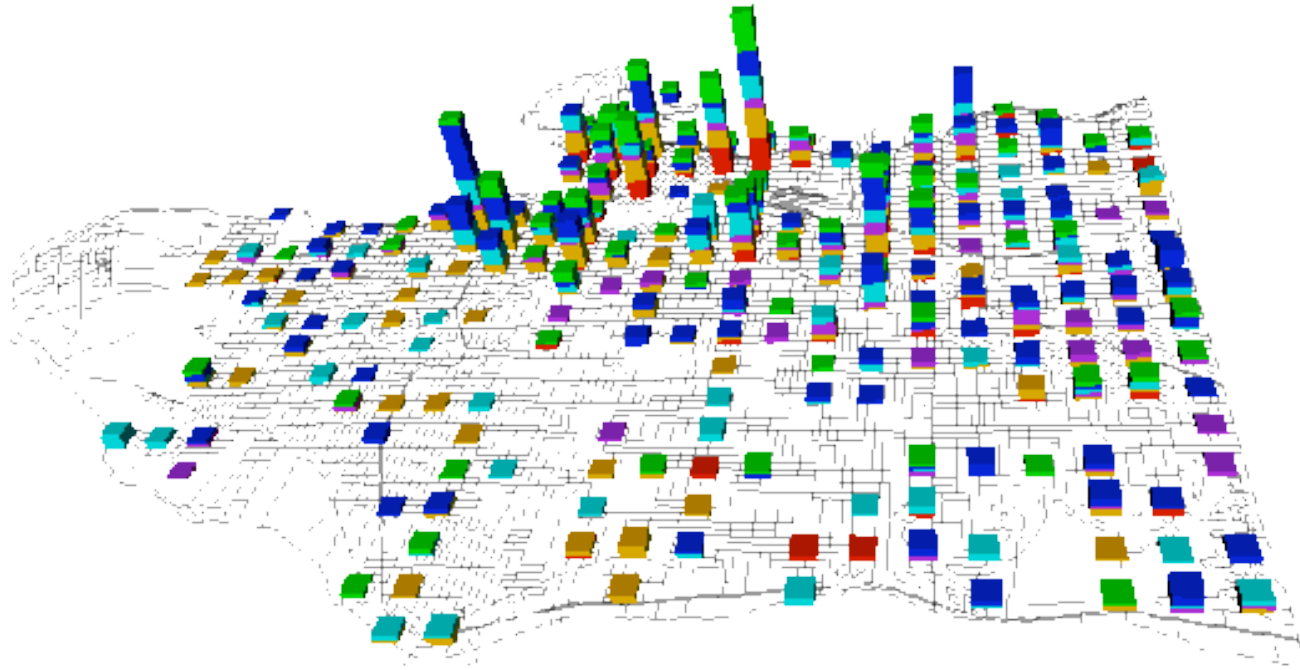
🔓 New types of electronic records

🔓 Rising and changing user expectations

🔓 Creative approaches to meeting the challenges of electronic records and demanding users.



New Types of Records: Geographic Information Systems

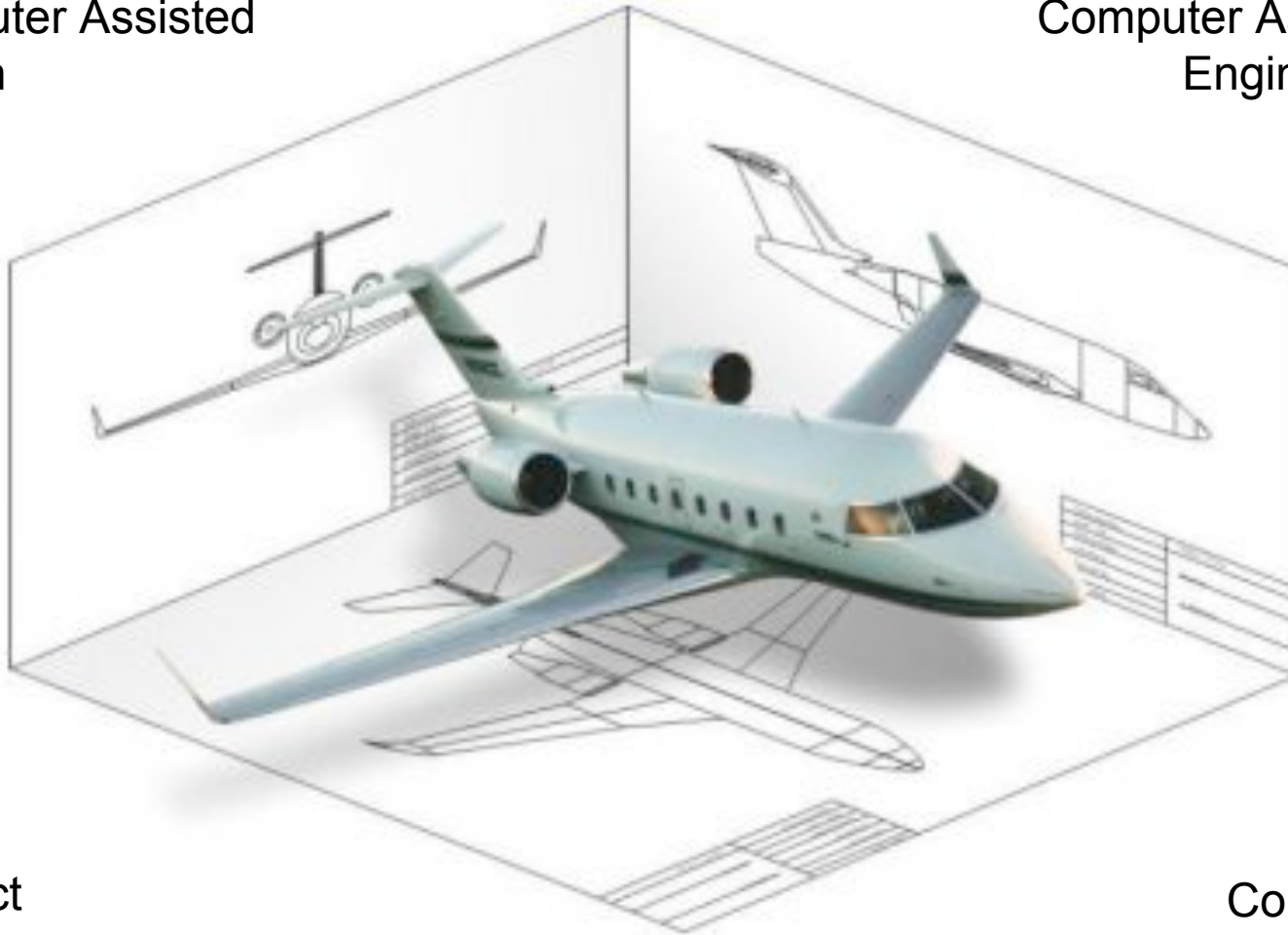




🔓 New Types of Records: Product Data

Computer Assisted
Design

Computer Assisted
Engineering



Product
Analysis and Testing

Computer
Assisted Manufacture



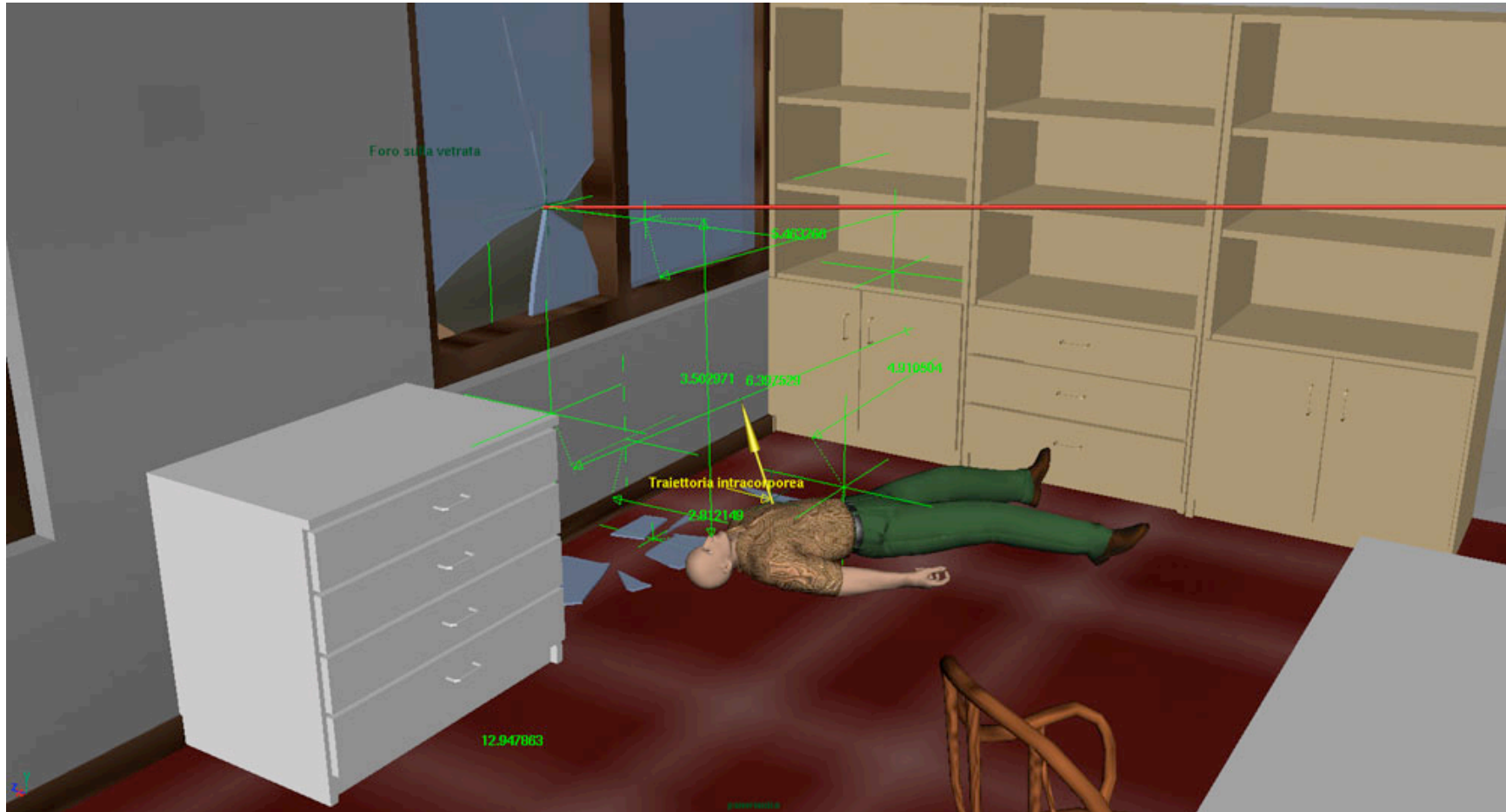
New Types of Records: Critical Infrastructure Data



Source: CLindberg http://commons.wikimedia.org/wiki/Image:135_Bridge_Collapse_4crop.jpg

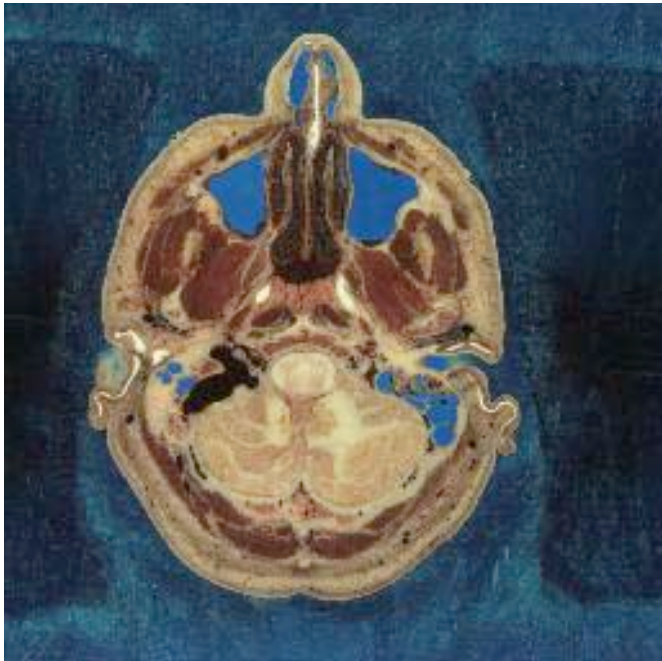


New Types of Records: Virtual Reality: Crime Scene Investigation





🔗 New Types of Records: Medical Tests and Observations



Open

Grow

Close

Evolve

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



Open

Grow

Close

Evolve

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.



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



San Diego Supercomputer Center



National Computational Science Alliance



Global Grid Forum



The Library of Congress



Army Research Laboratory



DIGITAL LIBRARY FEDERATION



Open

Grow

Close

Evolve

Keys to the Digital Future

⌘ Openness

⌘ Growth

⌘ Evolution

⌘ Closure

Open

Grow

Close

Evolve

Growth

→ 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



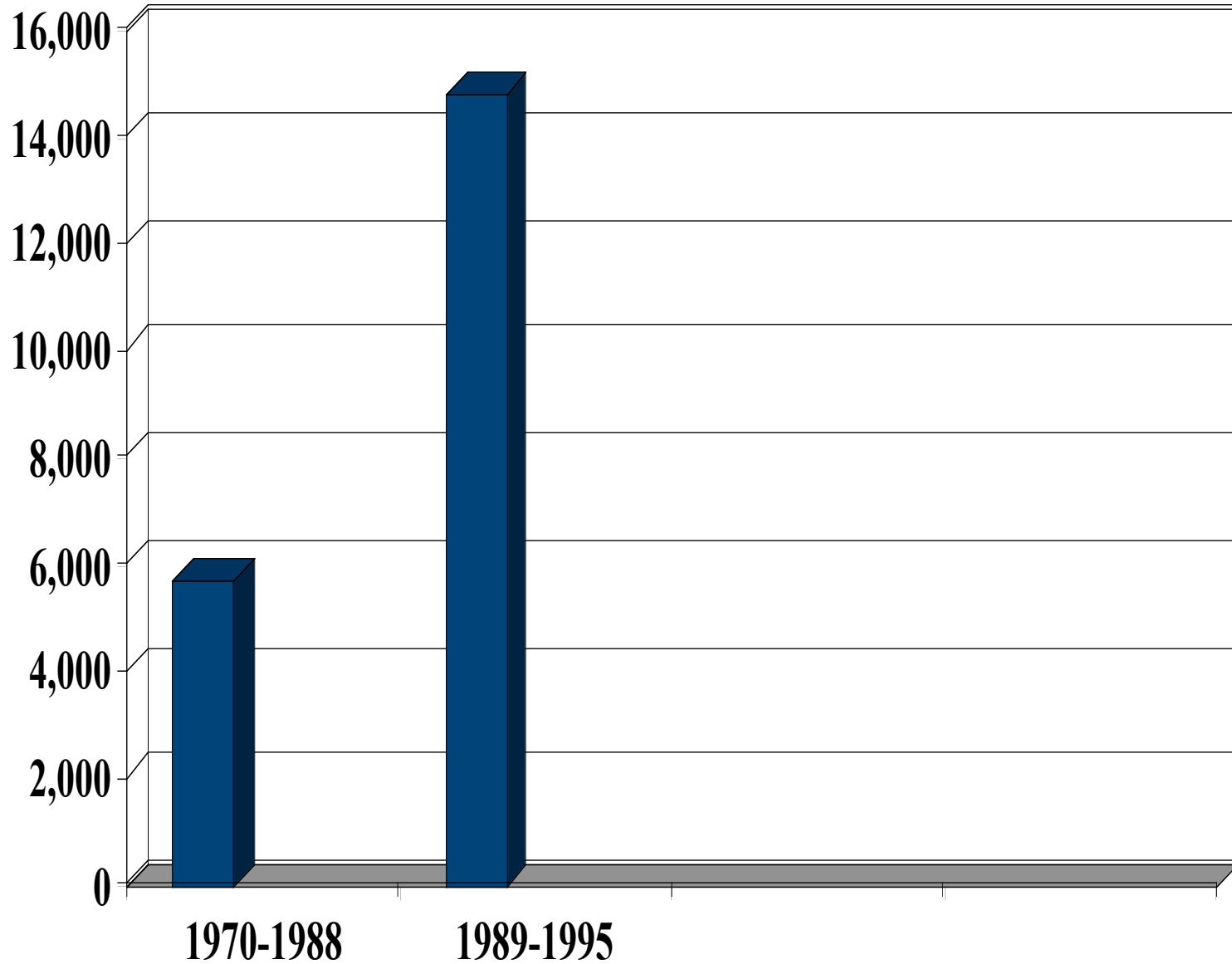
Increasing Volumes of Digital Information

- 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

– IDC. The Diverse and Exploding Digital Universe. An Updated Forecast of Worldwide Information Growth Through 2011. March 2008

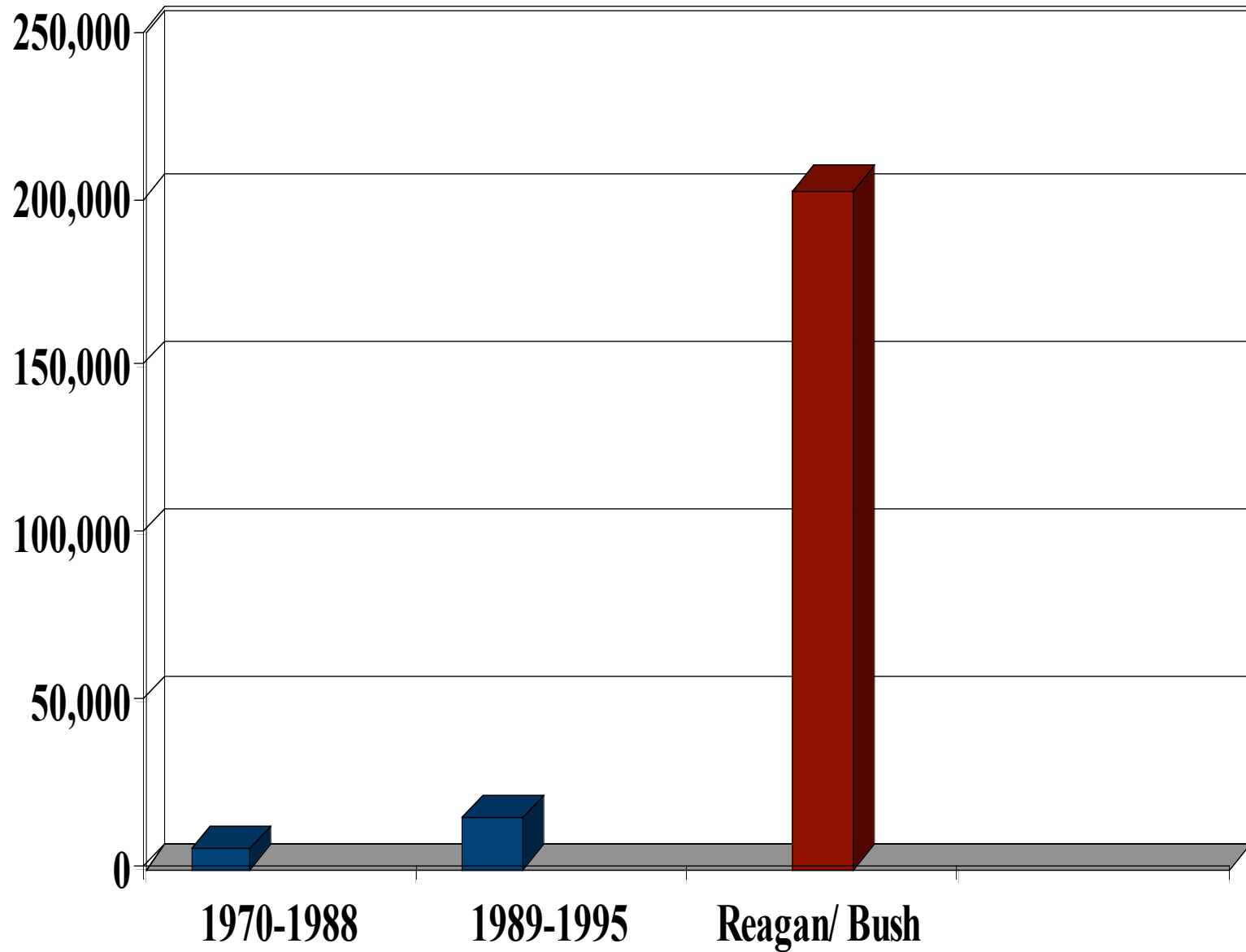


Transfers of Digital Files to NARA



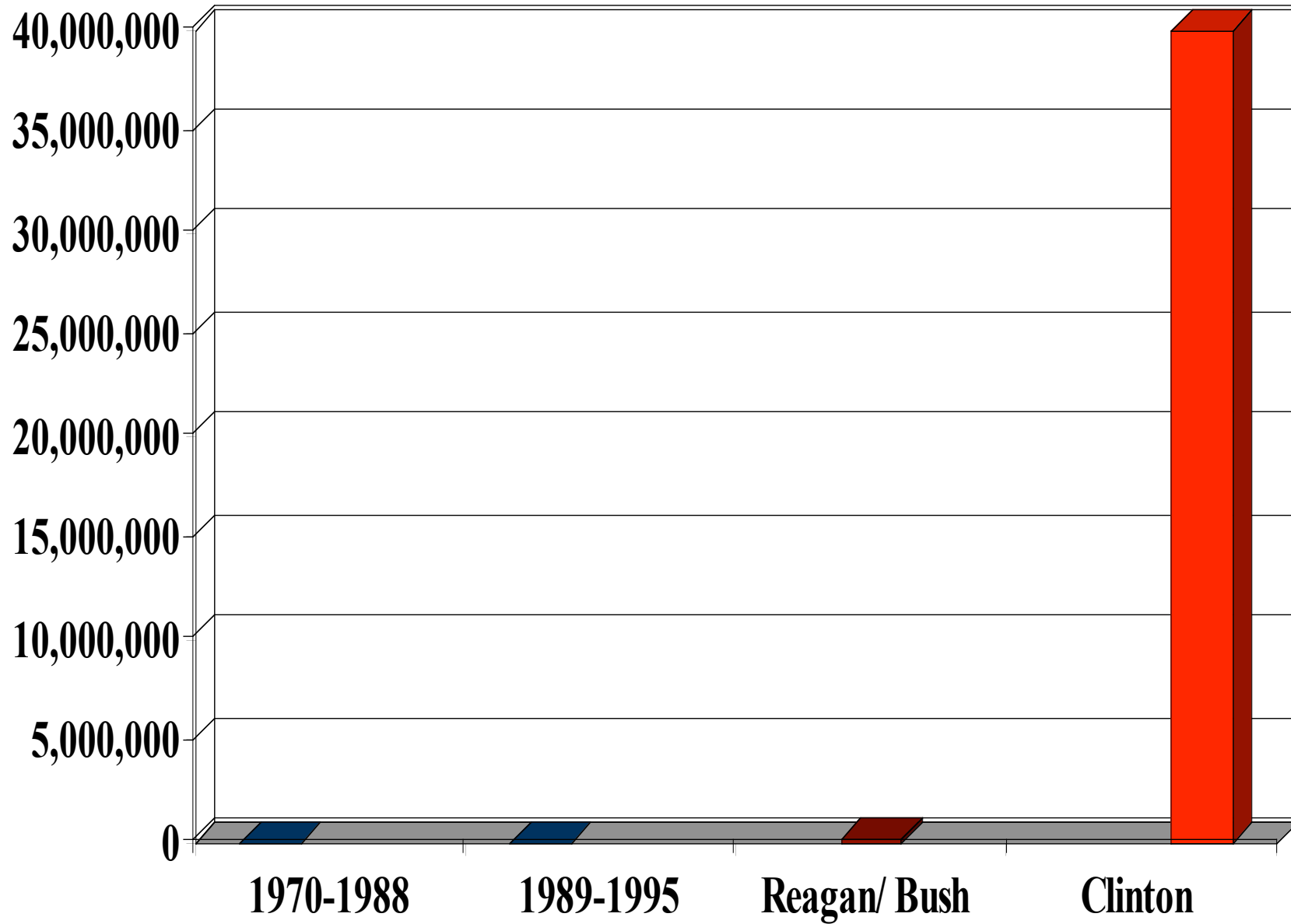


Transfers of Digital Files to NARA



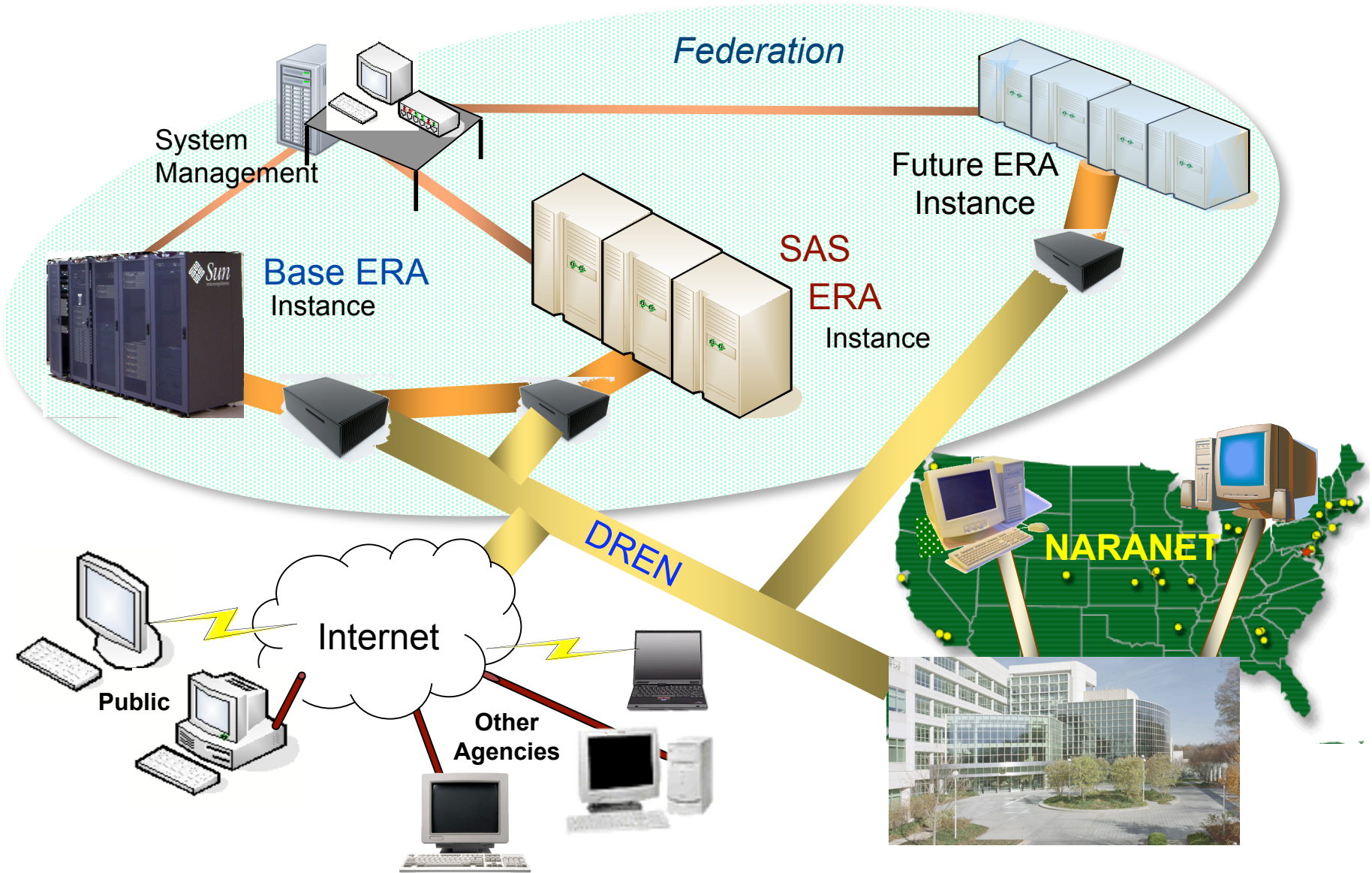


Transfers of Digital Files to NARA





Planning for Open-ended Growth



Open

Grow

Close

Evolve

Keys to the Digital Future

🔑 Openness

🔑 Growth

🔑 Evolution

🔑 Closure

Open

Grow

Close

Evolve

Evolution

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

🔧 Changing Information Technology

- Obsolescence
- Opportunities

🔧 Changing business requirements

Open

Grow

Close

Evolve

Evolution

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

🔗 Changing Information Technology

- Obsolescence
- Opportunities

🔗 Changing business requirements

Evolve



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

Evolve



Obsolescence of Formats of Electronic Records

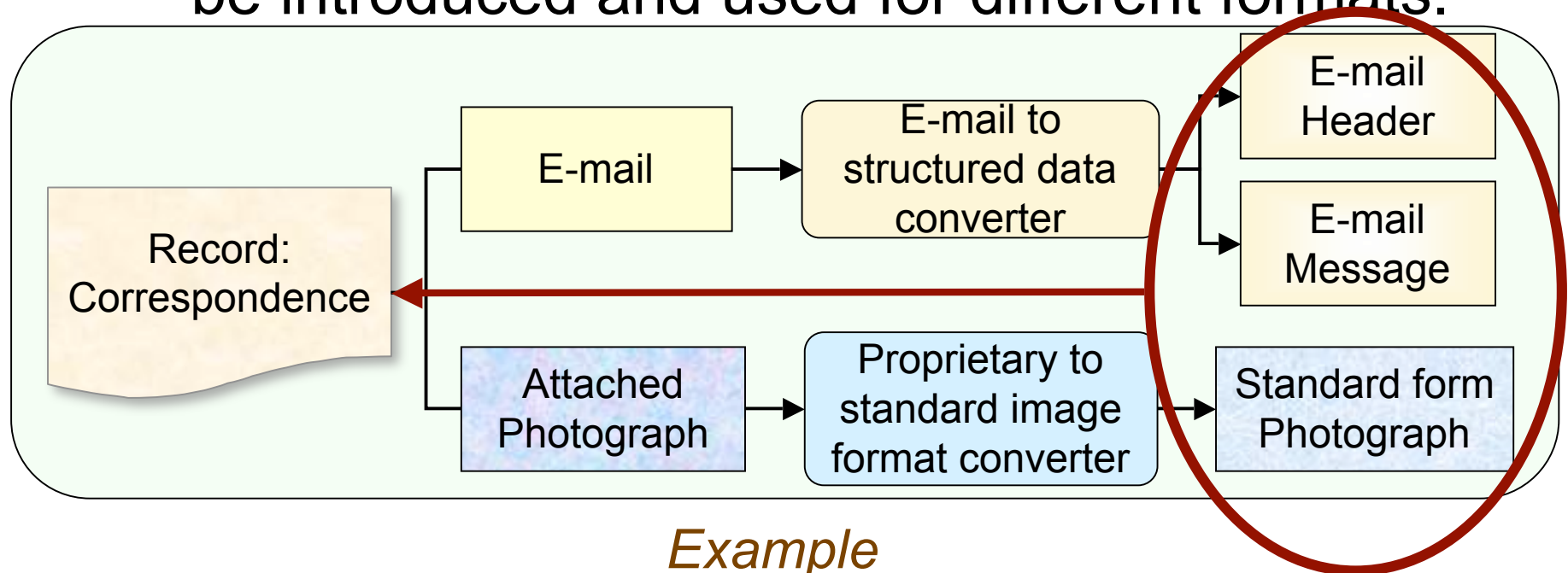
- ERA Preservation Framework
(Architecture Component)
 - Does not prescribe specific preservation solutions
 - Allows a variety of different software tools to be introduced and used for different formats.
 - Enforces archival requirements

Evolve



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.



Open

Grow

Close

Evolve

Evolution

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

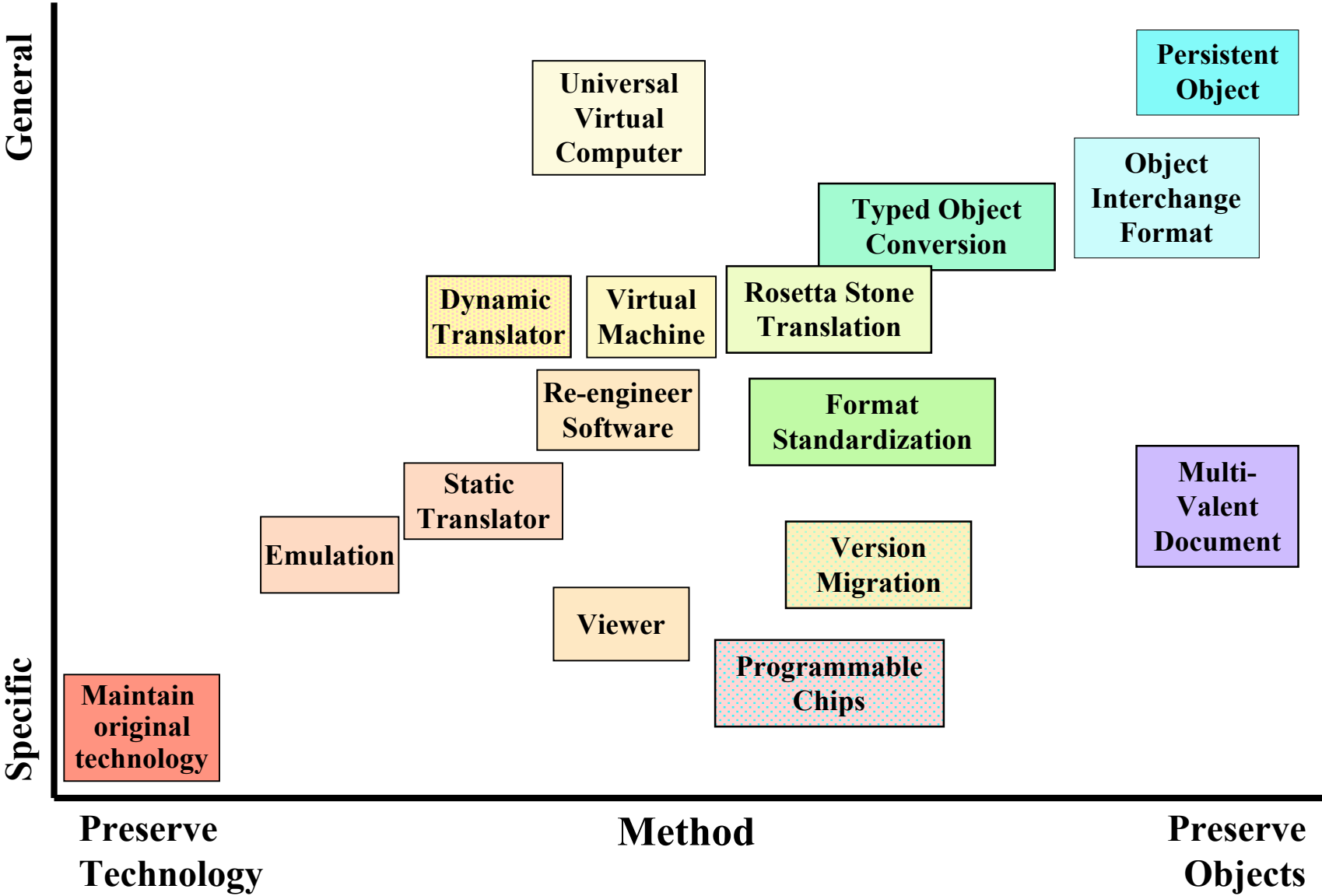
🔗 Changing Information Technology

- Obsolescence
- Opportunities

🔗 Changing business requirements

Evolve

Preservation Options



Open

Grow

Close

Evolve

Evolution

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

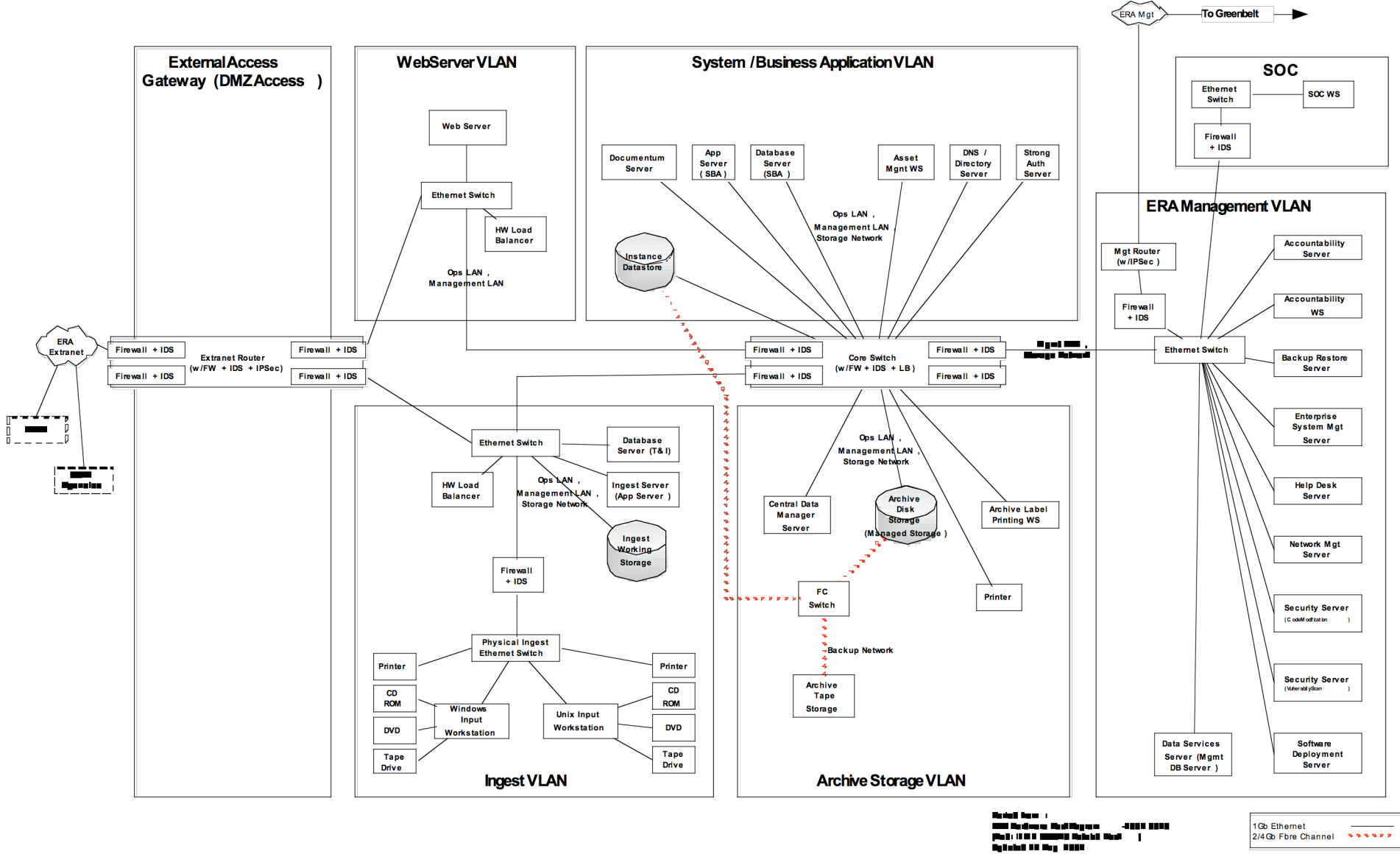
🔗 Changing Information Technology

- Obsolescence
- Opportunities

🔗 Changing business requirements

Evolve

Changing Information Technology: Service Oriented Architecture

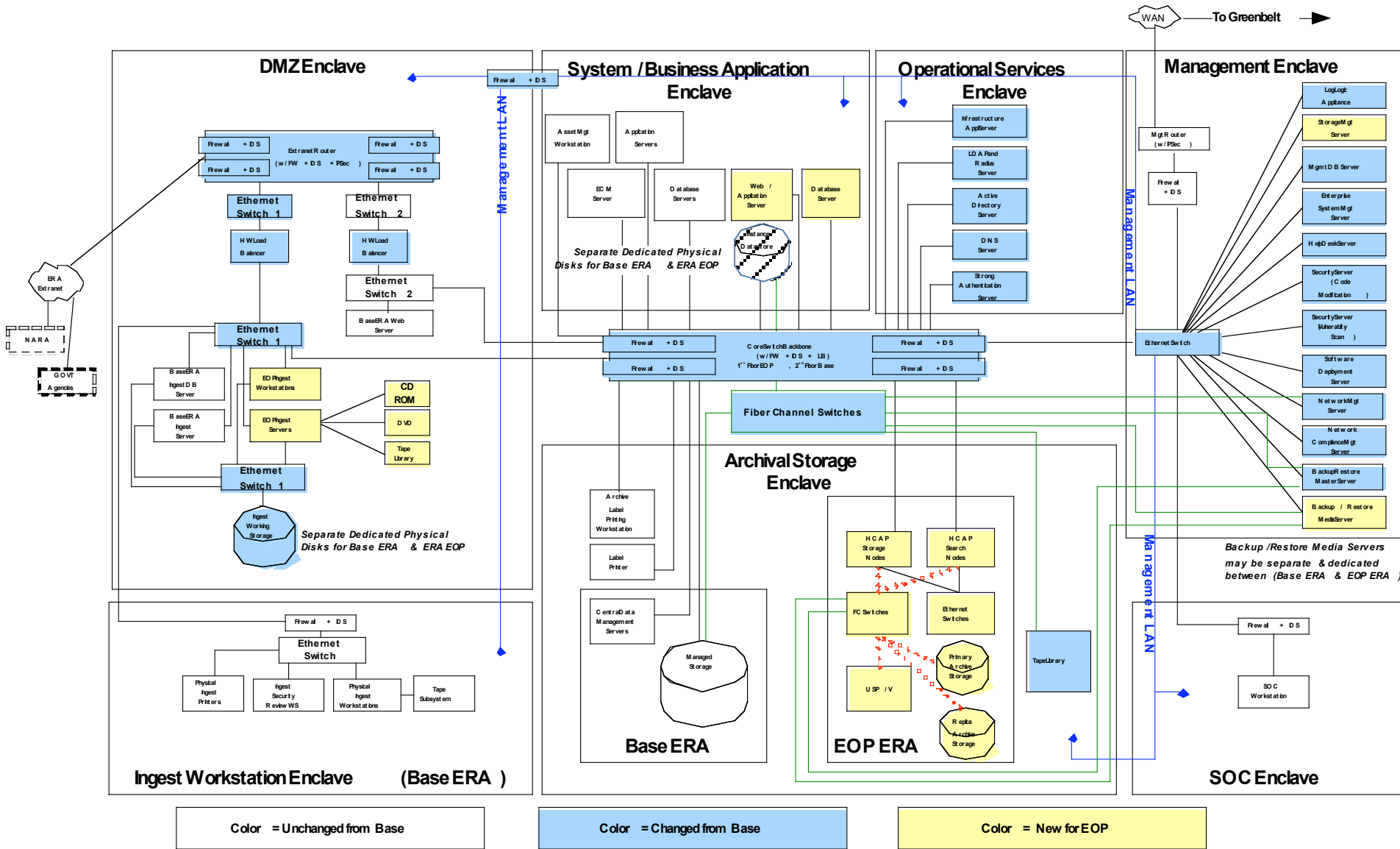


Legend:
 Solid line: 1Gb Ethernet
 Dashed red line: 2/4Gb Fibre Channel

Legend:
 Solid line: 1Gb Ethernet
 Dashed red line: 2/4Gb Fibre Channel

Evolve

Service Oriented Architecture As Built



Evolution

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

– Changing Information Technology

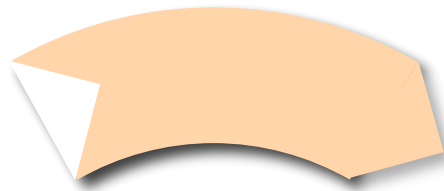
- Obsolescence

- Opportunities

– Changing business requirements



🔗 Evolution of Business Requirements

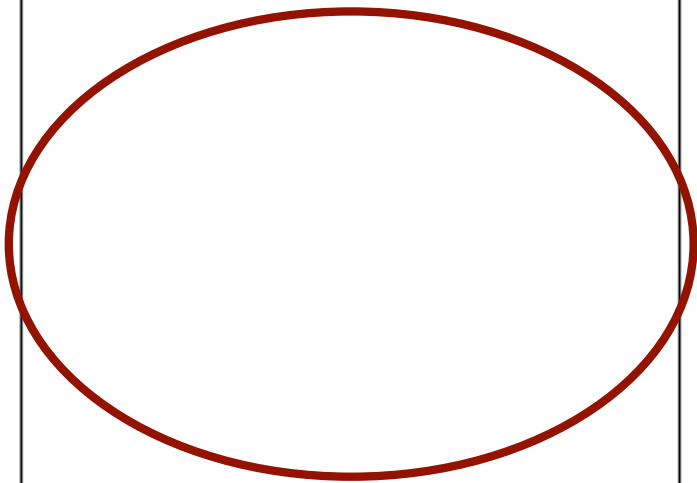


Requirements

**Technical
Solutions**

Evolve

Records Schedule: Current

Request for Records Disposition Authority <small>(See instructions on reverse)</small>		Leave Blank (NARA Use Only)	
To: National Archives and Records Administration (NIR) Washington, DC 20408		Job Number	
1. From: (Agency or establishment)		Date Received	
2. Major Subdivision		Notification to Agency In accordance with the provisions of 44 U.S.C. 3302a, the disposition request, including amendments, is approved except for items that may be marked "disposition not approved" or "withdrawn" in column 10.	
3. Minor Subdivision			
4. Name of Person with whom to confer	5. Telephone (include area code)	Date	Archivist of the United States
6. 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 not now 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: <input type="checkbox"/> is not required <input type="checkbox"/> is attached <input type="checkbox"/> has been requested			
Signature of Agency Representative		Date (mm/dd/yyyy)	
7. Item Number	8. Description of Item and Proposed Disposition	9. GRS or Superseded Job Citation	10. Action taken (NARA Use Only)
			



Create Schedule Item

Temporary Records

Permanent Records

General		7. Item Number	8. Description of Item and Proposed Disposition	9. GRS or Superseded Job Citation	10. Action Taken (NARA Use Only)
Item ID: <input type="text"/>					
*Description: <input type="text"/>					
Does agency have an associated manual? <input type="radio"/>					
*Manual ID: <input type="text"/>					
*Manual Version: <input type="text"/>					
*Manual Item ID: <input type="text"/>					
Is this a change to an approved schedule? <input type="radio"/>					
Is this item media neutral? <input type="radio"/> Yes <input type="radio"/> No					
Final Disposition					
*Final Disposition: <input type="radio"/> Permanent					
Temporary Disposition Instructions					
Cutoff Instructions: <input type="text"/>					

Transfer Instructions		Records to which these transfer instructions apply:		Accession Instructions	
Transfer to: <input type="text"/>	Time after cutoff when transfer occurs: <input type="text"/>	<input type="text"/>	*Transfer to: <input type="text"/>	<input type="radio"/> Accession immediately on cut-off	
*Retention Period			*Time after cutoff when transfer occurs: <input type="text"/>	<input type="radio"/> Accession <input type="text"/> after cut-off	
<input type="radio"/> Destroy immediately on cut-off			*Estimated First Transfer: <input type="text"/>	<input type="radio"/> Accession between <input type="text"/> years and <input type="text"/> years after cut-off	
<input type="radio"/> Destroy <input type="text"/> after cut-off				<input type="radio"/> Accession in <input type="text"/> year blocks <input type="text"/> years after cutoff of most recent records in the block	
<input type="radio"/> Destroy between <input type="text"/> years and <input type="text"/> years after cut-off				<input type="radio"/> Other <input type="text"/>	
<input type="radio"/> Retain at least <input type="text"/> years after cut-off, but longer is authorized				*Estimated First Transfer: <input type="text"/>	
<input type="radio"/> Retain no more than <input type="text"/> years after cut-off				*If records are not transferred to NARA physical custody when legal custody is transferred, specify institution that will maintain physical records: <input type="text"/>	
<input type="radio"/> Destroy when no longer needed					
<input type="radio"/> Destroy <input type="text"/> years after cut-off or when <input type="text"/> occurs, whichever is sooner					
<input type="radio"/> Destroy <input type="text"/> years after cut-off or when <input type="text"/> occurs, whichever is later					
<input type="radio"/> Destroy <input type="text"/> years after cut-off or <input type="text"/> years after <input type="text"/> occurs, whichever is sooner					
<input type="radio"/> Destroy <input type="text"/> years after cut-off or <input type="text"/> years after <input type="text"/> occurs, whichever is later					
<input type="radio"/> Other <input type="text"/>					

Additional Information		Annual Accumulation	
*Estimated Current Volume		*Electronic/Digital: <input type="text"/>	
<input type="radio"/> Electronic/Digital: <input type="text"/>		<input type="radio"/> Paper: <input type="text"/> cubic feet	
<input type="radio"/> Paper: <input type="text"/> cubic feet		<input type="radio"/> Microform: <input type="text"/> microfiche <input type="text"/> microfilm	
<input type="radio"/> Microform: <input type="text"/> microfiche <input type="text"/> microfilm		<input type="radio"/> Traditional Special Media: <input type="text"/> Units: <input type="text"/>	
<input type="radio"/> Traditional Special Media: <input type="text"/> Units: <input type="text"/>		<input type="radio"/> Unknown: <input type="text"/>	
<input type="radio"/> Unknown: <input type="text"/>			
Date Span		*First year of records accumulation: <input type="text"/>	*End year of records accumulation: <input type="radio"/> Records ceased accumulation in <input type="text"/>
			<input type="radio"/> Records are still being accumulated

Open

Grow

Close

Evolve

Keys to the Digital Future

🔑 Openness

🔑 Growth

🔑 Evolution

🔑 Closure

Open

Grow

Close

Evolve

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

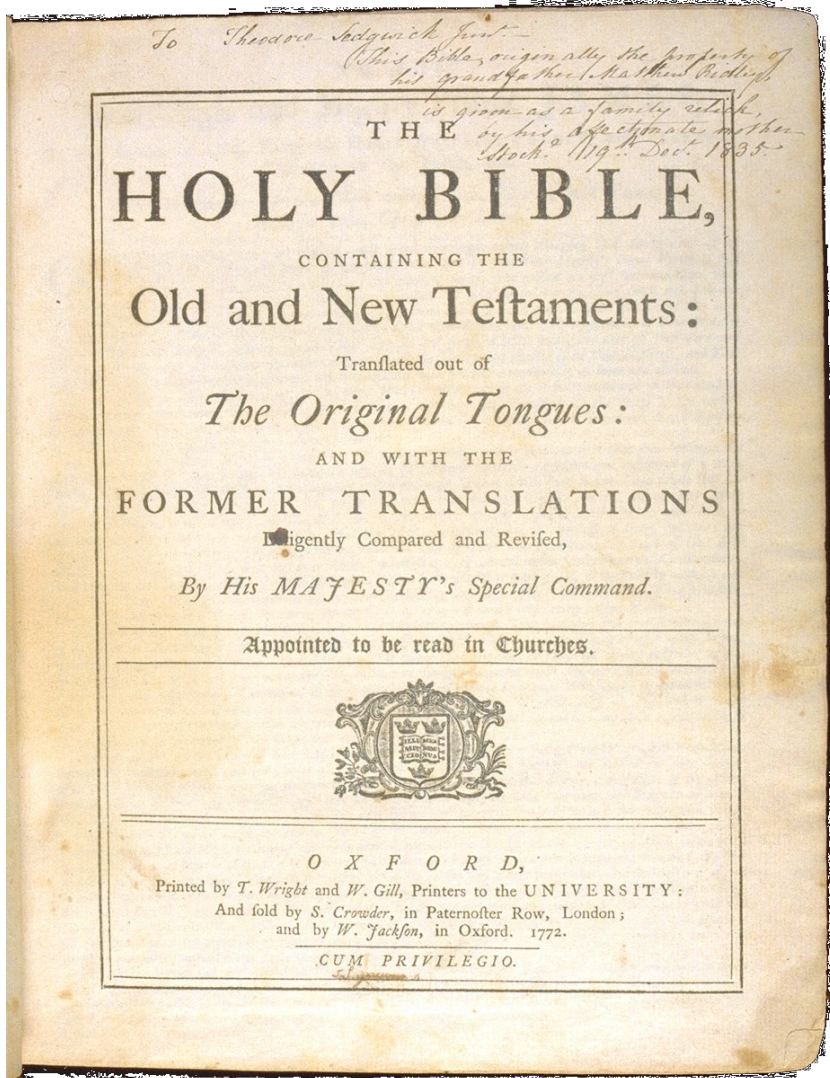


Document v. Record

- 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. ...
 - en.wikipedia.org/wiki/Document
- 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
- The information communicated by a record depends on its content, structure, and **context**.



Document



What does this document tell us?



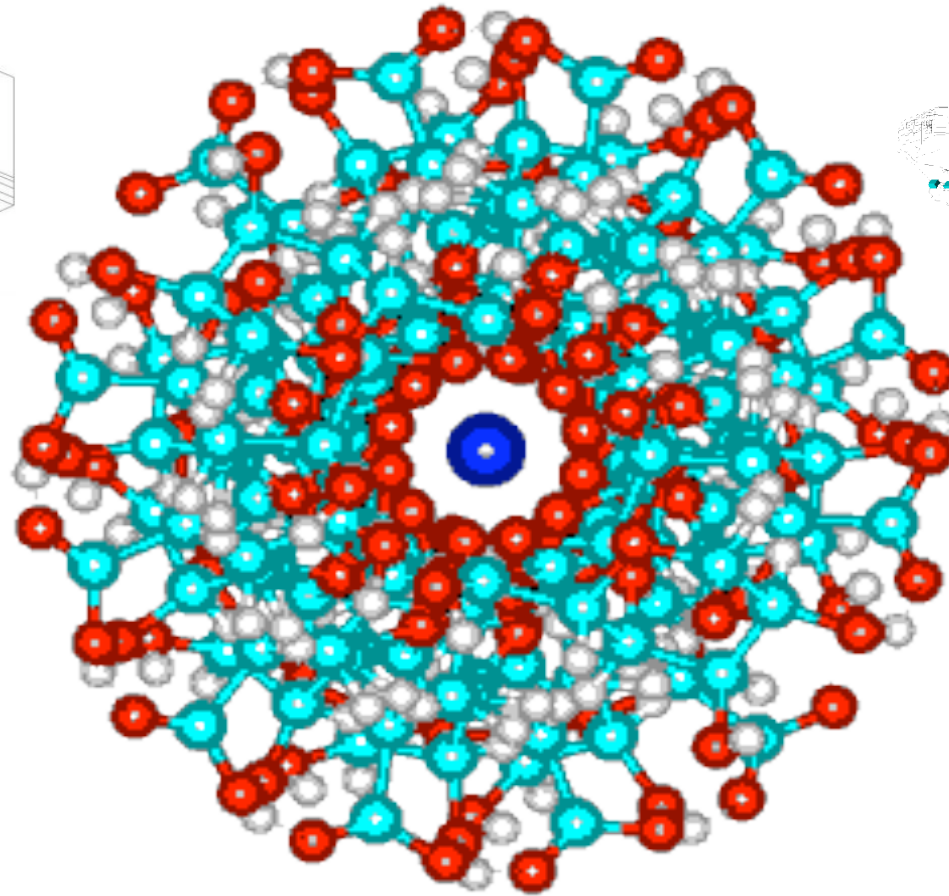
Preservation

- 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 Packagesare managed to respect the original order of records.

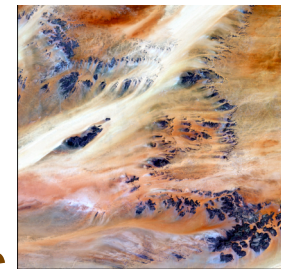
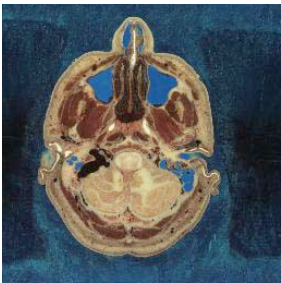
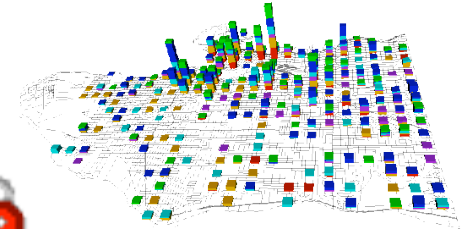
Close

Electronic Records

Open



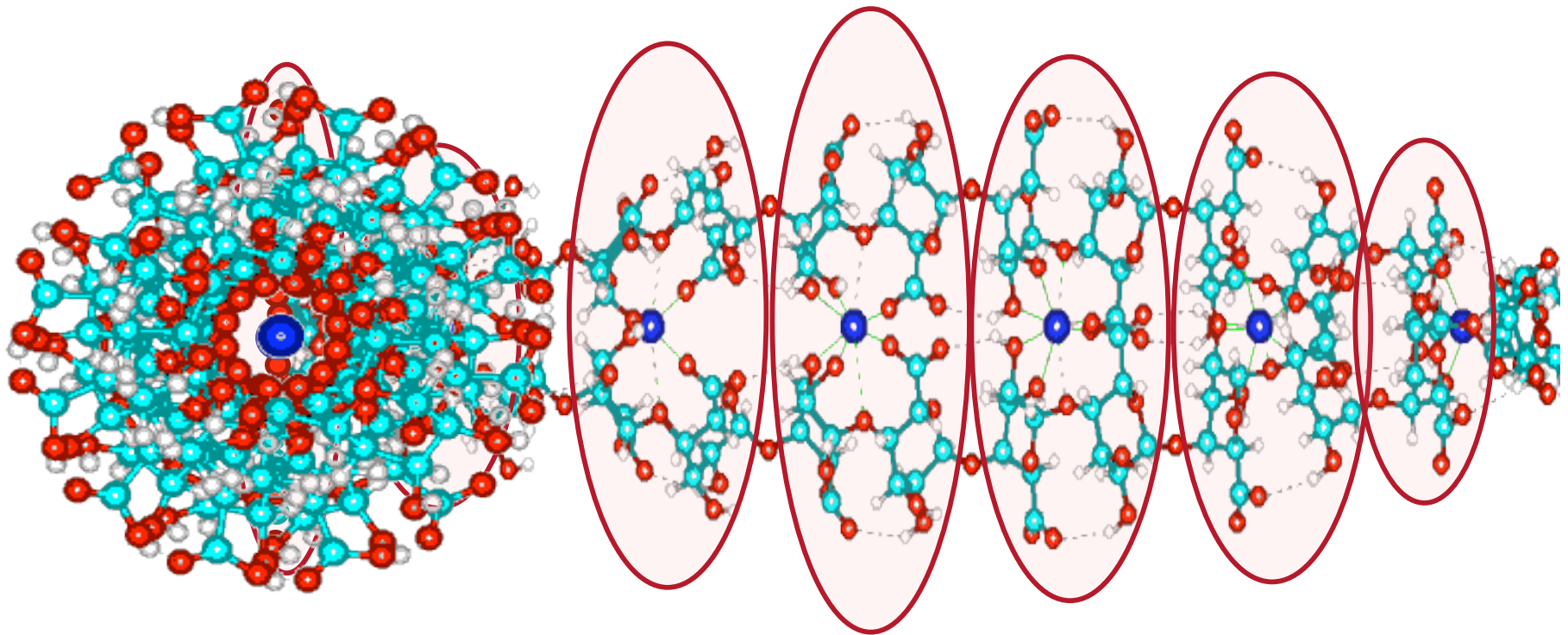
Start ASLT FAMIRB TFAUTO
BNE MSCHE THEFT



May be instantiated as subsets
of complex ordered sets

Close

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.

Open

Grow

Close

Evolve

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.

Open

Grow

Close

Evolve

Thank you.



For more information:

www.archives.gov/era