

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

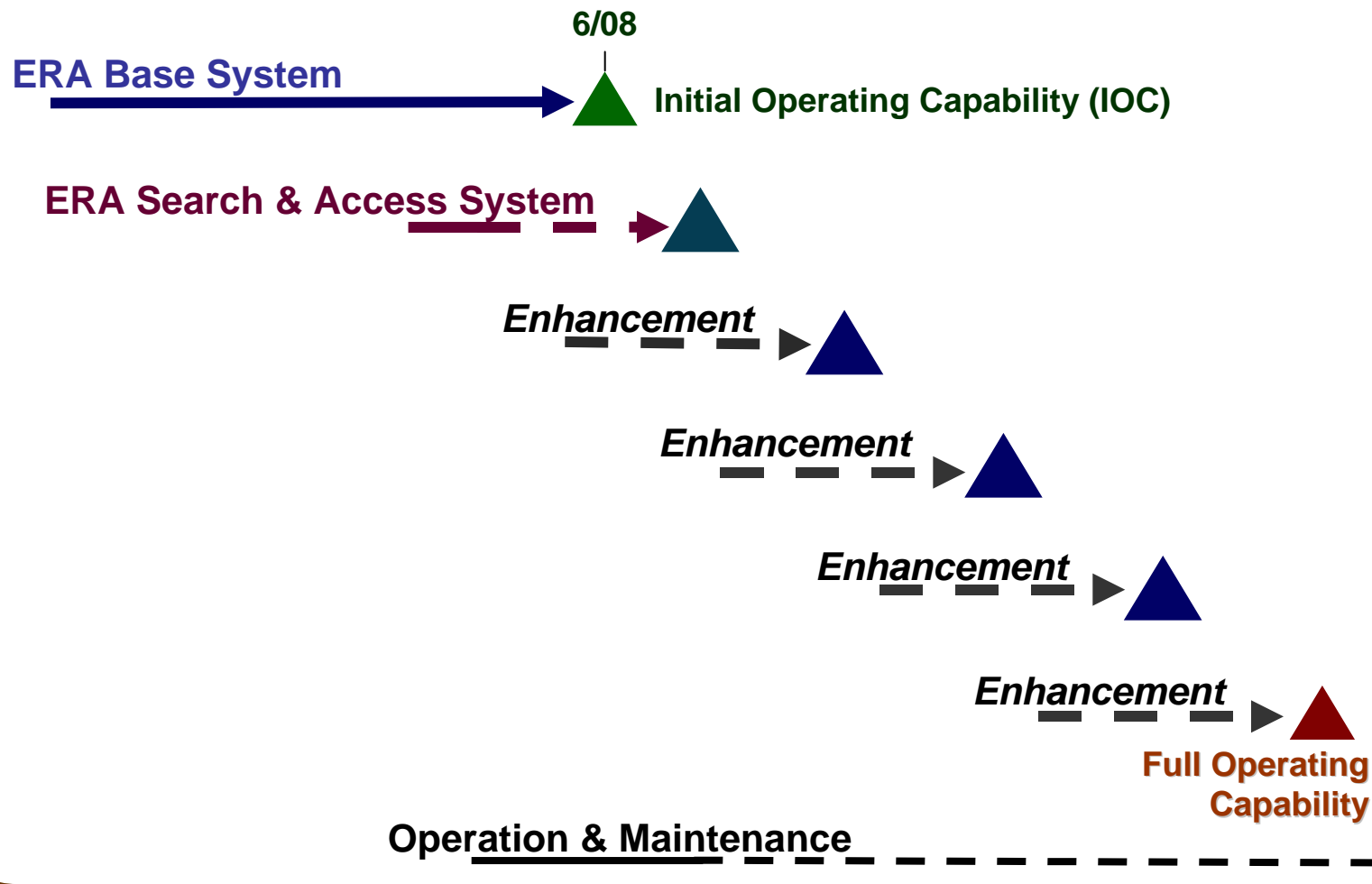
9/07

9/08

9/09

9/10

9/11



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



**U.S. Bureau of  
Labor Statistics**



**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

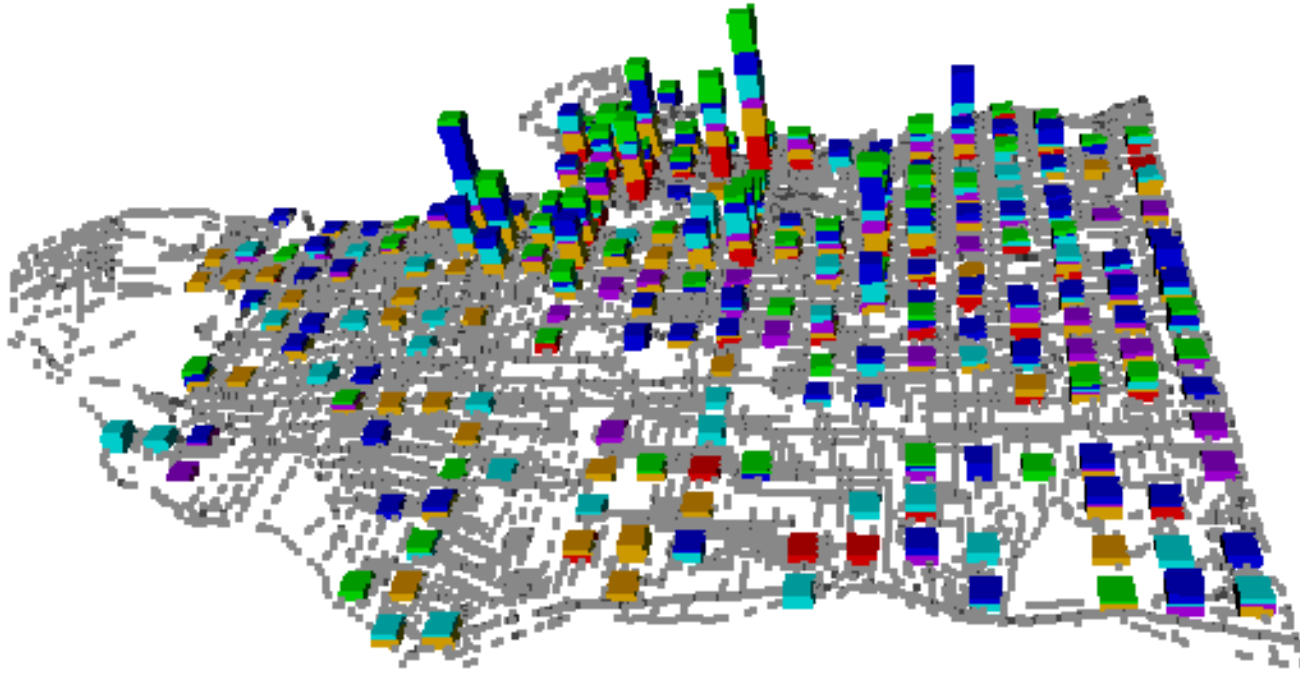
# Openness

🔑 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

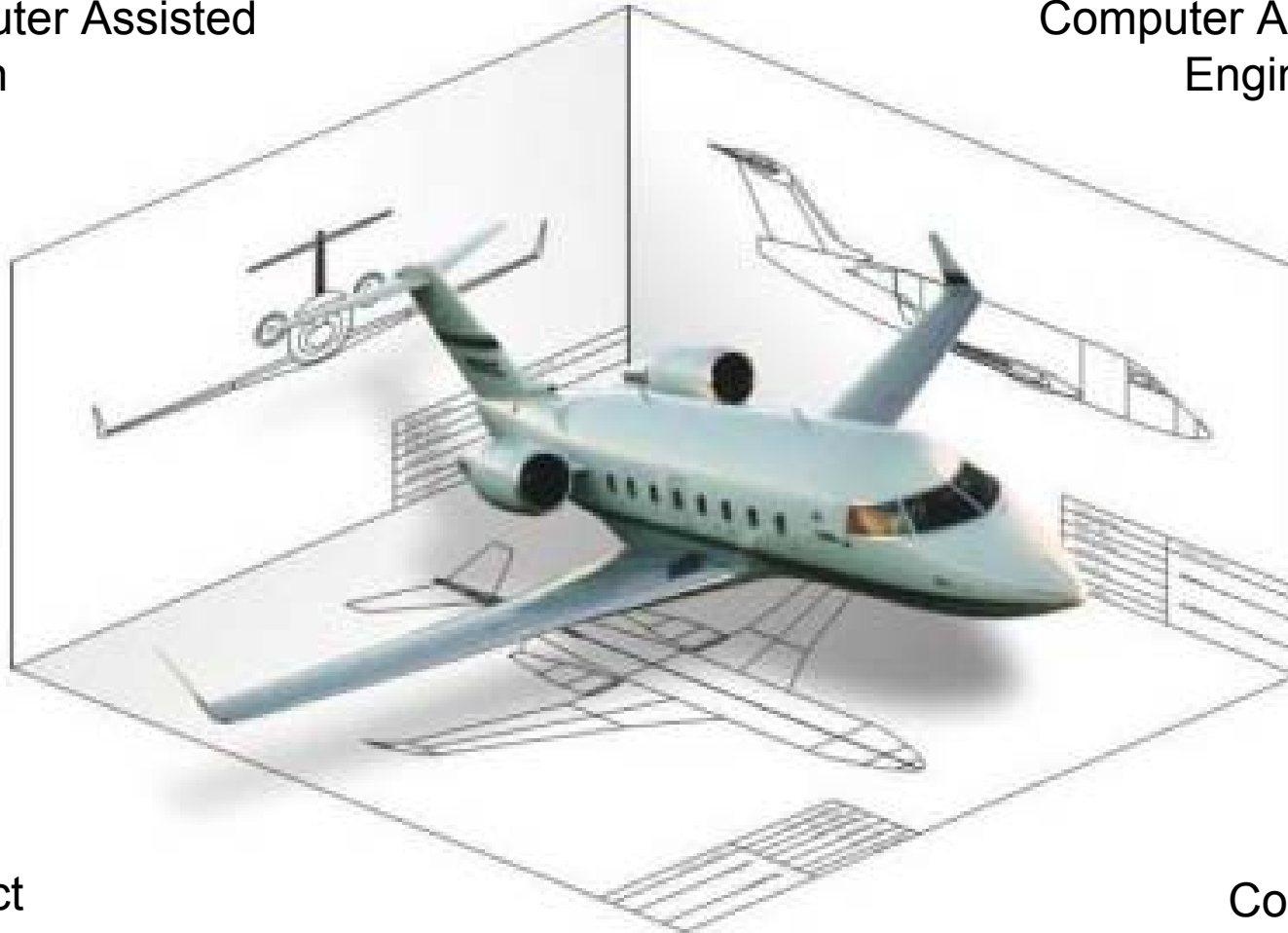




# 🔗 New Types of Records: Product Data

Computer Assisted  
Design

Computer Assisted  
Engineering



Product  
Analysis and Testing

Computer  
Assisted Manufacture



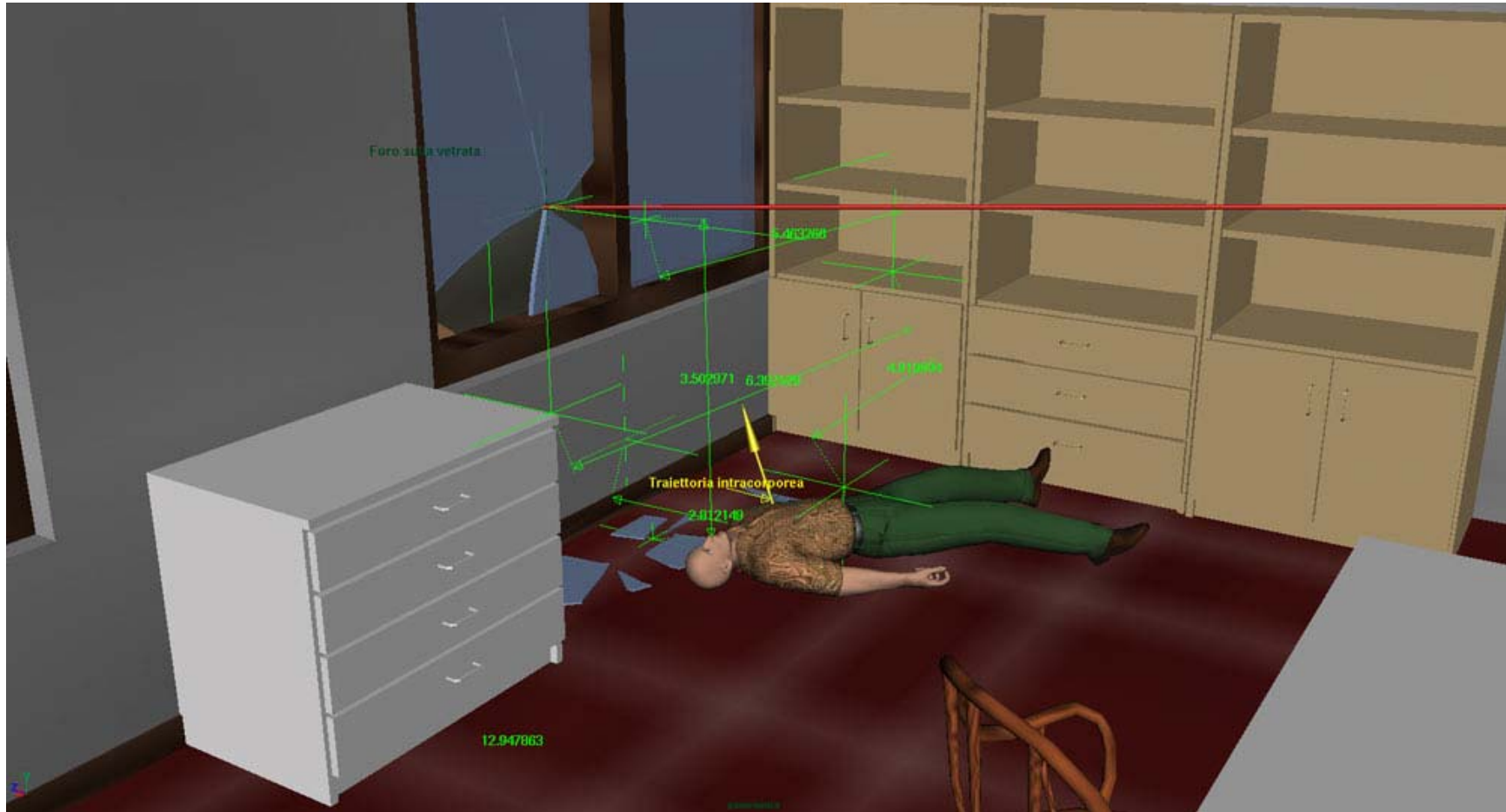
# New Types of Records: Critical Infrastructure Data



Source: CLindberg [http://commons.wikimedia.org/wiki/Image:I35\\_Bridge\\_Collapse\\_4crop.jpg](http://commons.wikimedia.org/wiki/Image:I35_Bridge_Collapse_4crop.jpg)

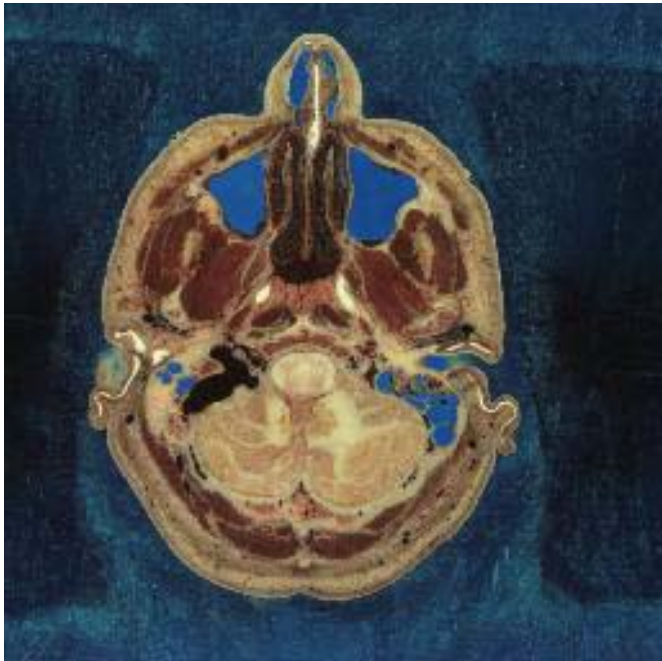


# New Types of Records: Virtual Reality: Crime Scene Investigation





# 🔓 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.

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.

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 of users and frequency of use



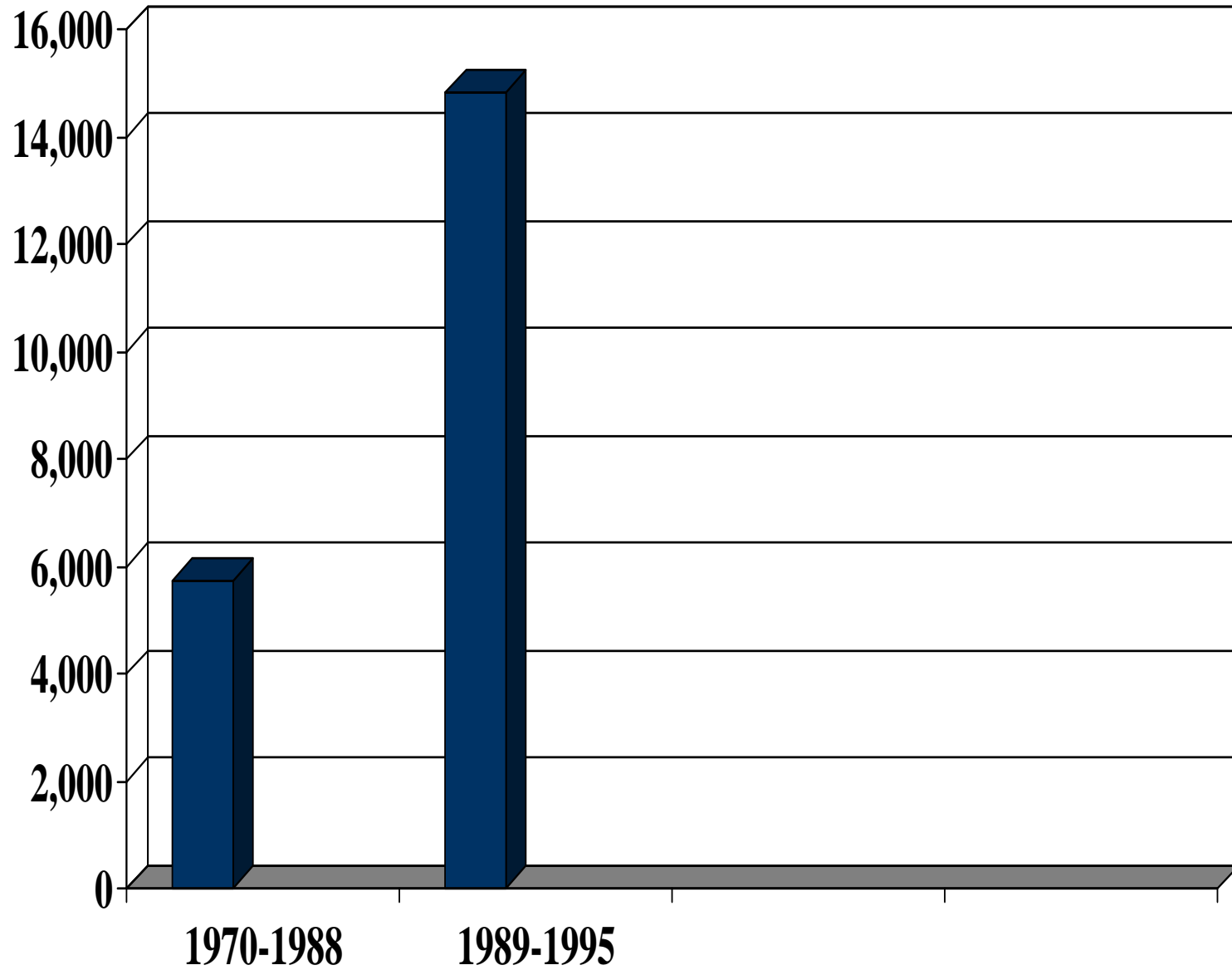
# Increasing Volumes of Digital Information

- 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

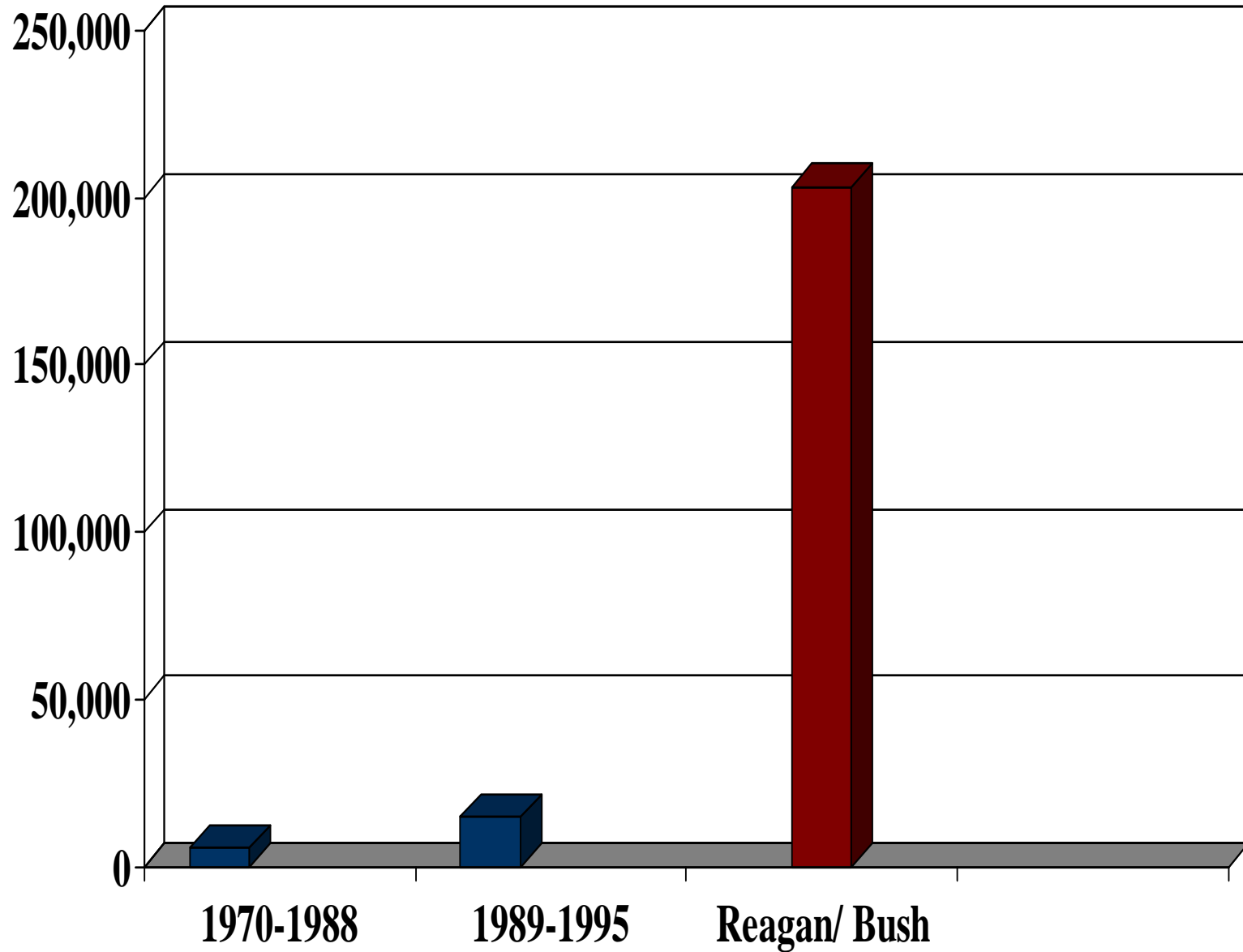


# Transfers of Digital Files to NARA



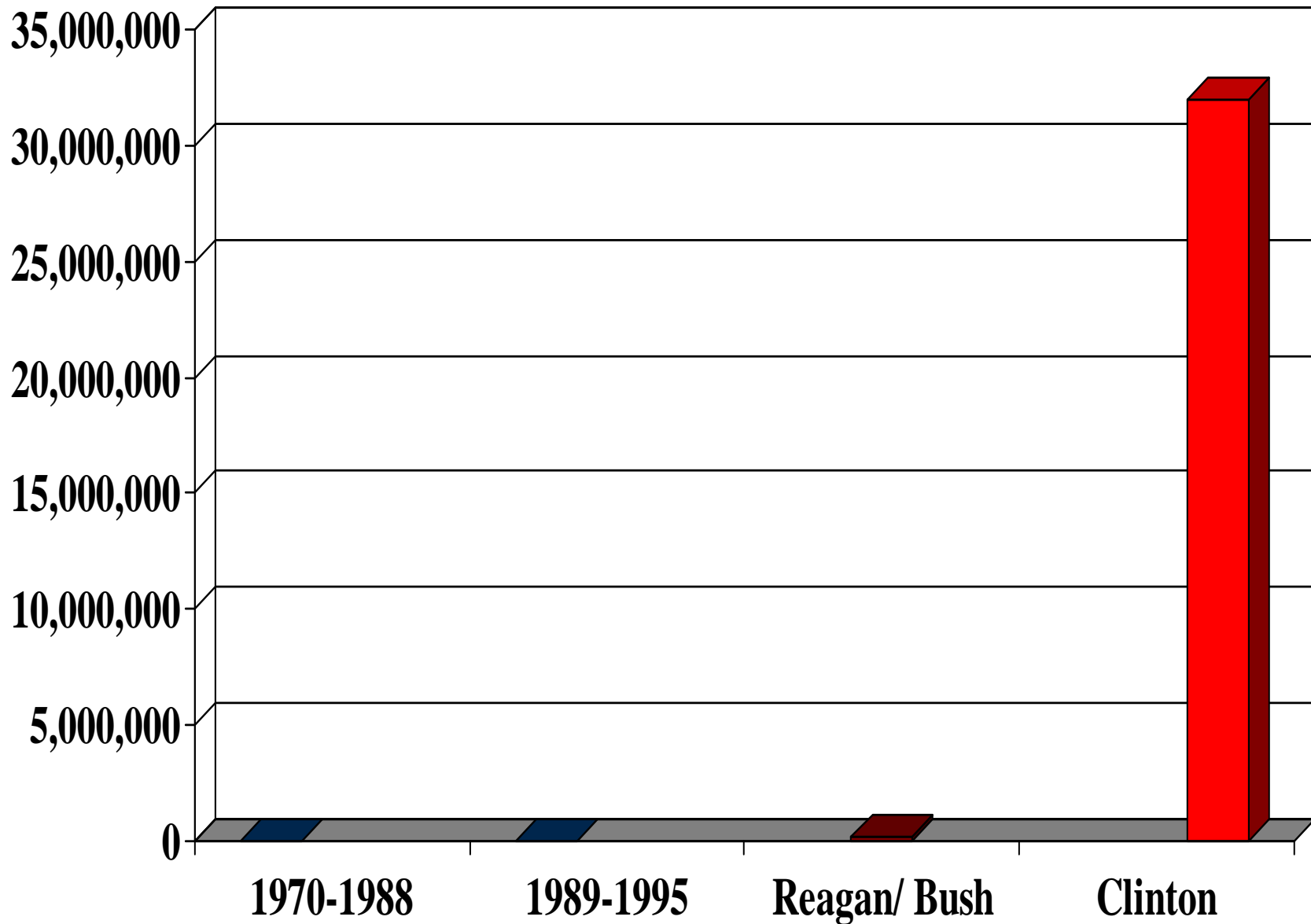


# Transfers of Digital Files to NARA



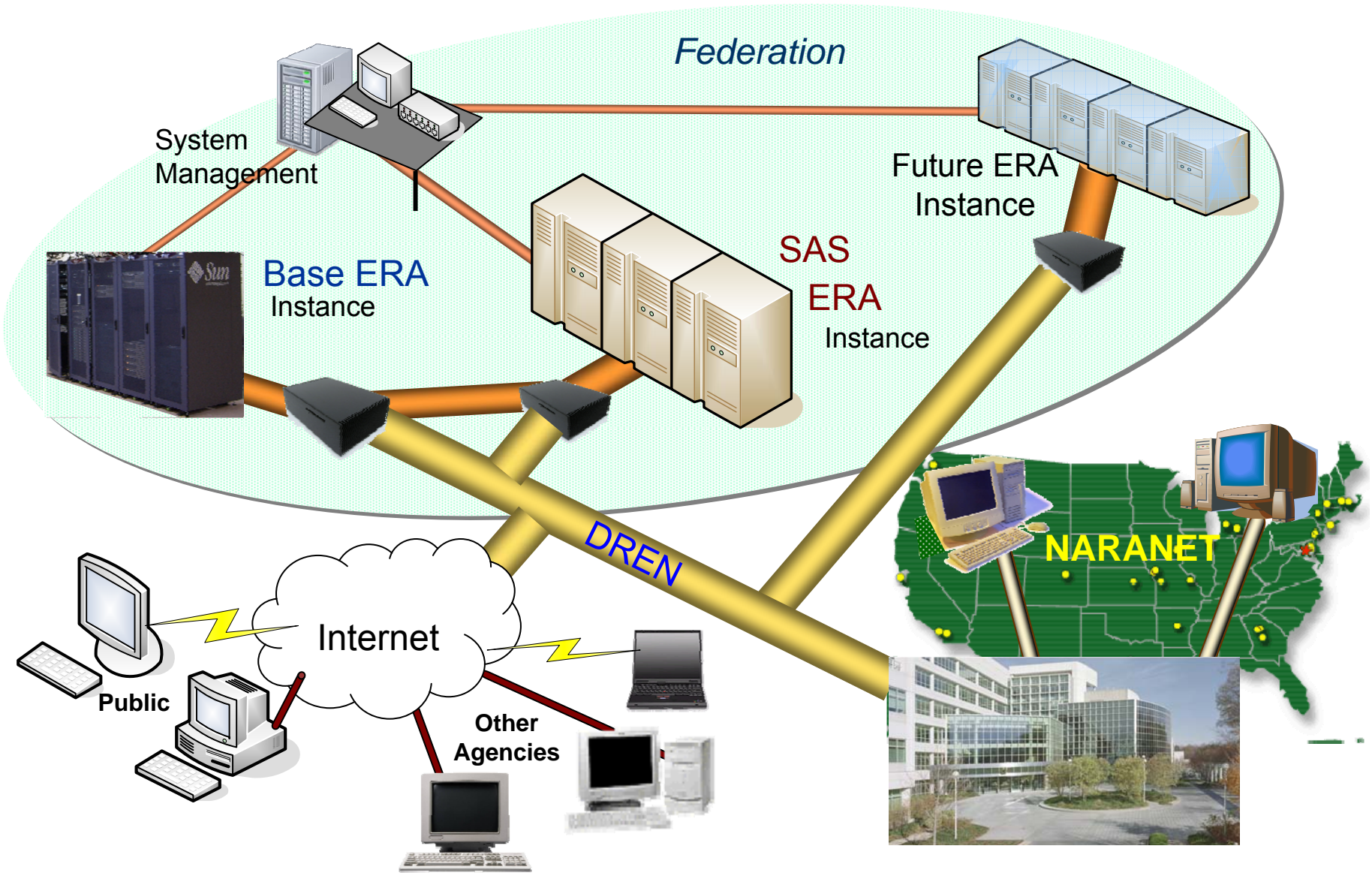


# Transfers of Digital Files to NARA





# 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



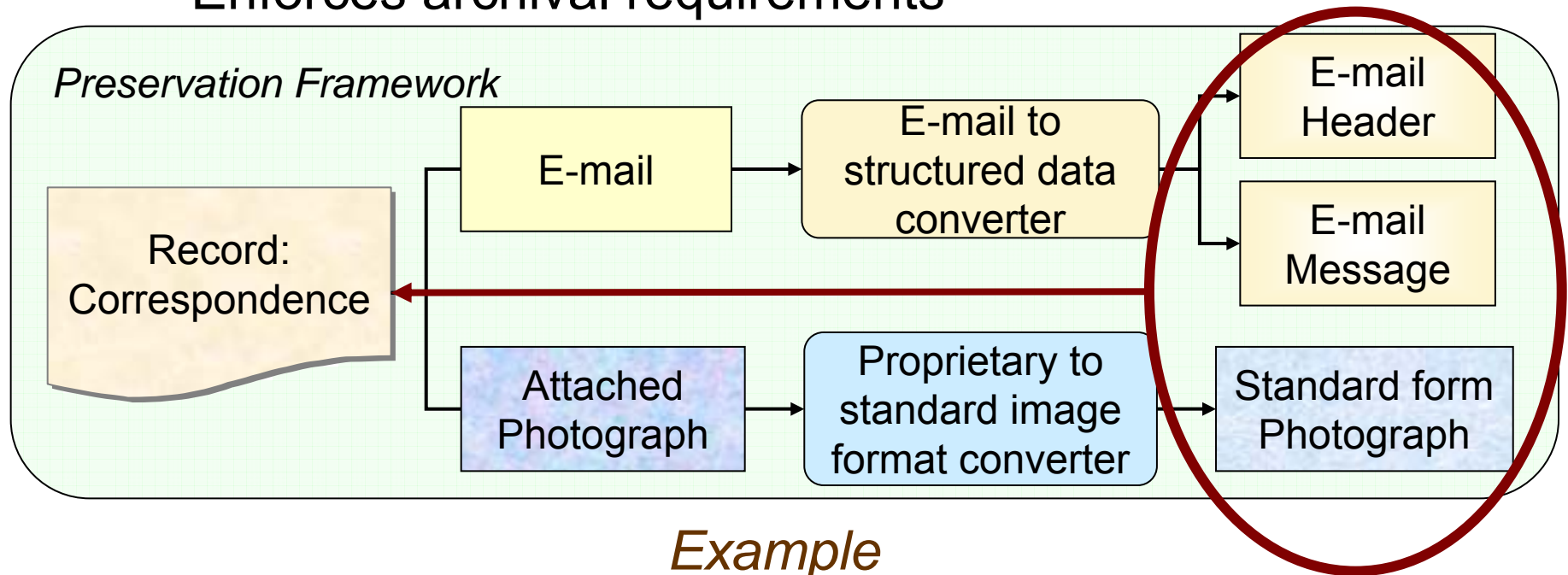
# Obsolescence of Formats of Electronic Records

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

Open

Grow

Close

Evolve

# Evolution

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

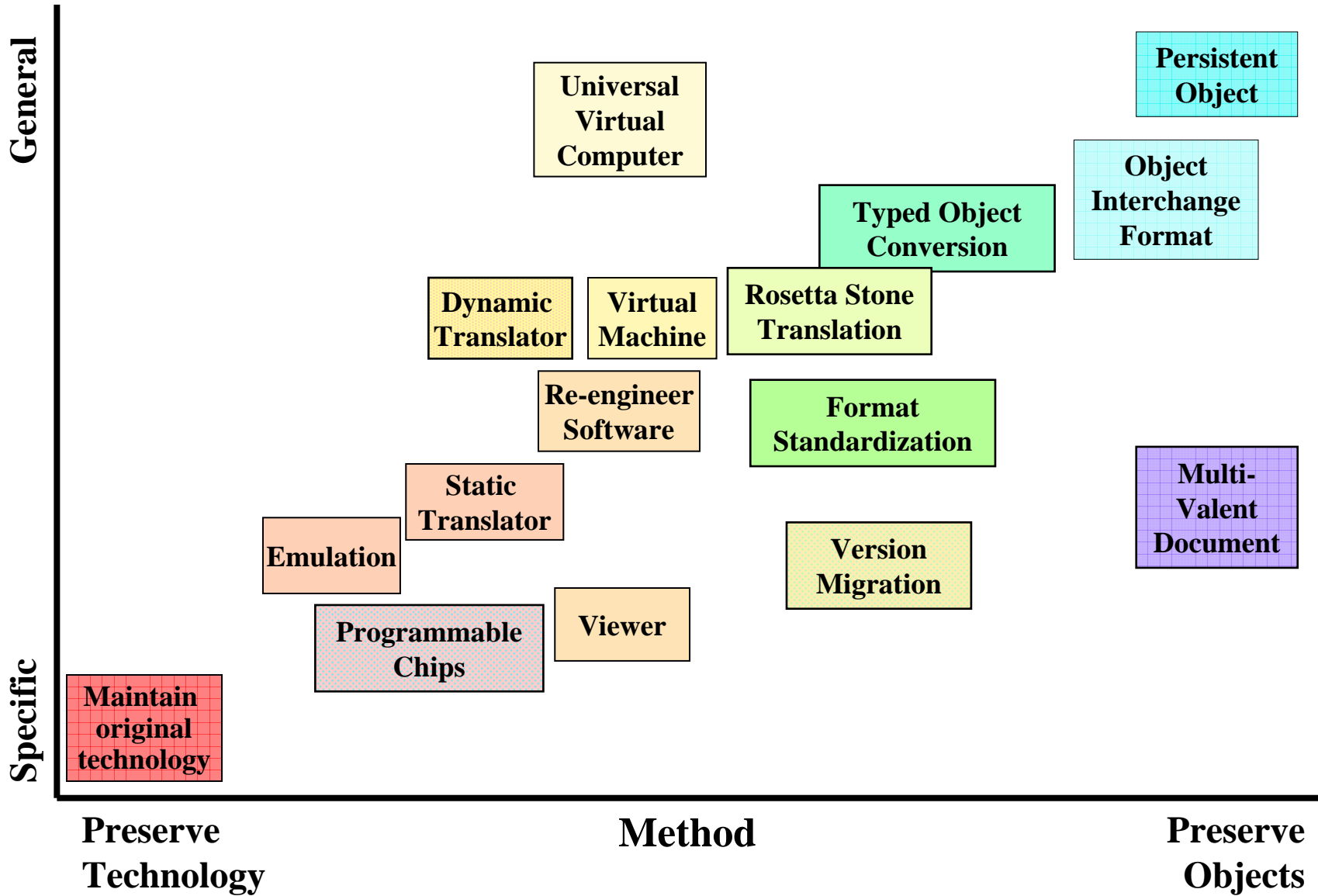
🔗 Changing Information Technology

- Obsolescence
- Opportunities

🔗 Changing business requirements

Evolve

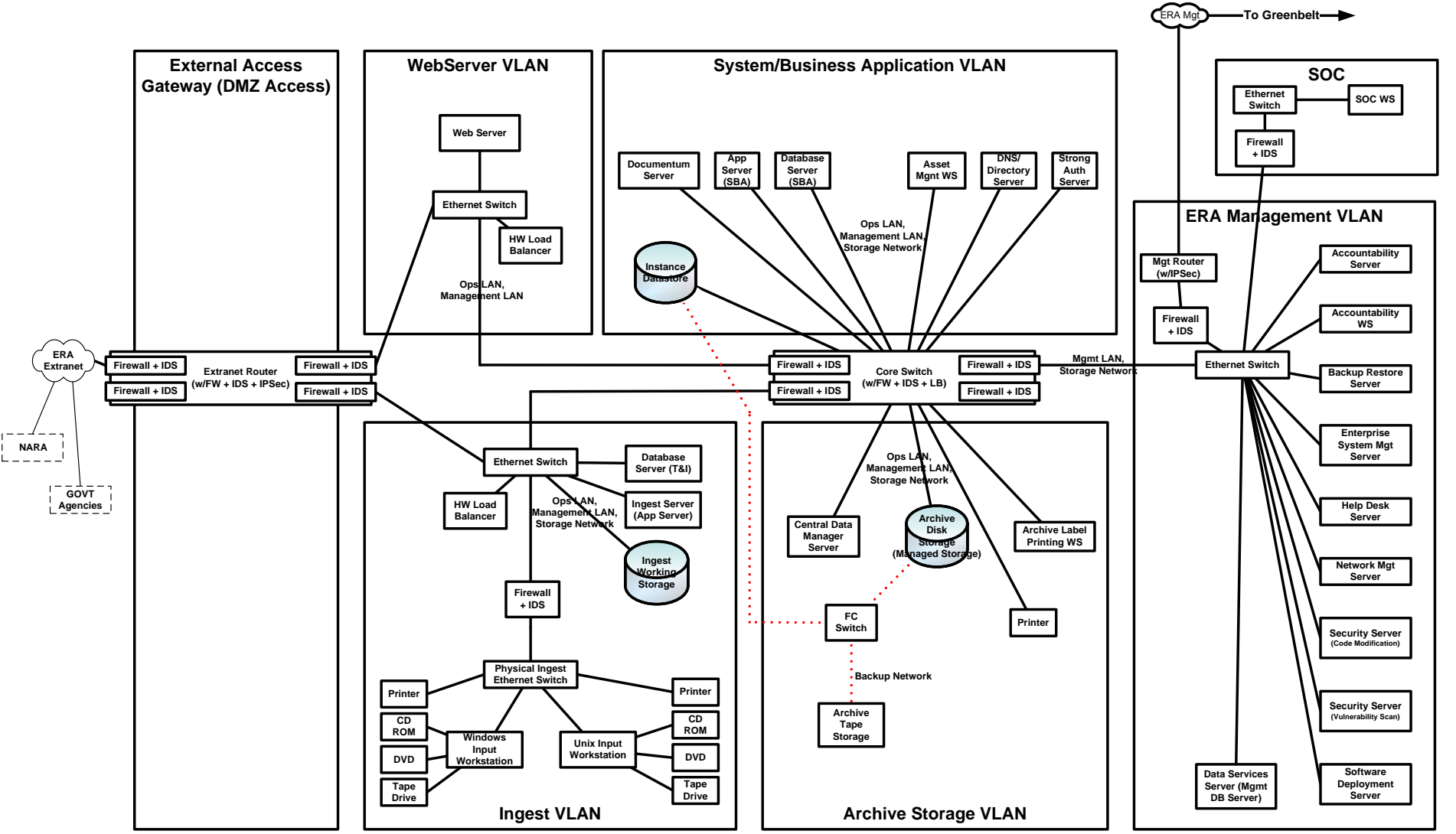
# Preservation Options



Evolve



# Changing Information Technology: Service Oriented Architecture

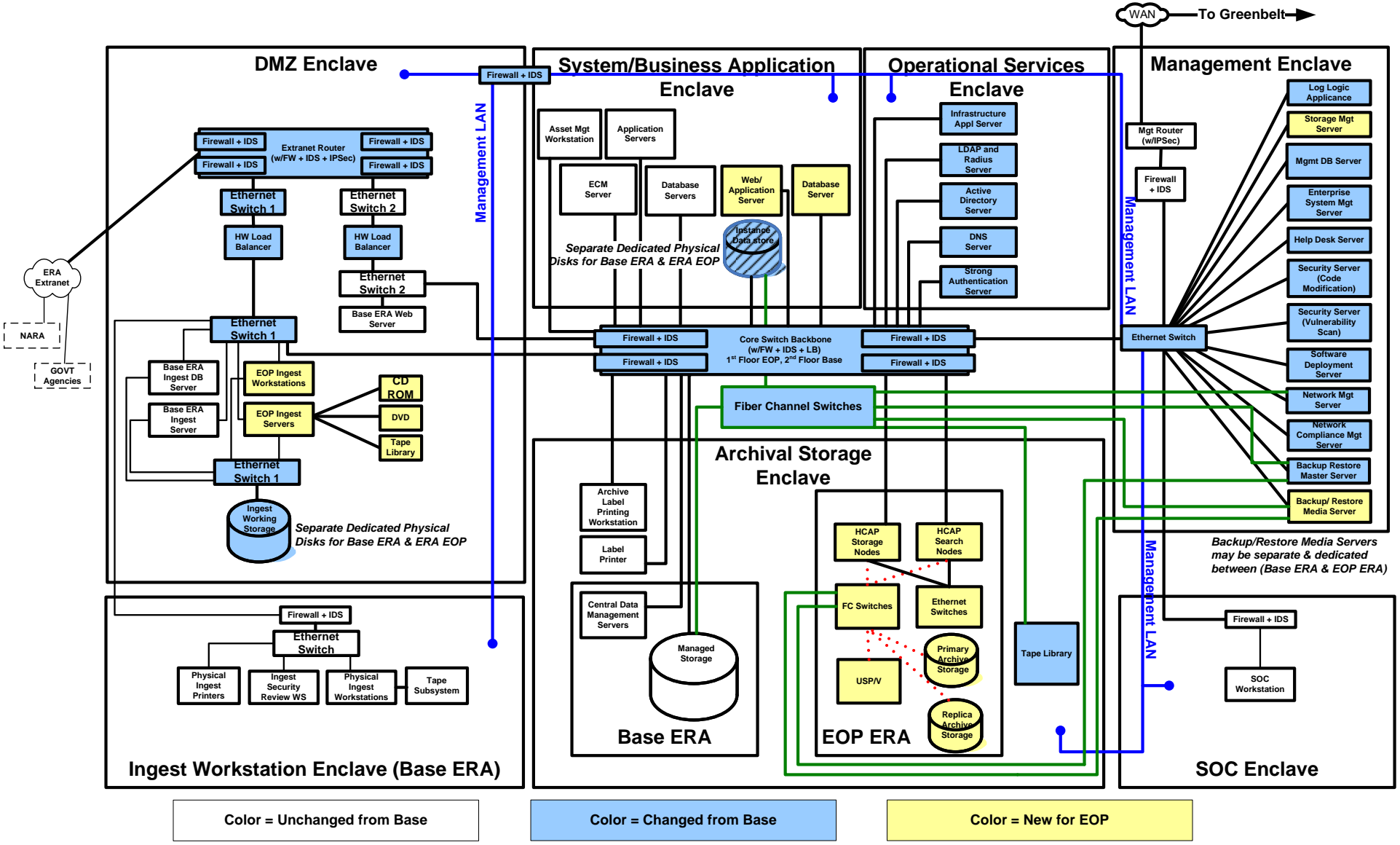


Derived from:  
ERA Hardware Block Diagram -2007 0823  
(Tab: 11R2 U/USBU Detailed Block)  
Updated 24 Aug 2007

1Gb Ethernet ———  
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

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

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

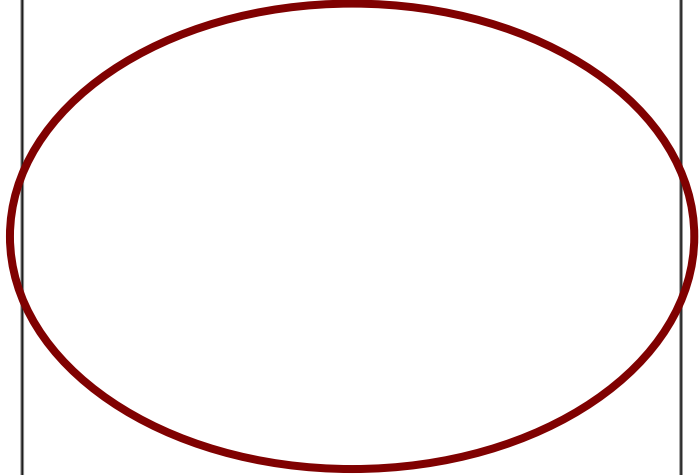


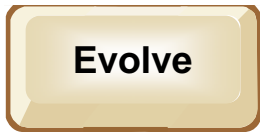
# 🔗 Evolution of Business Requirements



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		<b>Notification to Agency</b> 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
<b>6. Agency Certification</b> 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

**General**

Item ID:  \*Title:

\*Description:

Does agency have an associated manual?  Yes  No

Records Schedule ID: DAI-PENDING-2008-0051

Legacy Data: No

\*Manual ID:

\*Manual Version:

\*Manual Item ID:

\*GAO Concurrence Required:

Is this a change to an approved schedule?  Yes  No

Is this item media neutral?  Yes  No

Do any of the records covered by this item currently exist in electronic format(s) other than e-mail and word processing?  Yes  No

**Final Disposition**

\*Final Disposition:  Permanent  Temporary

**Temporary Disposition Instructions**

Cutoff Instructions:

**Transfer Instructions**

Transfer to:  Time after cutoff when transfer occurs:

**\*Retention Period**

Destroy immediately on cut-off

Destroy  after cut-off

Destroy between  years and  years after cut-off

Retain at least  years after cut-off, but longer is authorized

Retain no more than  years after cut-off

Destroy when no longer needed

Destroy  years after cut-off or when  occurs, whichever is sooner

Destroy  years after cut-off or when  occurs, whichever is later

Destroy  years after cut-off or  years after  occurs, whichever is sooner

Destroy  years after cut-off or  years after  occurs, whichever is later

Other

## Permanent Records

**General**

\*Title:  Item ID:

\*Description:

Does agency have an associated manual?  Yes  No

Records Schedule ID: DAI-PENDING-2008-0044

Legacy Data: No

\*Manual ID: Records Management Handbook

\*Manual Version: Version 1.0

\*Manual Item ID: A240314

\*GAO Concurrence Required:

Is this a change to an approved schedule?  Yes  No

Is this item media neutral?  Yes  No

Do any of the records covered by this item currently exist in electronic format(s) other than e-mail and word processing?  Yes  No

**Final Disposition**

\*Final Disposition:  Permanent  Temporary

**Permanent Disposition Instructions**

\*Cutoff Instructions:

**Transfer Instructions**

Records to which these transfer instructions apply:

\*Transfer to:

\*Time after cutoff when transfer occurs:

\*Estimated First Transfer:

**\*Accession Instructions**

Accession immediately on cut-off

Accession  after cut-off

Accession between  years and  years after cut-off

Accession in  year blocks  years after cutoff or most recent records in the block

Other

\*Estimated First Transfer:

\*If records are not transferred to NARA physical custody when legal custody is transferred, specify institution that will maintain physical records:

**Additional Information**

**\*Estimated Current Volume**

Electronic/Digital:

Paper:  cubic feet

Microform:  microfiche  microfilm

Traditional Special Media: Units:

Unknown:

**Annual Accumulation**

Electronic/Digital:

Paper:  cubic feet

Microform:  microfiche  microfilm

Traditional Special Media: Units:

Unknown:

**Date Span**

\*First year of records accumulation:

\*End year of records accumulation:

Records ceased accumulation in

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



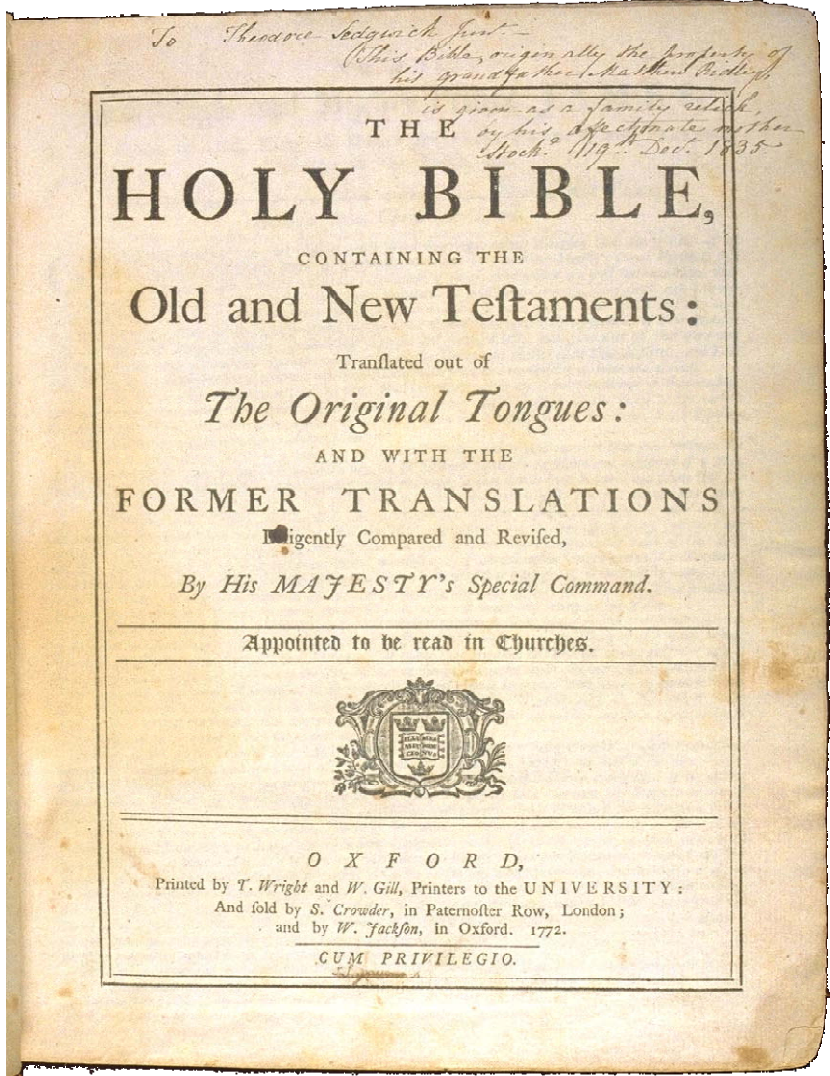


# Document v. Record

- A **document** is a bounded physical representation of a 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](http://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.inter pares.org/ip2/ip2\\_terminology\\_db.cfm](http://www.inter pares.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?

Close

# Record

Ebenezer Prior of Fairfield in the County of  
 Hartford in the State of Connecticut of lawful age  
 do hereby testify that I was before the  
 Revolutionary War, well acquainted with  
 Ebenezer Prior formerly of Danbury in the  
 name of Fairfield that in the year 1779 he  
 purchased a Town of Danbury at New London  
 in a Company of Militia commanded by  
 Capt. Ebenezer Prior in a Regiment  
 commanded by Col. Jonathan Wells of  
 Hartford that I served in the same  
 Company & was resolved that I should  
 serve there as afore said but cannot say how  
 long I further he said to be

Ebenezer Prior

subscribed & sworn to the day & year afore  
 said in the presence of  
 Wm. D. [unclear] & [unclear]

And this I do certify to be true & correct  
 after the best of my knowledge & belief  
 the interposition of the War Office  
 that the above named Ebenezer Prior  
 is the same person as he is called to the Court  
 & that the preceding Report of Ebenezer  
 Prior was taken in open Court & that Ebenezer  
 Prior is known to the Court as a credible person &  
 that the Report of Charles Prior & Ebenezer  
 Prior & also Ebenezer Prior's name  
 by the affidavit of Ebenezer Prior upon a Request  
 duly authorized to take the same & that the  
 Charles Prior & Ebenezer Prior are known to the  
 Court as credible persons Wm. D. [unclear] Judge

3-1789

SERVICE

Prior Ebenezer

NUMBER

W 17496

Mary

Mary

**Register**

Of Ebenezer Prior, FAMILY He was MARRIED

BORN May 21 1745

Mary Thompson She was born Oct 21 1744

by her had the following Children viz.

NAME	BORN	DIED
Thompson Prior	May 10 1775	April 8 1795
Mary Prior	March 30 1776	September 15 1812
Married Sarah Child	Oct 15 1779	
Charlotte Prior	March 25 1780	February 8 1789
Married Rufus Stillings	Jan 10 1781	
Harriet Prior	November 20 1785	

M. Ebenezer Prior M.  
 DIED January 12 1811 DIED

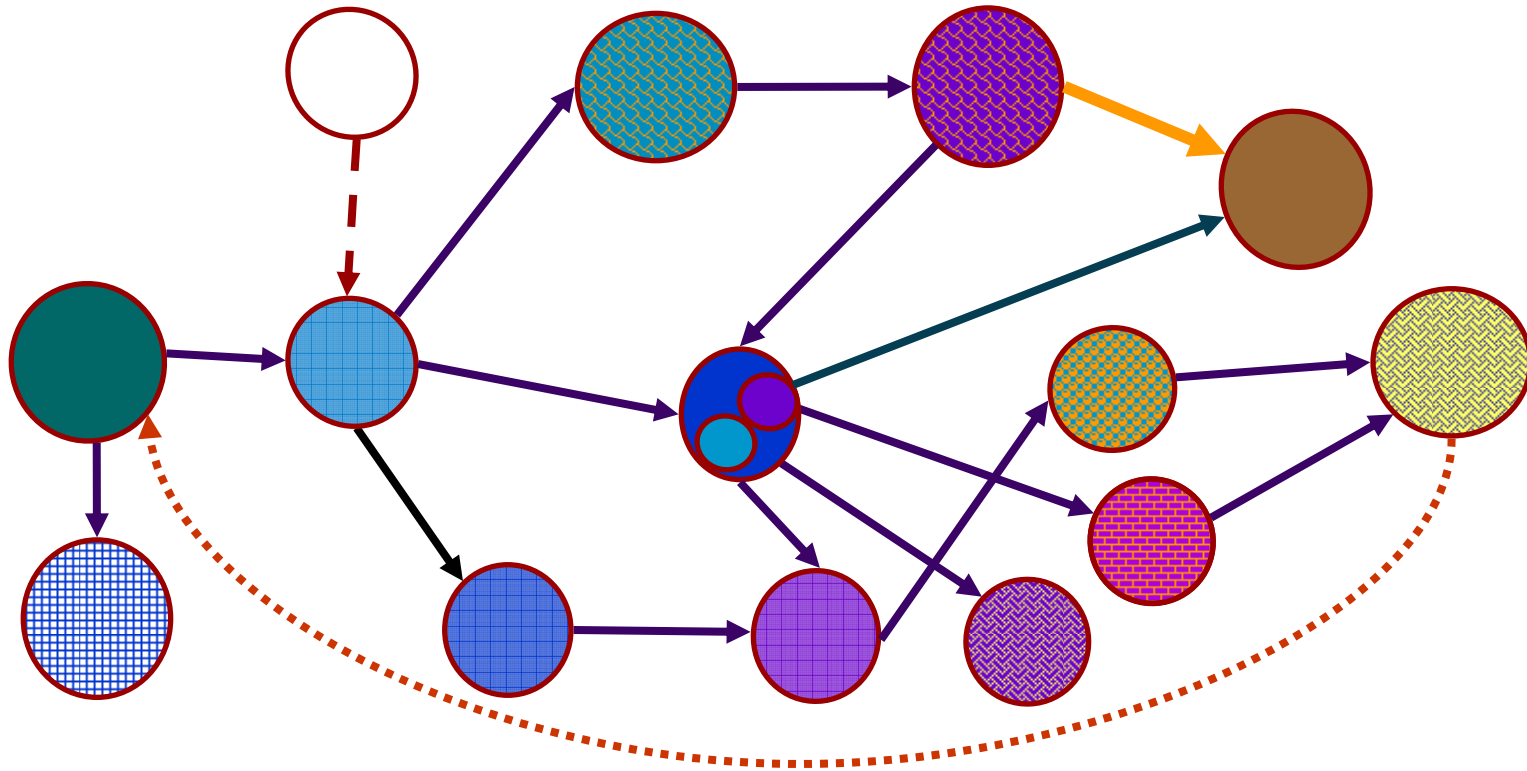
I Ebenezer Prior of Fairfield in the County of  
 Hartford State of Connecticut of lawful age  
 do hereby testify that I was well acquainted with  
 Ebenezer Prior formerly of Danbury now residing  
 in Danbury in the State of Massachusetts  
 before I during the Revolutionary War, that  
 in the year 1779 in the month of August  
 I & Sarah Wells went to New London in  
 a Company of Men drafted from the Militia  
 under the command of a Capt. Ebenezer  
 Nelson or Patton Ebenezer Wells Capt. L  
 Samuel Pancraft Surgeon that he  
 said Ebenezer Prior had been absent two months  
 according to my best recollection that  
 Col. Jonathan Wells of East Hartford  
 commanded the Regiment to which he belonged  
 that I served in the same Company with  
 him during the War as a Sergeant  
 I further do depose and testify that

Ebenezer Prior

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.



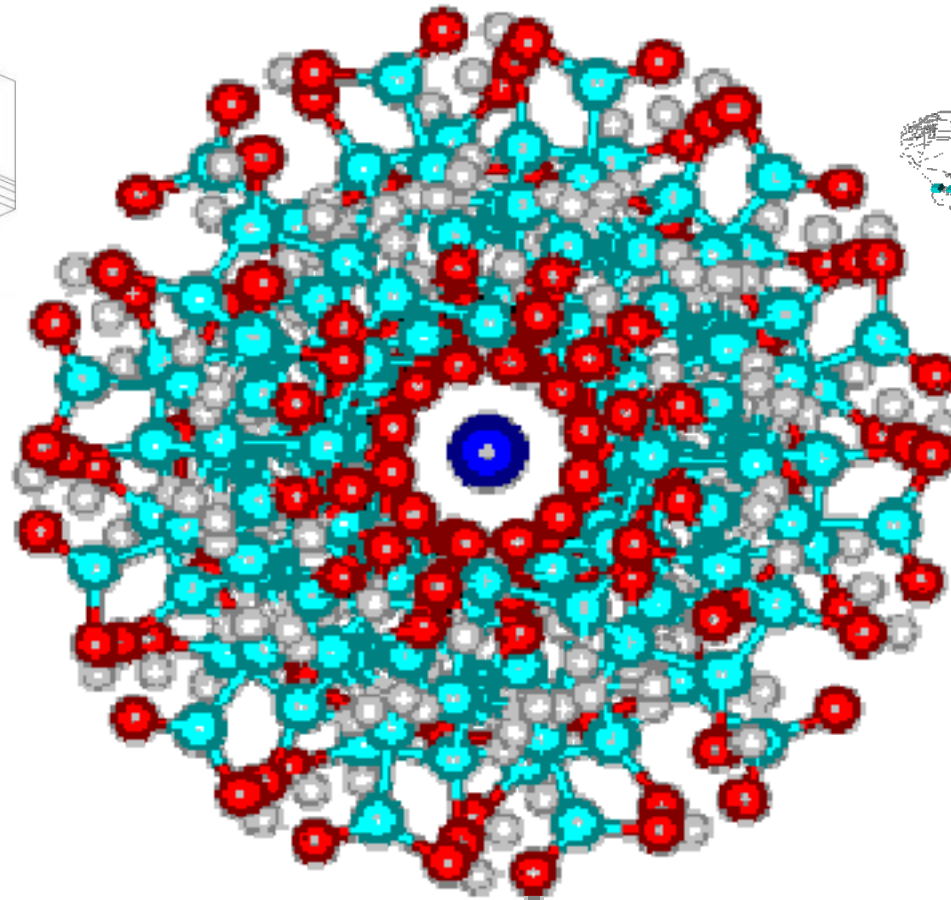
# Preservation

- 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 Packagesare managed to respect the original order of records.

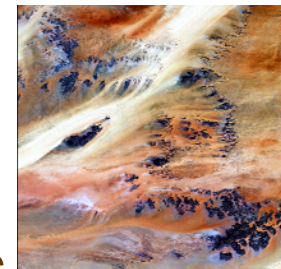
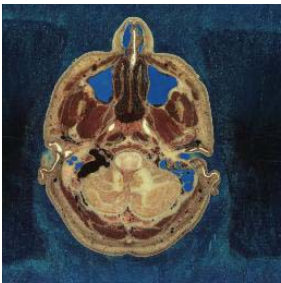
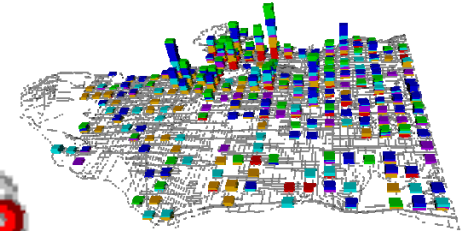
Close

# Electronic Records

Open



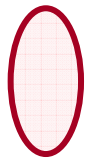
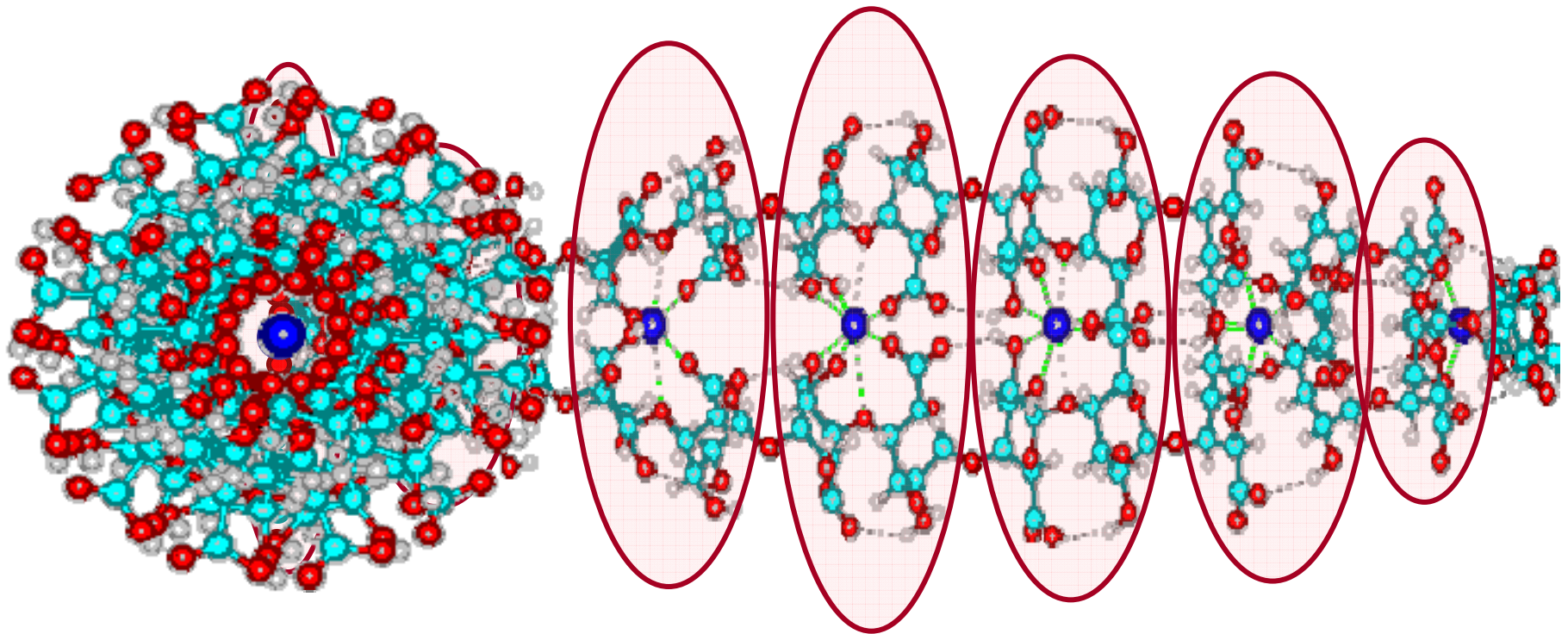
Start ASLI FAMTB TFAUTO  
BNE MSCHE THEFT



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

# SUMMARY

## ➤ Openness

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

## ➤ Growth

- 🔗 Exponential increase in volumes of data
- 🔗 Increasing users and usage

## ➤ 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](http://www.archives.gov/era)