

# NARA's Electronic Records Archives Program



Presentation to the  
National Digital Strategy Advisory Board  
Library of Congress

June 28, 2004

Kenneth Thibodeau, Ph.D.  
Director

Electronic Records Archives Program

**National Archives and Records Administration**

# NARA Mission

- “NARA ensures, for the citizen and the public servant, for the President and for the Congress and the Courts, ready access to essential evidence.”
  - *essential evidence*: material generated by or received by the Federal Government, that documents
    - the rights of citizens,
    - the actions of Federal officials, and
    - the national experience.

# NARA's Electronic Records Challenge

- Preserve any type of record,
- Created using any type of application,
- On any computing platform
- From any entity in the Federal Government and any donor
- Provide discovery and delivery to anyone with an interest and legal right of access
- Now and for the life of the Republic, and
- Provide effective policies, strategies, standards, guidance, and tools for federal agencies to manage electronic records in support of their lines of business

# The Electronic Records Archives Vision

“The Electronic Records Archives will authentically preserve and provide access to any kind of electronic record, free from dependency on any specific hardware or software, enabling NARA to carry out its mission into the future.”





# The ERA Program

- Acquisition of the ERA system to
  - Support NARA's end-to-end process for lifecycle management of ***all federal records***
  - Preserve, and provide access to valuable ***electronic records***
- Organizational change management
  - Ensure NARA successfully implements the system to achieve its strategic goals
- Research and Exploratory Development
  - Address new challenges posed by continuing change in Information Technology and its use in government
  - Capitalize on progress in technology

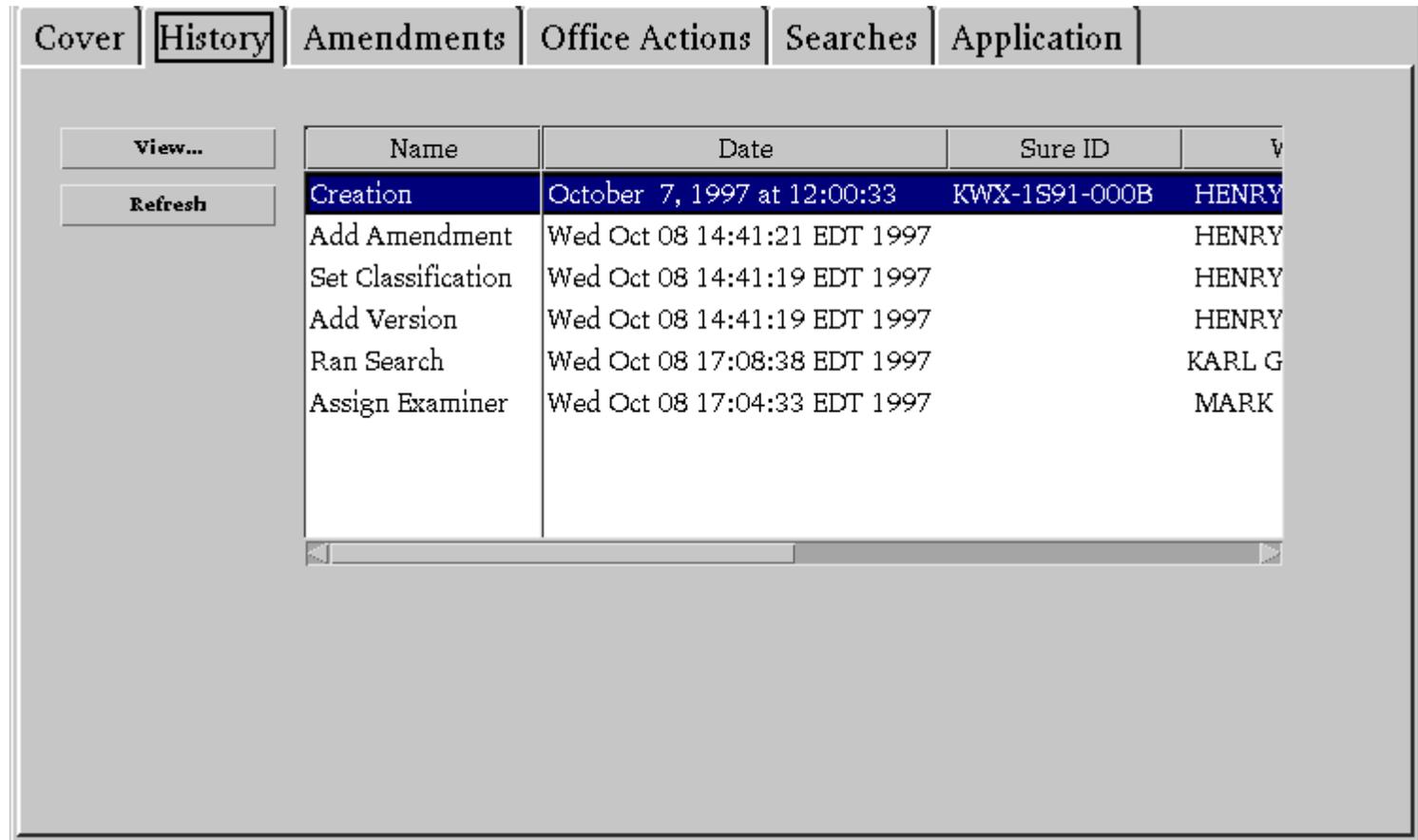


# NARA's Strategy

- Attack the critical preservation problem
- Define the requirements in terms of the lifecycle management of records
- Find solutions in commercially viable, mainstream technologies being developed to support e-commerce, e-government and the next generation national information infrastructure
- Align with overall direction of IT in the U.S. Government

# Some Specific Aspects of Preserving Records

# Records are preserved as members of ordered sets



The screenshot shows a software interface with a menu bar at the top containing 'Cover', 'History', 'Amendments', 'Office Actions', 'Searches', and 'Application'. The 'History' menu item is highlighted. Below the menu bar, there are two buttons: 'View...' and 'Refresh'. The main area contains a table with the following data:

Name	Date	Sure ID	v
Creation	October 7, 1997 at 12:00:33	KWX-1S91-000B	HENRY
Add Amendment	Wed Oct 08 14:41:21 EDT 1997		HENRY
Set Classification	Wed Oct 08 14:41:19 EDT 1997		HENRY
Add Version	Wed Oct 08 14:41:19 EDT 1997		HENRY
Ran Search	Wed Oct 08 17:08:38 EDT 1997		KARL G
Assign Examiner	Wed Oct 08 17:04:33 EDT 1997		MARK

e.g. Patent Application Files

# Sets of records can span long time periods

United States Patent: 6,597,288 - Microsoft Internet Explorer provided by NARA

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Search Favorites Media Mail Print Mailbox

Address <http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=/netahtml/search-bool.html&r=1&f=G&l=50&co1=AND&> Go

**References Cited** [Referenced By](#)

---

**U.S. Patent Documents**

<a href="#">4195286</a>	Mar., 1980	Galvin.	
<a href="#">4749986</a>	Jun., 1988	Otani et al.	340/587.
<a href="#">4871999</a>	Oct., 1989	Ishii et al.	340/587.
<a href="#">5260687</a>	Nov., 1993	Yamauchi et al.	340/522.
<a href="#">5592147</a>	Jan., 1997	Wong.	
<a href="#">5691703</a>	Nov., 1997	Roby et al.	340/628.
<a href="#">5767776</a>	Jun., 1998	Wong	340/628.
<a href="#">5818326</a>	Oct., 1998	Winterble et al.	
<a href="#">6154142</a>	Nov., 2000	Kosugi et al.	
<a href="#">6195011</a>	Feb., 2001	Winterble et al.	340/589.

**Foreign Patent Documents**

4-270493	Sep., 1992	JP.	
----------	------------	-----	--

*Primary Examiner:* Wu, Daniel J.  
*Attorney, Agent or Firm:* Oblon, Spivak, McClelland, Maier & Neustadt, P.C.

Internet

# Examples of Electronic Records

## NARA preserves

- Bureau of Economic Analysis, Dept. of Commerce
  - Regional Economic Information System Time Series, 1969-89.
- Watergate Special Prosecution Force, Department of Justice
  - Abstracts of Testimony, Interviews, Diaries, 1973-77.
- Military Assistance Command Vietnam, DoD
  - Psychological Operations Information System, 1970-73
- U.S. Army and Joint Services Environmental Support Group
  - Army Nuclear Testing (1945-1962) Personnel Review
- NARA
  - Electronic Army Serial Number Merged File, ca. 1938 - 1946
- Agency For Health Care Policy And Research,
  - National Medical Care Expenditure Survey, 1977, 1987
- Executive Office of the President
  - Automated Records Management System (email), 1994-2000

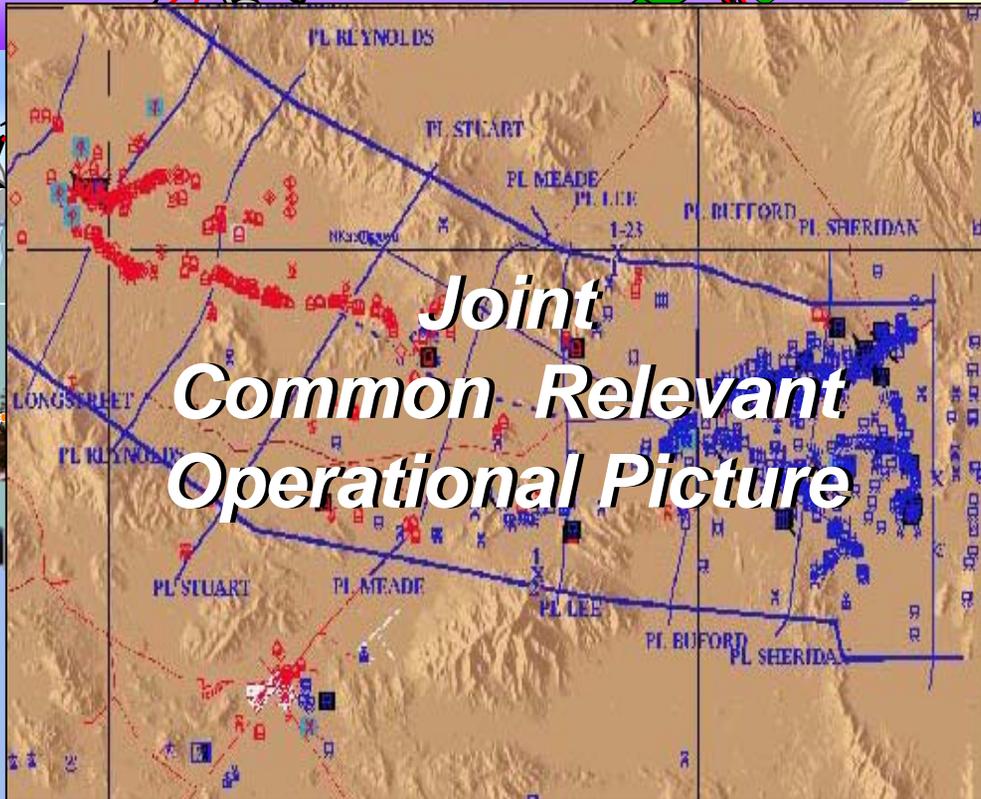


# Exploiting the Enterprise

Bandwidth Expansion



Joint, Interagency, & Multi-National Interoperability



**Joint  
Common Relevant  
Operational Picture**

Employment  
(Tactical Level)

Global Hawk

UE (XX)

Predator

Joint Tactical Network

Unit(s) of Action  
(Tactical Level)

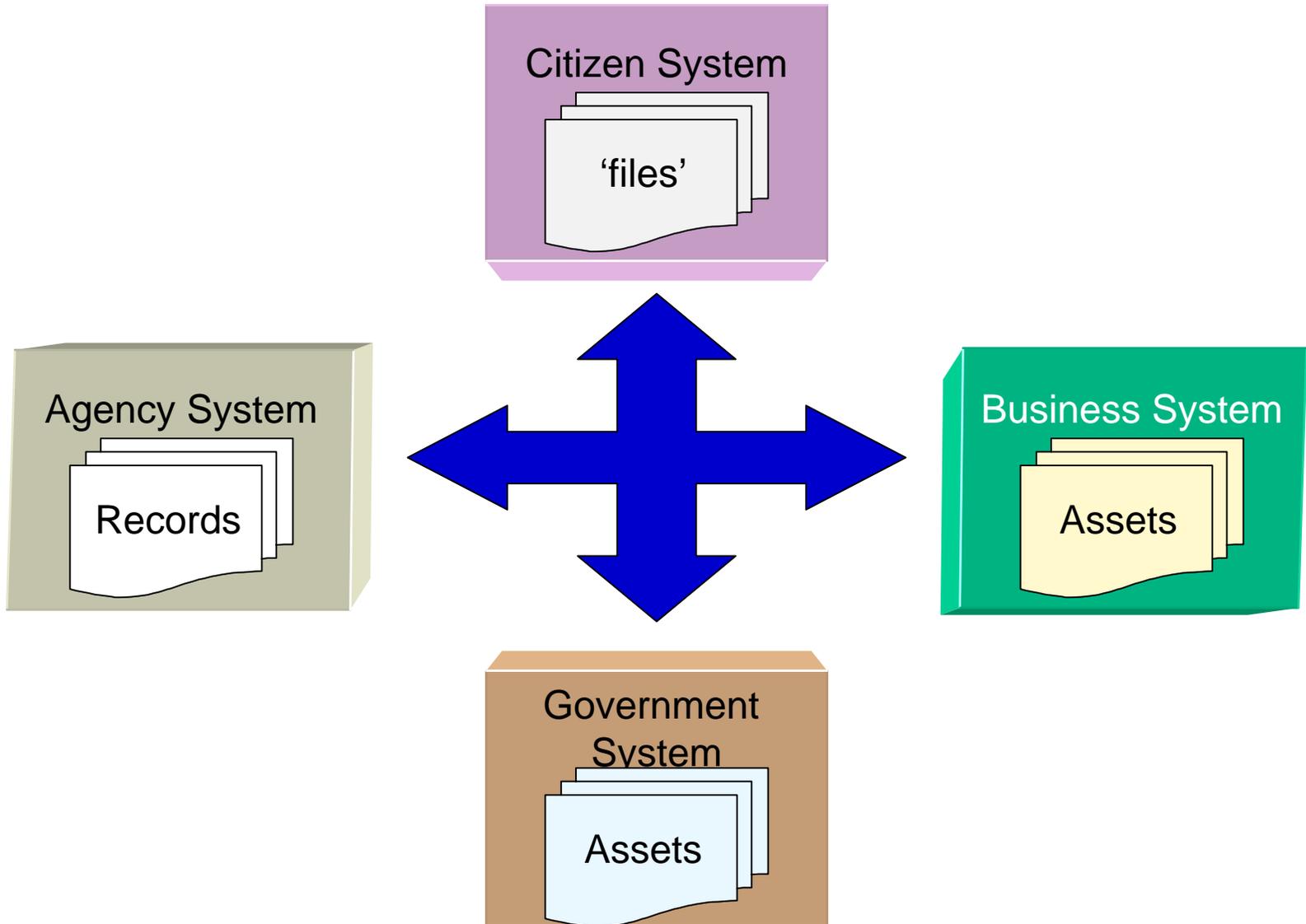
Unit of Employment  
(Tactical Level)



"Space to Mud"

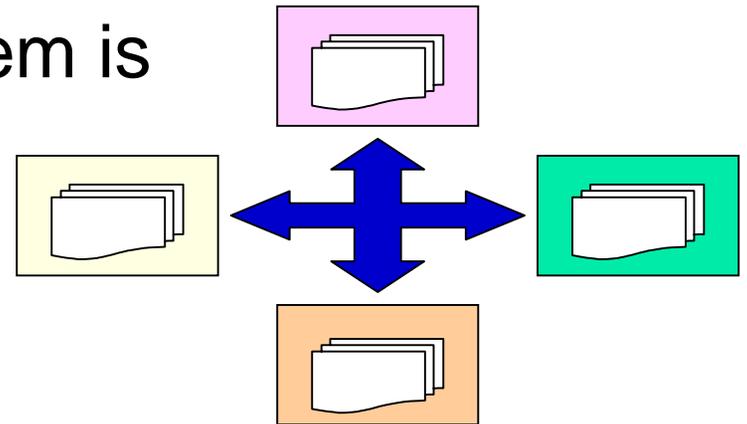
"Factory to Foxhole"

# E-transactions

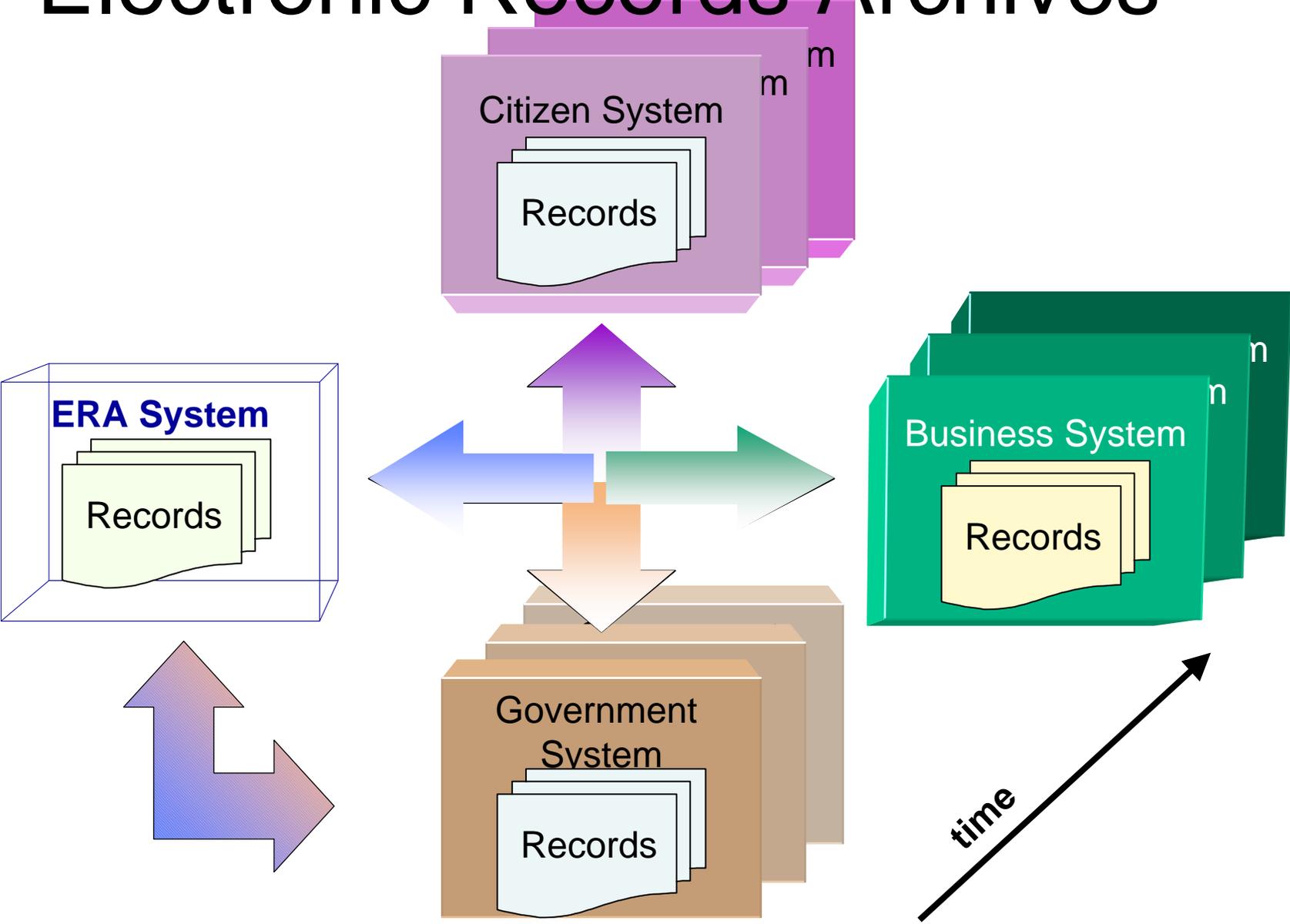


# E-transactions: basic assumptions

- ***Different*** systems interact in conducting e-transactions.
- Anything which must be true about an information asset in one system must be true about that asset in another system involved in the e-transaction.
- The only thing one system needs to know about another system is that it can use the same intermediary.

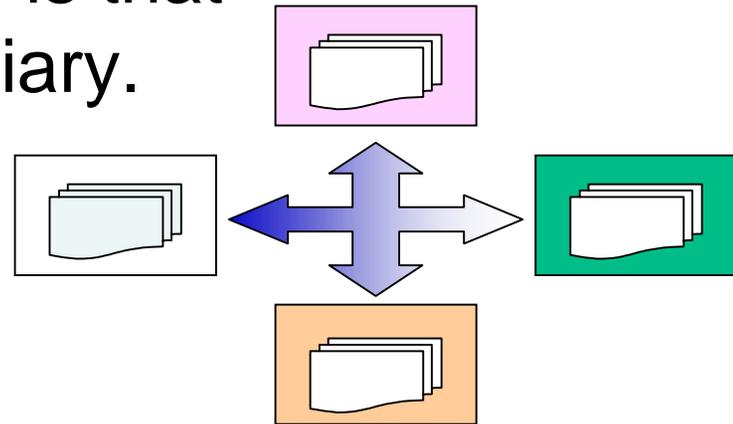


# Electronic Records Archives

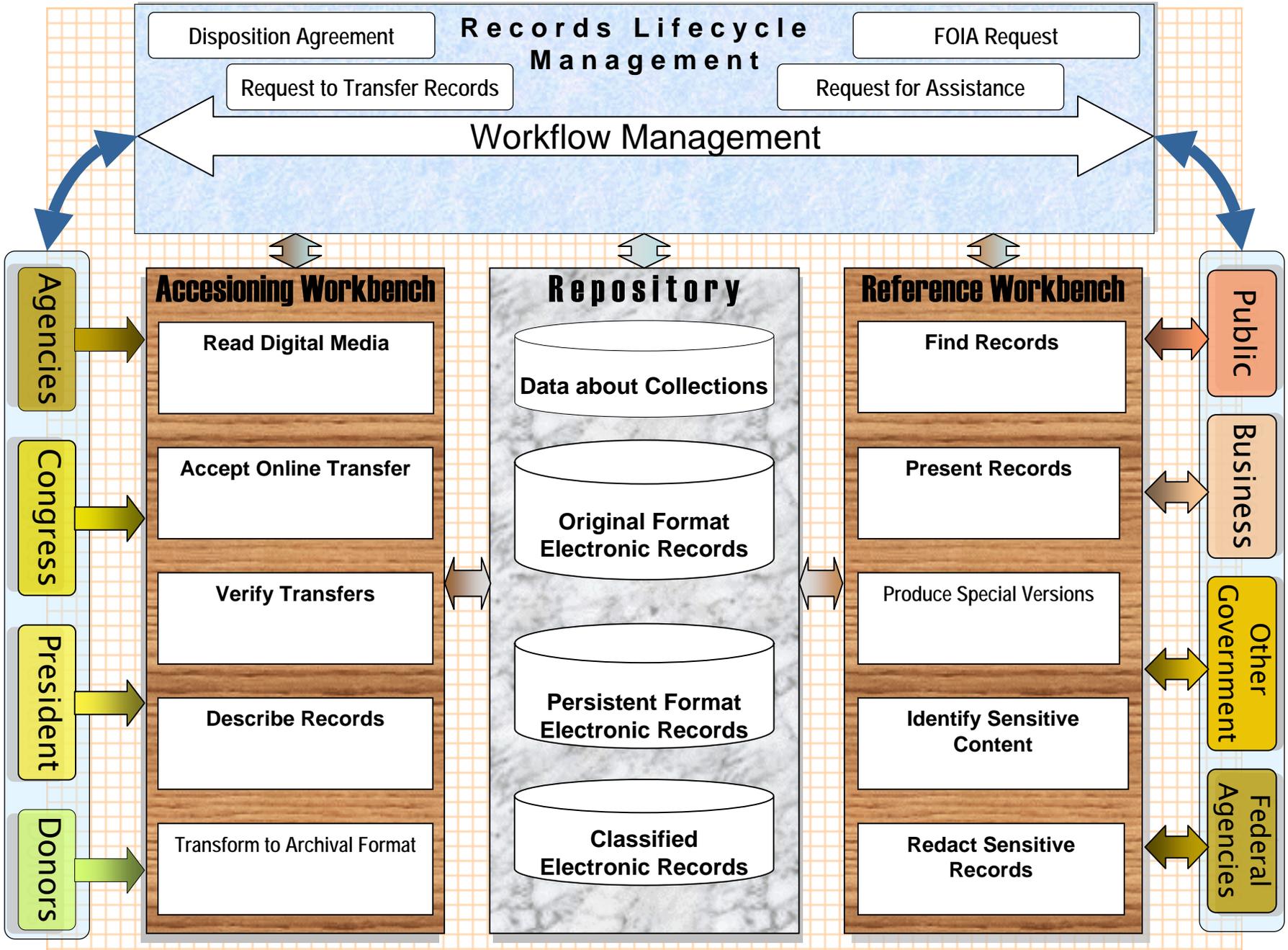


# Electronic Records Archives: basic assumptions

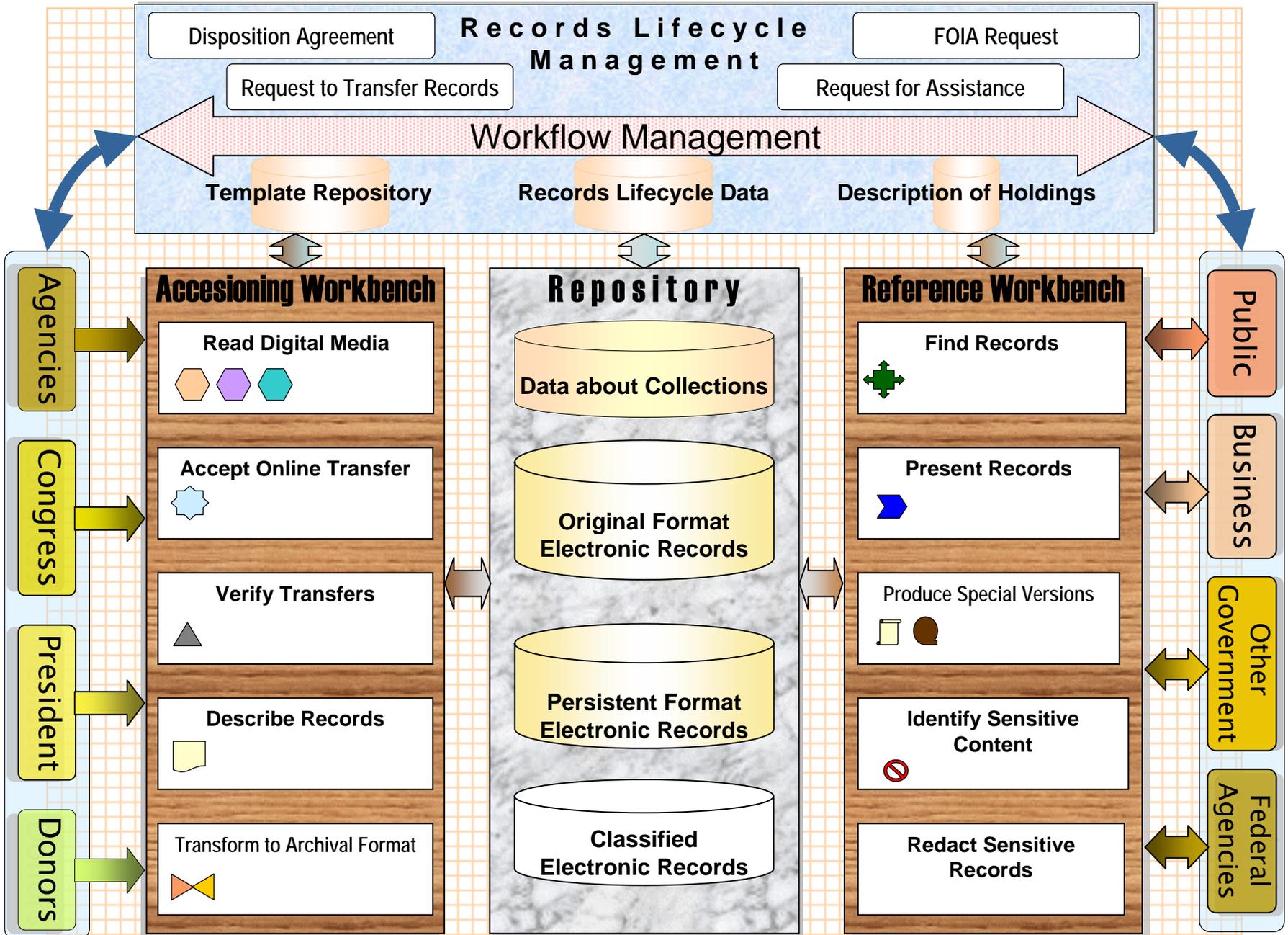
- The ERA system must be able to interact with ***different*** systems.
- Any record in the ERA system must be an authentic copy of that record.
- Any record delivered from ERA to another system must be an authentic copy.
- At any time, the only thing the ERA system needs to know about another is that it can use the same intermediary.
- Over time, the ERA system cannot know what mediators other systems will use.



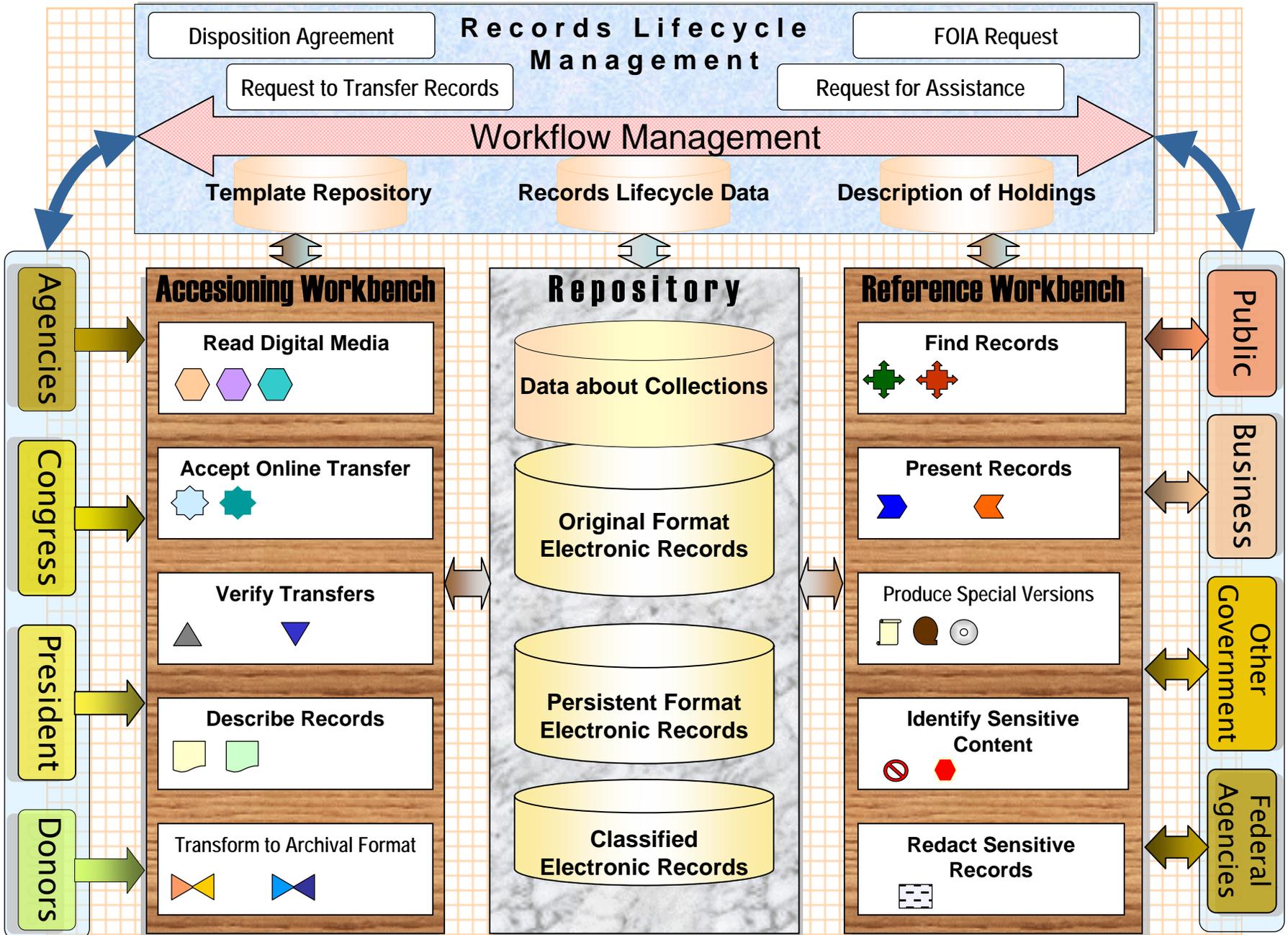
# ERA Virtual Workspaces



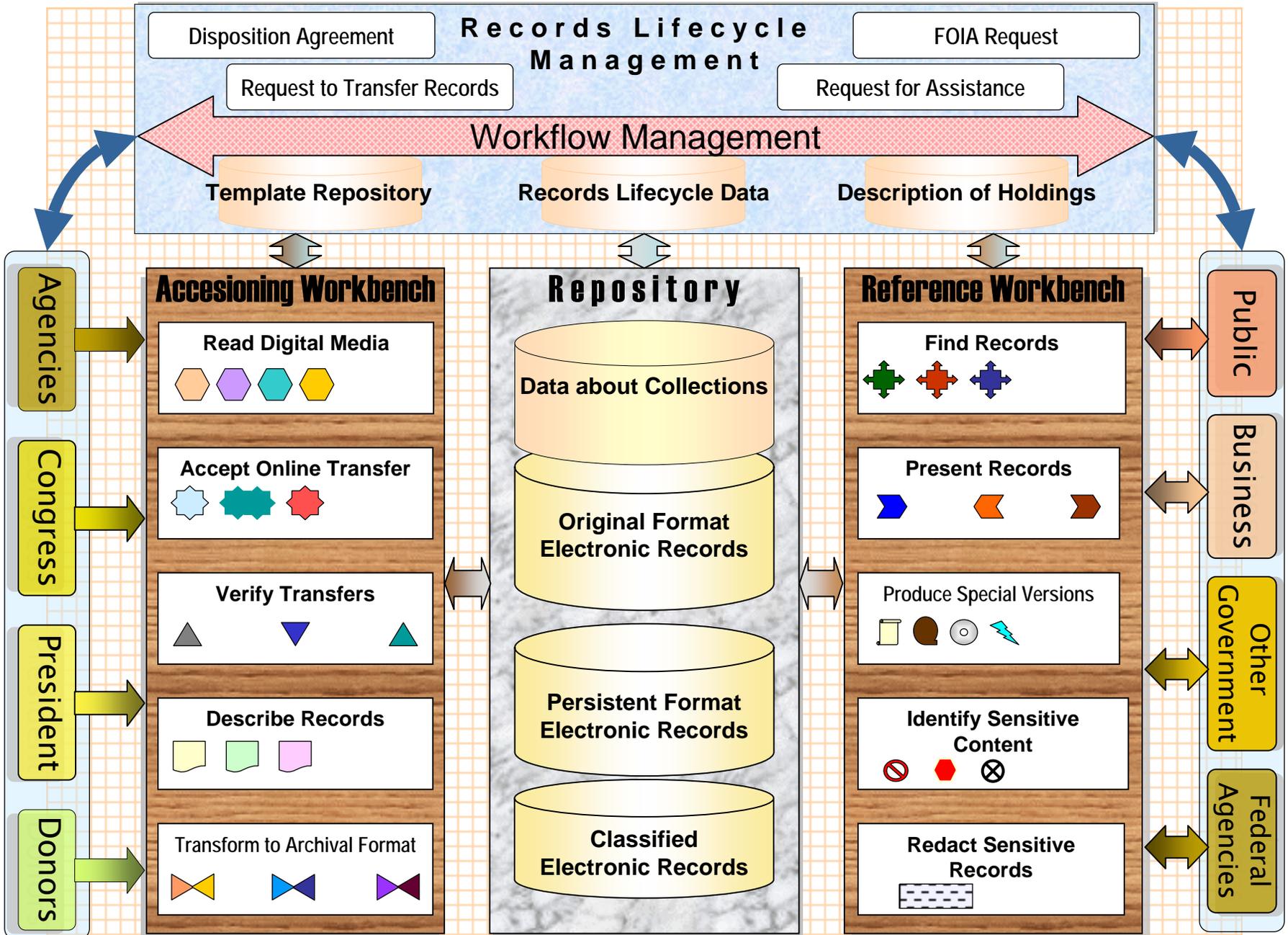
# ERA Virtual Workspaces Increment 1



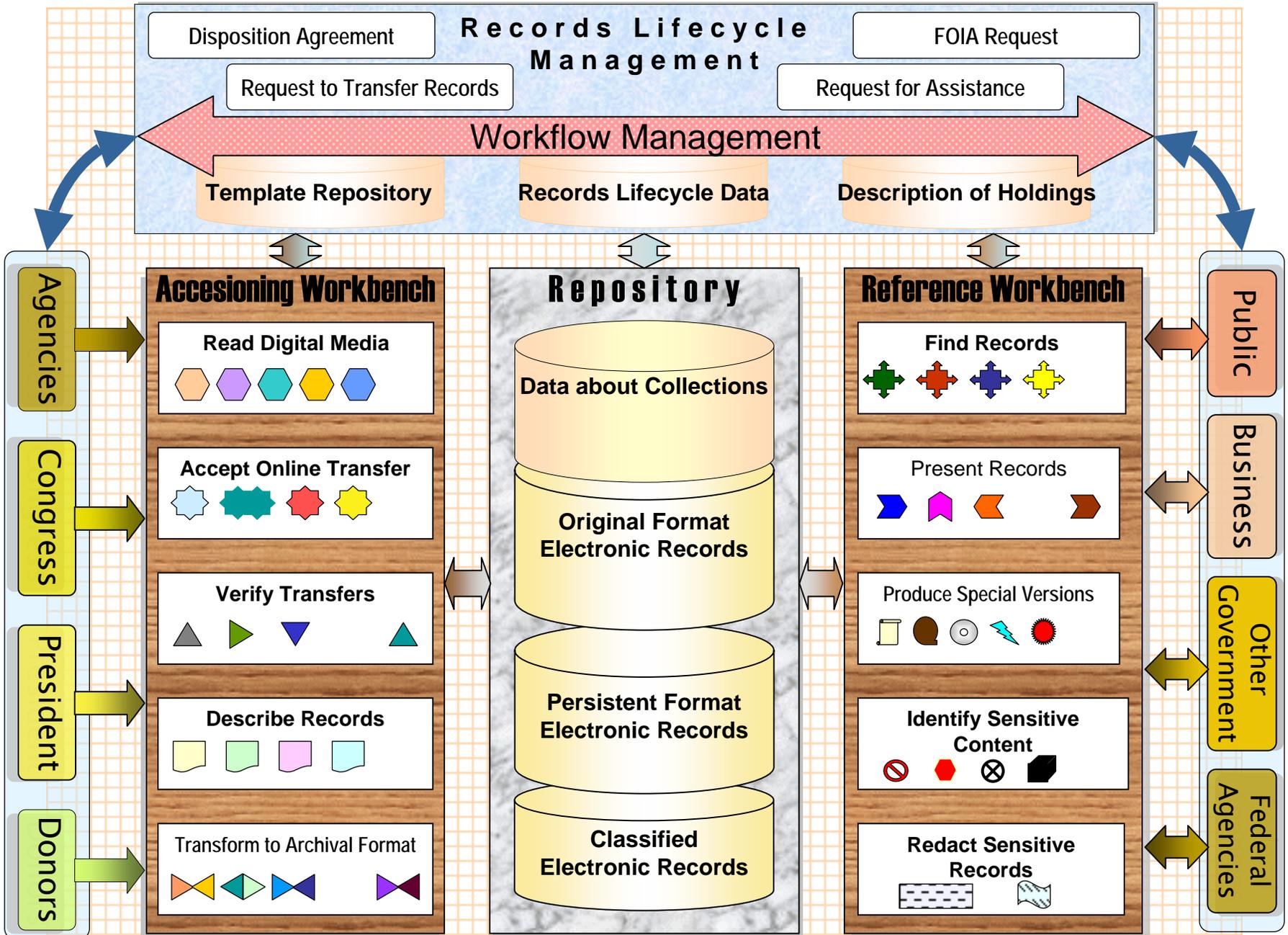
# ERA Virtual Workspaces Increment 2



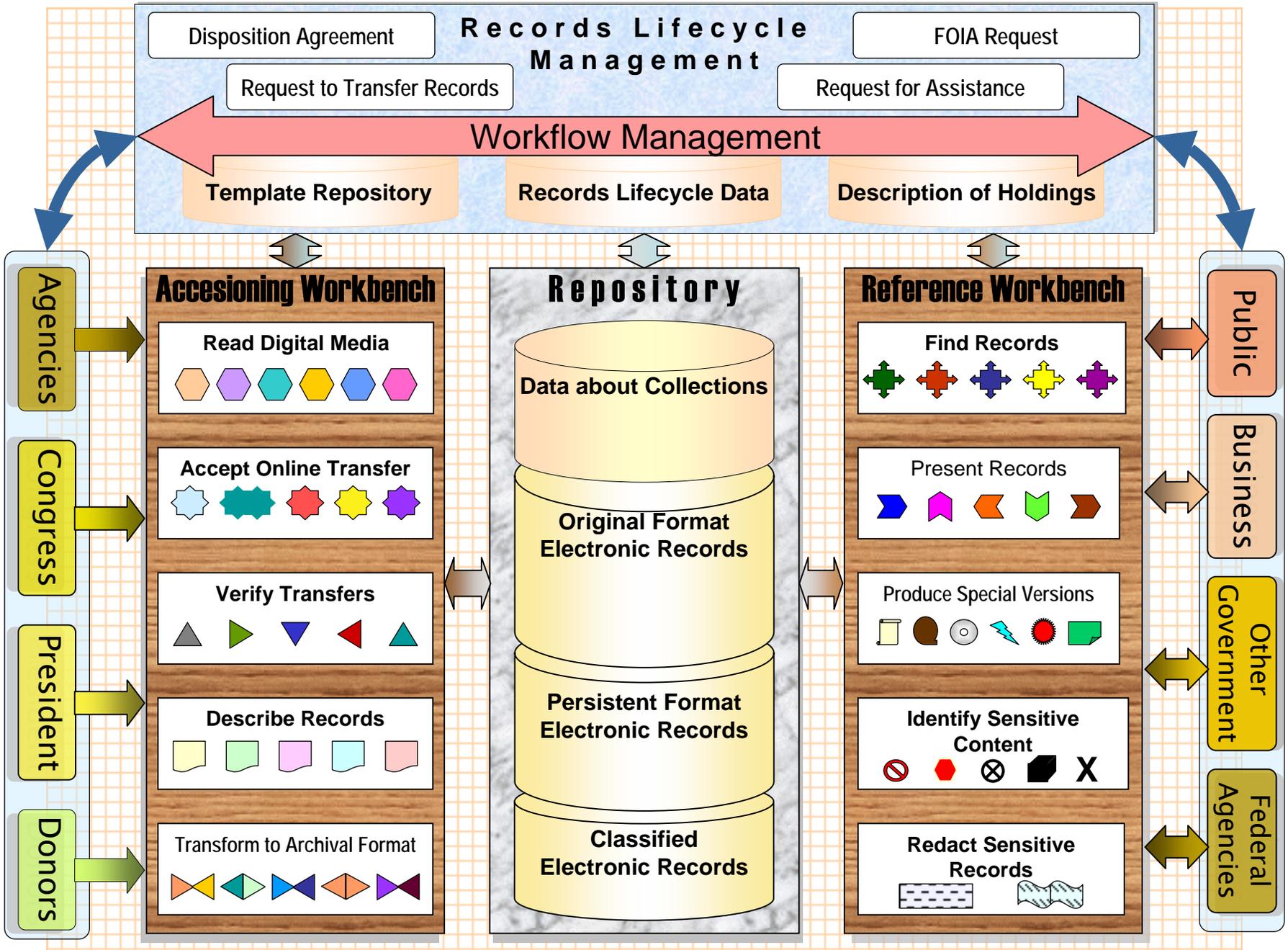
# ERA Virtual Workspaces Increment 3



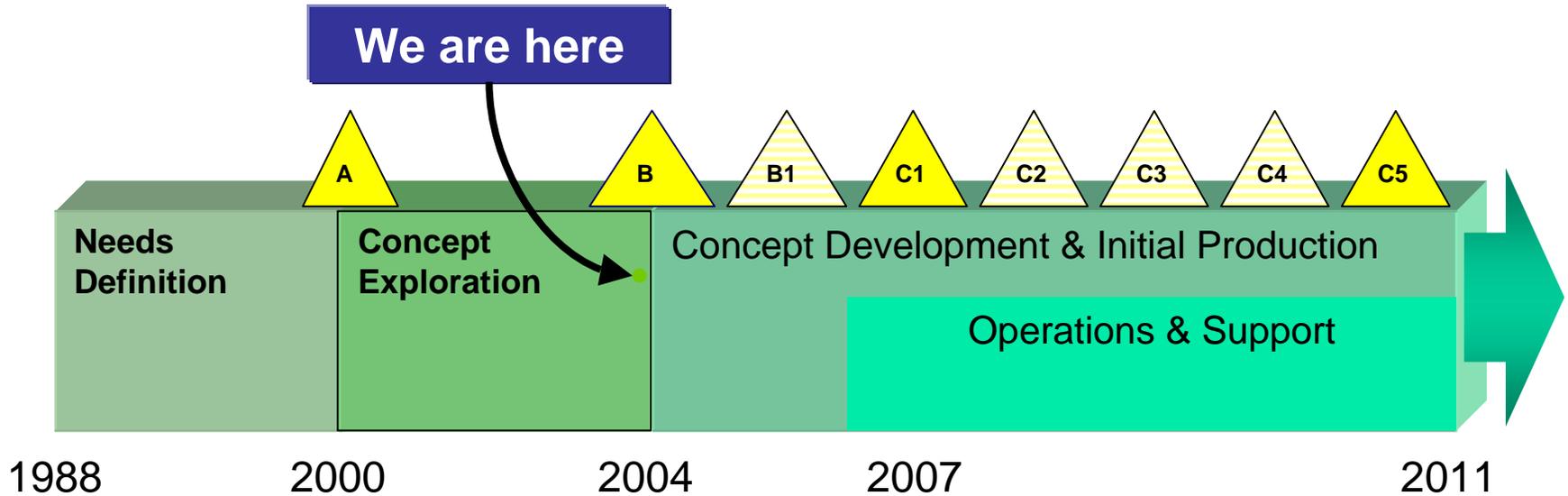
# ERA Virtual Workspaces Increment 4



# ERA Virtual Workspaces Increment 5



# ERA System Acquisition Lifecycle



## Milestones

- A Mission Need, Vision Statement, Analysis of Possibilities
- B Contract Award
- B1 Design Selection
- C1 – C5 Increments 1 to 5



# ERA Partnerships



National Science Foundation



Global Grid Forum



San Diego Supercomputer Center



National Computational Science Alliance

The Library of Congress



Army Research Laboratory



National Agricultural Library



National Partnership for Advanced Computational Infrastructure

# NARA RESEARCH PROTOTYPE PERSISTENT ARCHIVE

*"Preservation & Performance Through Federation"*

## NARA DIGITAL COLLECTIONS and FEDERATED SITES

### Federating Distributed SRB Zones

Multiple types of federation are possible

#### • Replication Zone

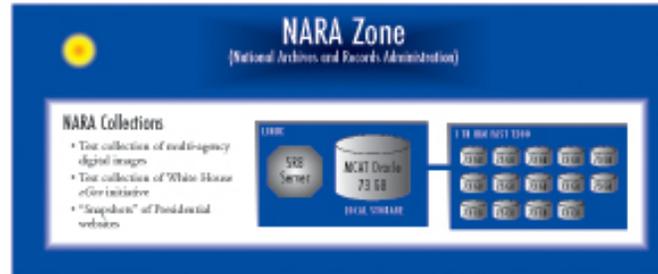
- Resources are synchronized every seven known days/years
- Users are synchronized every user known days/years
- Collections and Data can be partially or completely synchronized (flexible)

#### • Deep Archive

- No shared resources or users
- Access only by data grid administrators
- Collections and Data are replicated

### Federation of Persistent Archives Through Zone SRB for

- Preservation
- Management
- Publication



## Zone SRB Federated Data Grid

### NETWORK CONNECTION

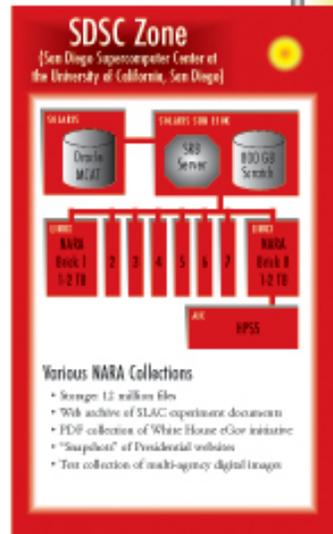
OC3 (255 Mbps) to Internet-Address

### Benefits of Federation through Zone SRB

- High Level of Redundancy
- Risk Mitigation through Data Backup for Disaster Recovery
- Data Available from Multiple Sites
- Latency Management & Load Balancing
- Fault Tolerance & Failover

**Stanford  
Linear  
Accelerator  
Center**

**MIT Libraries  
DSpace™**

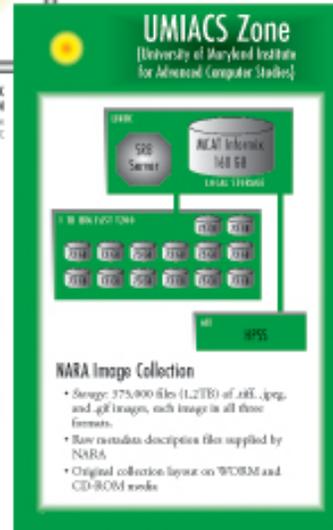


### NETWORK CONNECTION

Internet-Address  
TimeGrid  
CERN  
ESnet

### NETWORK CONNECTION

Internet-Address  
MAX



[www.sdsc.edu/DICE/SRB](http://www.sdsc.edu/DICE/SRB)



For additional information:

[http://www.archives.gov/electronic\\_records\\_archives/  
index.html](http://www.archives.gov/electronic_records_archives/index.html)