



# Building the Archives of the Future

## An Overview of the Electronic Records Archives (ERA) Program

Department of the Navy  
June 26, 2008

Michael Carlson

Director, Electronic and Special Media Records  
Services Division

National Archives and Records Administration

11/4/2008

National Archives and Records  
Administration



# Overview

- **Electronic Records Challenges & Strategies**
- **The ERA Timeline: *Where we are now***
- **Projected Full Operating Capability**



# Electronic Records NARA's Challenges

- **Scope** The entire U.S. Federal Government
- **Obsolescence** Constantly Changing Technology
- **Access** Ability to view records over time
- **Volume** Large amounts of records arriving to NARA
- **Variety** Different/Complex Types of Records
- **Complexity** and Records Formats



# Challenges in Preserving The Records of Our Times

1. The volume of information in digital form is increasing exponentially.
2. The variety and complexity of digital information are increasing.
3. No one knows how to preserve and provide sustained access to authentic electronic records for most types of electronic records.
4. No one knows what information technology will be in the future .



# Examples of Variety & Complexity

## The Air Force Health Study

### Office Automation Files

- PDF documents
- Scanned paper documents

### Complex Formats

- Databases\*
- Digital Photography\*
- **Web pages\***



# Examples of Variety & Complexity

## The 9-11 Commission Records

### Office Automation Files

- Word processing documents
- Spreadsheets
- Presentations
- E-mail w/ attachments
- Scanned paper documents

### Complex Formats

- Databases
- Digital Photography
- Satellite Imagery
- **Digital audio files**
- **HDTV**
- **Web pages**
- **Geospatial Information Systems**



# Examples of the Challenge: Volume

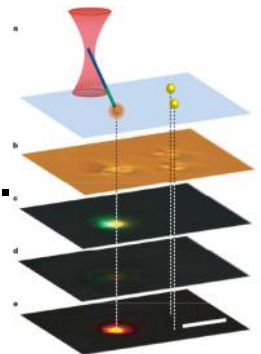
- **Clinton Administration**  
40 million email messages
- **State Department**  
25 million electronic diplomatic messages
- **Department of Defense**  
54 million images from electronic official military personnel files annually
- **Census Bureau (2000 Census)**  
600 - 800 million image (TIFF) files

## 1. Anticipate change in

- The characteristics of electronic records,
- The technologies used to preserve and provide access, and
- The expectations and behaviors of researchers

## 2. Recognize those things that will not or should not change

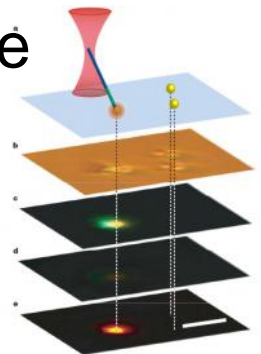
- Archival science provides stable principles, concepts, requirements and understanding.
- NARA's mission and the functions that it must carry out in order to accomplish that mission.





## 3. Make reasonable assumptions about the future

- Computers will continue to become more common in the activities of institutions, the lives of individuals, and the interactions of groups of people.
- The power, speed, capacity and usefulness of information technology will continue to grow
- Prices for Information Technology will continue to decline.
- The Internet will continue to grow



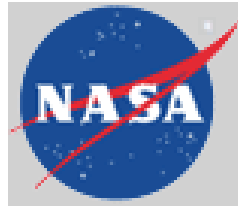


# NARA's Strategic Response: The ERA Program

1. **Research** and exploratory development on technologies that offer promise for addressing electronic records challenges.
2. **Acquiring and building a system** that meets our requirements and our mission for NARA, the Presidential Libraries, and Federal Records Centers
3. Organizational and cultural **Change Management**



# Research 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



InterPARES Project

International Research on Permanent Authentic Records in Electronic Systems



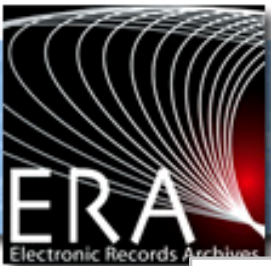
# The ERA Research Prototype: Data Grid Technology



**Electronic Records Stacks  
at Archives II**



**The *ERA Transcontinental  
Persistent Archive Prototype (TPAP)*  
-- or NARA's "Grid"**

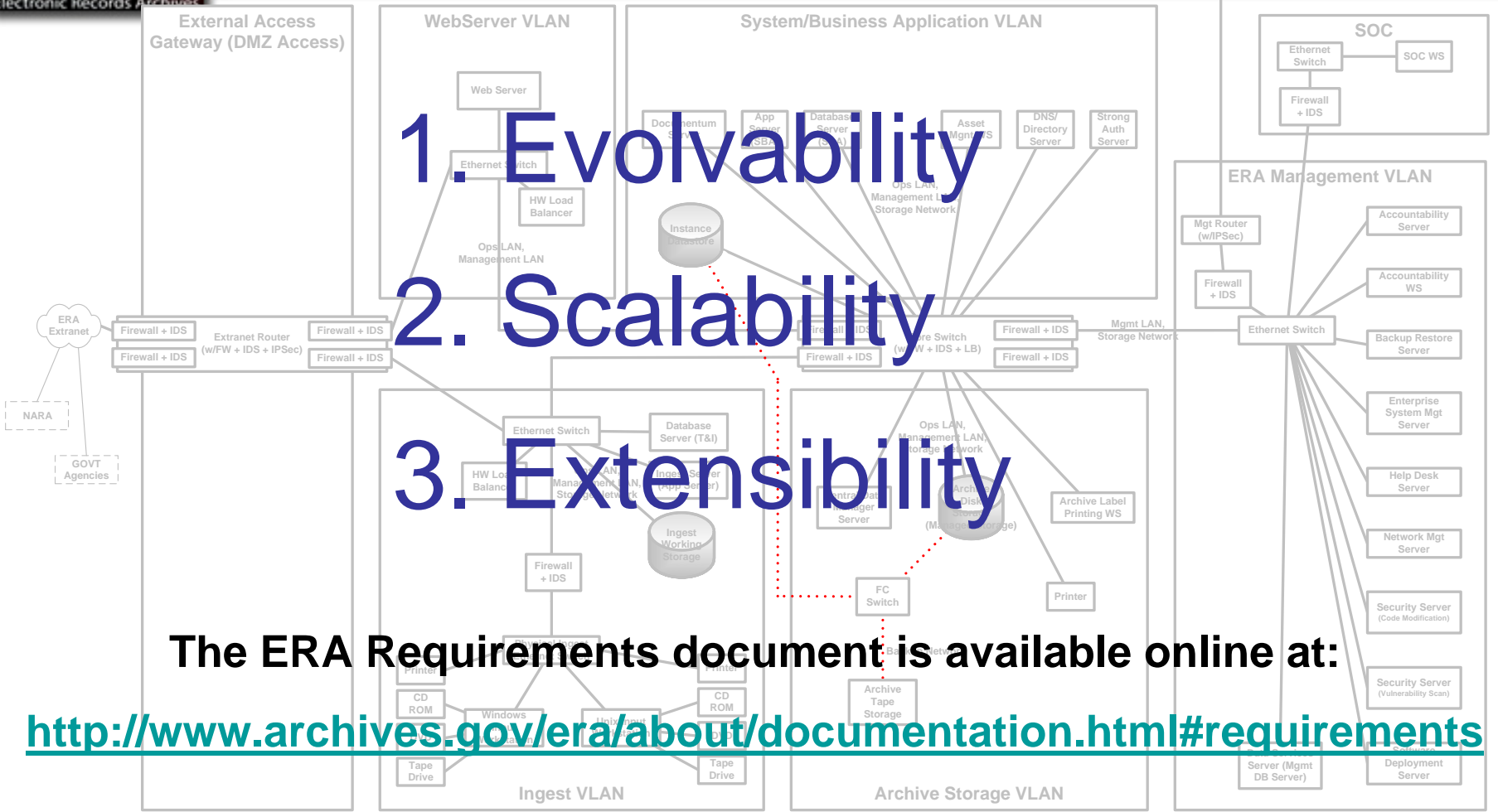


# Fundamental Requirements for The ERA System

- 1. Evolvability
- 2. Scalability
- 3. Extensibility

The ERA Requirements document is available online at:

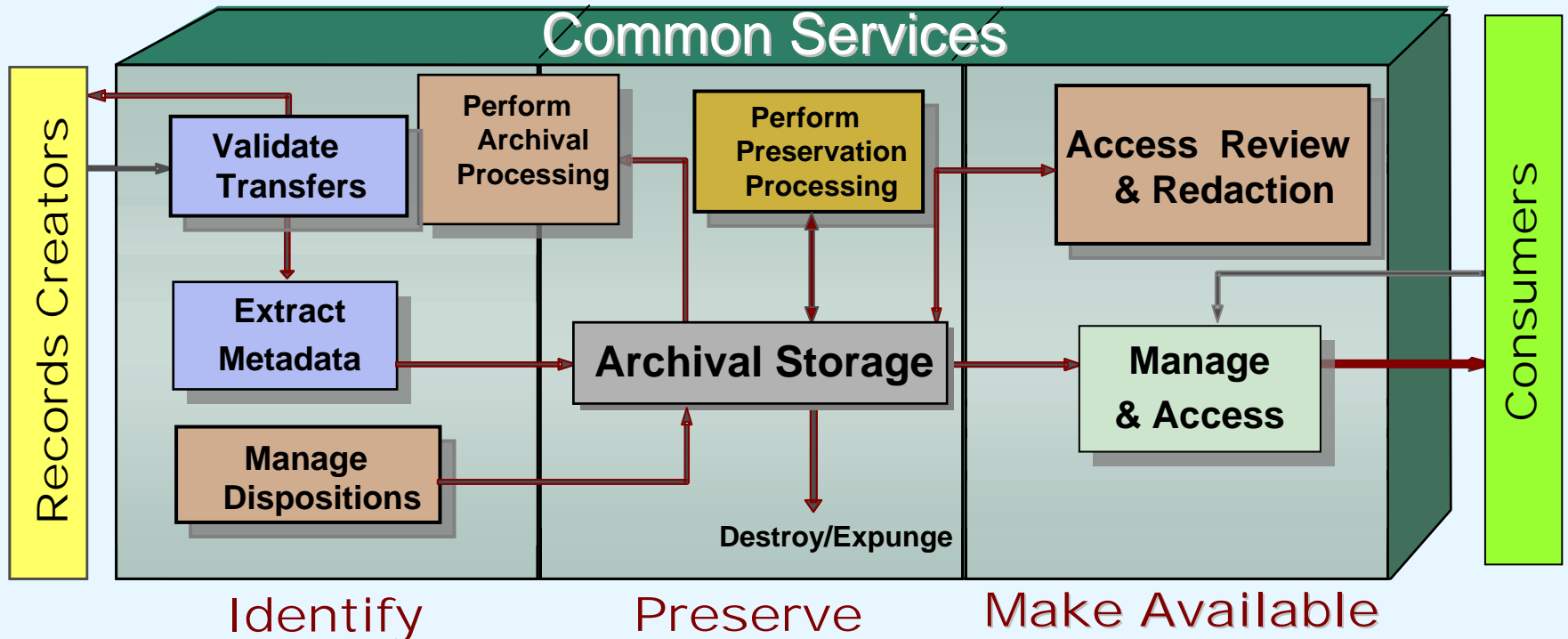
<http://www.archives.gov/era/about/documentation.html#requirements>



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

1Gb Ethernet  
2/4Gb Fibre Channel

# The ERA System is based on a Service-oriented architecture





# ERA will be a set of NESTED, mini-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

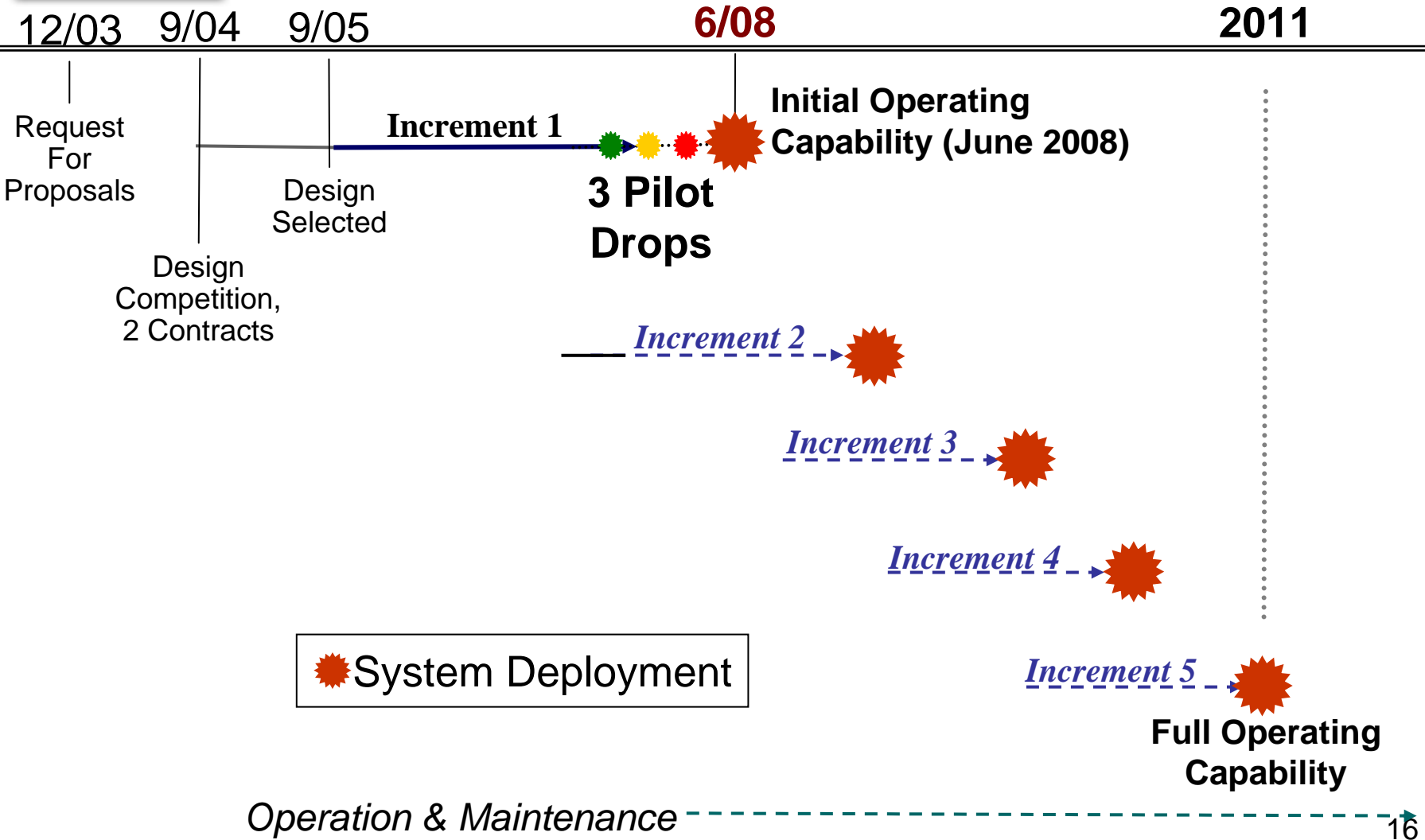
- Support a variety of different approaches to different needs.

## Archival “mini-systems”

- Specific, systematic management for each series or aggregate of electronic records.



# Acquiring and Building ERA Incrementally







# ERA System Functions at Each Drop

<b>Drop 1</b> <b>Sep 21, 2007</b>	<b>Drop 2</b> <b>Dec 19, 2007</b>	<b>Drop 3</b> <b>March 7, 2008</b>
<ul style="list-style-type: none"><li>• Create, modify, and delete <b><u>new records schedules</u></b></li><li>• Create, modify, and delete new <b><u>Legal Transfer Instrument</u></b> (proxy for agencies)</li><li>• Create, modify, and delete new <b><u>Transfer Requests</u></b> (proxy for agencies)</li></ul>	<ul style="list-style-type: none"><li>• Browse Asset Catalog</li><li>• Create, modify, and delete legacy records schedules</li><li>• Create and modify shipments of electronic records to ERA</li><li>• <b><u>Transfer and ingest</u></b> electronic records to ERA</li><li>• Create, modify, and delete legacy Transfer Requests</li><li>• Create, modify, and delete legacy Legal Transfer Instruments</li></ul>	<ul style="list-style-type: none"><li>• <b><u>Accept legal and physical custody</u></b> of non-electronic records</li><li>• <b><u>Verification</u></b> of electronic records</li><li>• Accept physical and legal custody of electronic and non-electronic records</li></ul>



# Projected Full Operating Capability

- Scalable to one exabyte of storage without a major design change.
- Provide dynamic, evolvable capabilities for preserving and providing long-term access to electronic records.
- Be able to ingest records from Records Management Applications (RMAs)
- Provide public access to unrestricted holdings.
- Support review and redaction of restricted holdings



# Current Preservation Strategy

- Software and Hardware Independence
  - When possible; some dependencies unavoidable at this time
  - Hardware/software rights may dictate access and sustainability
- Focus on evidential and research values
  - Records needed for evidence of agency activities, rights, and responsibilities
  - Records needed for future research purposes



# Transfer Guidelines

- Databases – (36 CFR 1228) flat file, ASCII, detailed documentation as to table relationships, linkages, layouts, and code lists.
- E-Mail –
- Scanned Images of Textual Records
- Born Digital Images



# Transfer Guidelines

- GIS
- Web Records
- PDF



## Next Steps

- Leverage the ERA research
- Leverage ERA preservation framework
- Determine path(s)