National Archives and Records Administration

The Electronic Records Archives Initiative

Digital Strategies 2000

Kenneth Thibodeau

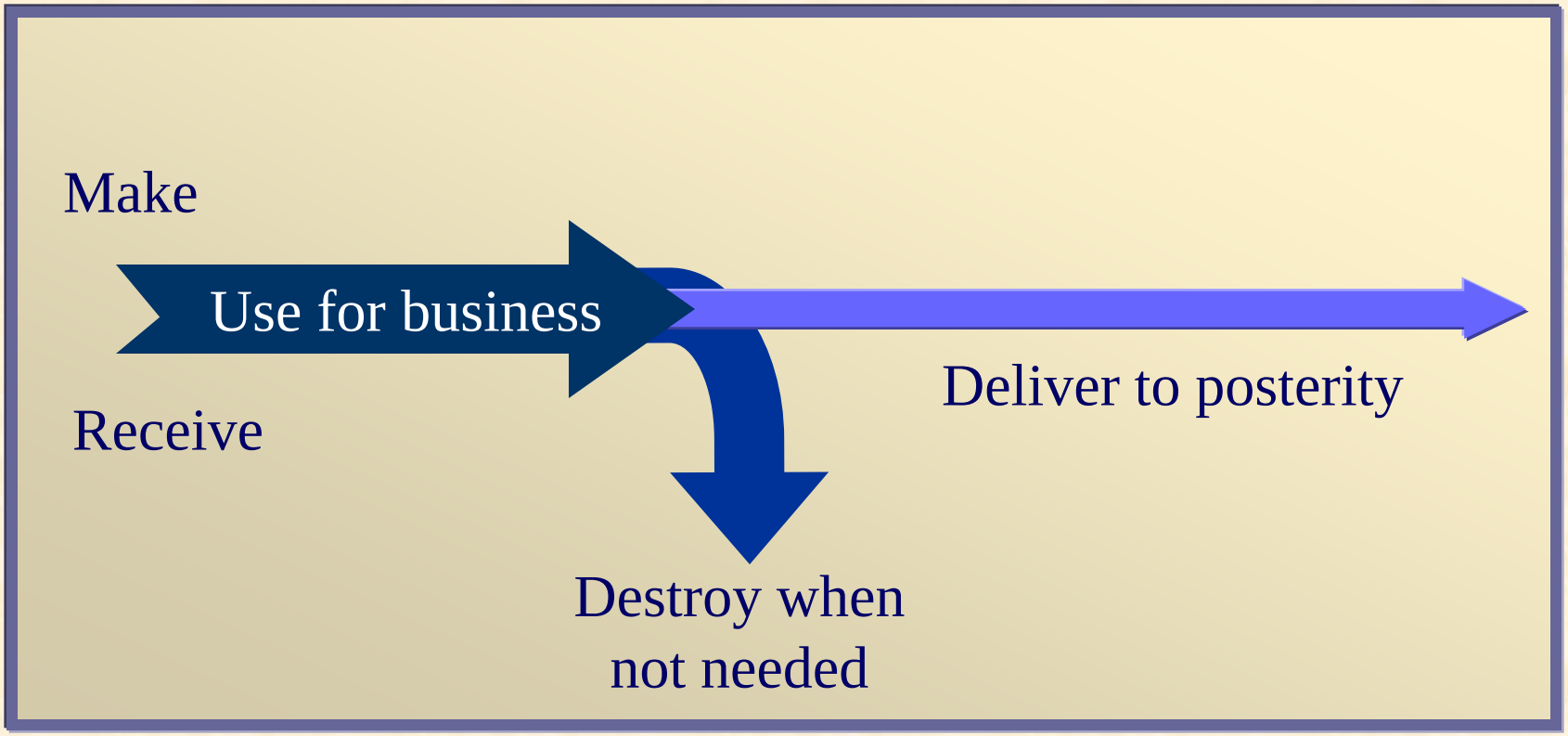
November 16, 2000

"Information Technology Research: Investing in Our Future"

"Our Nation's security, commerce, education, and well-being depend increasingly on our information infrastructure. It is thus critical to ensure the survivability of that infrastructure.... **Survivability includes long term preservation of information**; a document of historical importance should be preserved, even as underlying storage technologies and information representation evolve and even if it is not accessed for hundreds of years."

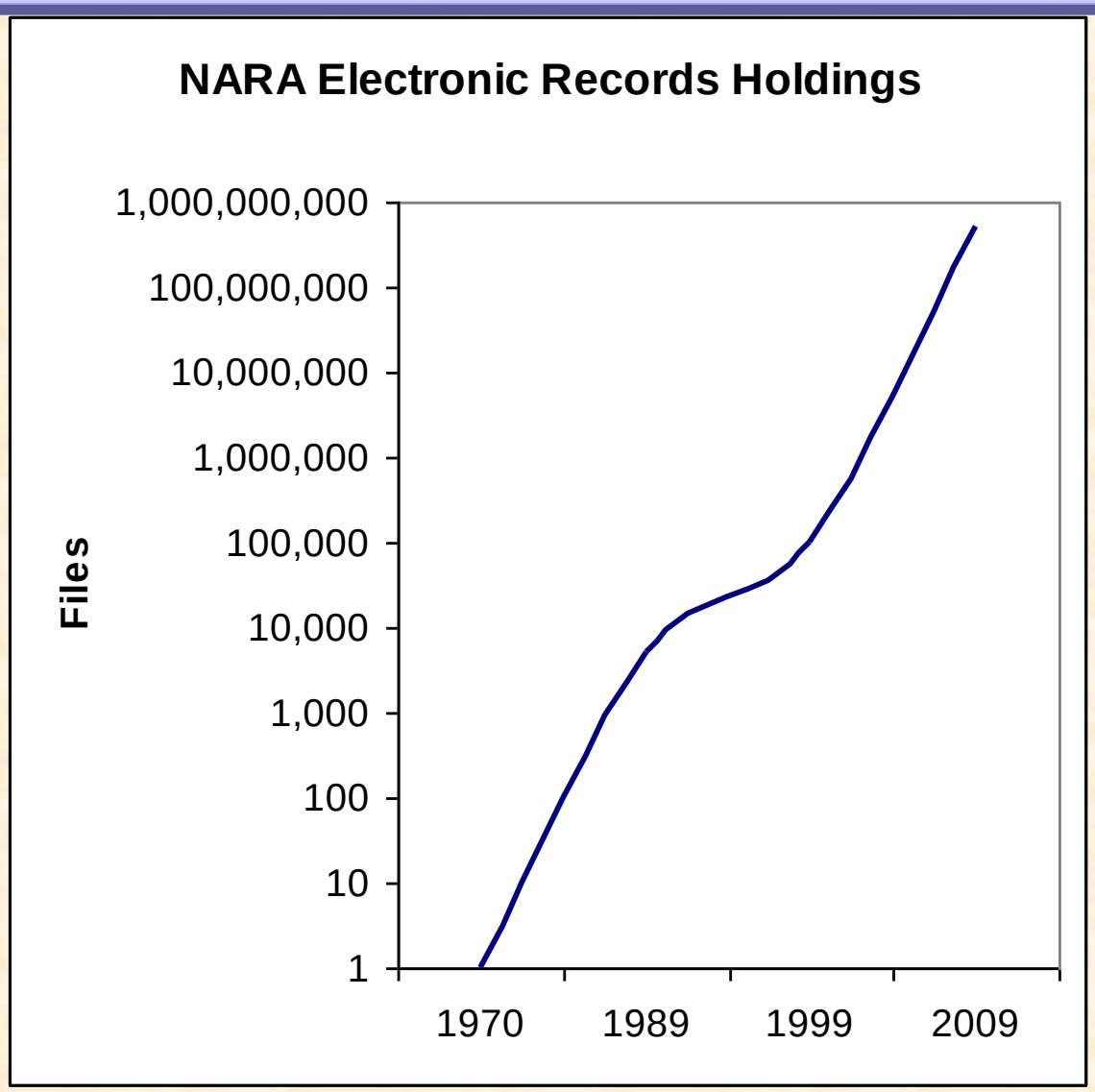
President's Information Technology Advisory Committee, Report To The President. February 1999

Electronic Records Archives Scope



The Life Cycle of Records

What is the size of the challenge?



What do we need to do?

- Overcome technological obsolescence in a way that preserves demonstrably authentic records.
- Build a dynamic solution that incorporates the expectation of continuing change in information technology and in the records it produces.
- Find ways to take advantage of continuing progress in information technology in order to maintain and improve both performance and customer service

What is the Electronic Records Archives?

The Electronic Records Archives is a comprehensive, systematic, and dynamic **means** of providing continuing access to authentic electronic records over time.

“Information Technology: The 21st Century Revolution”

“Today's fast-evolving capabilities of computing systems and networks make it possible to recreate, archive, display, and manipulate exponentially greater quantities of electronically generated documents, data, images, sounds, and video streams-and to offer potential instant access to this knowledge to a significant portion of the world's population. **But organizing "collections" of these many forms of electronic information and developing systems and software tools to make them available to end users require complex technical innovations, including collaboration among experts from widely disparate fields of knowledge.**

Digital Libraries Initiative “Phase Two activities are jointly supported by NSF, DARPA, NIH/NLM, the Library of Congress, NASA, the National Endowment for the Humanities, and the FBI, in partnership with the National Archives and Records Administration, the Smithsonian Institution, and the Institute of Museum and Library Services.”

Supplement to the President's Budget, FY 2001

Interagency Working Group on Information Technology Research and Development

National Science and Technology Council

“Information Technology: The 21st Century Revolution”

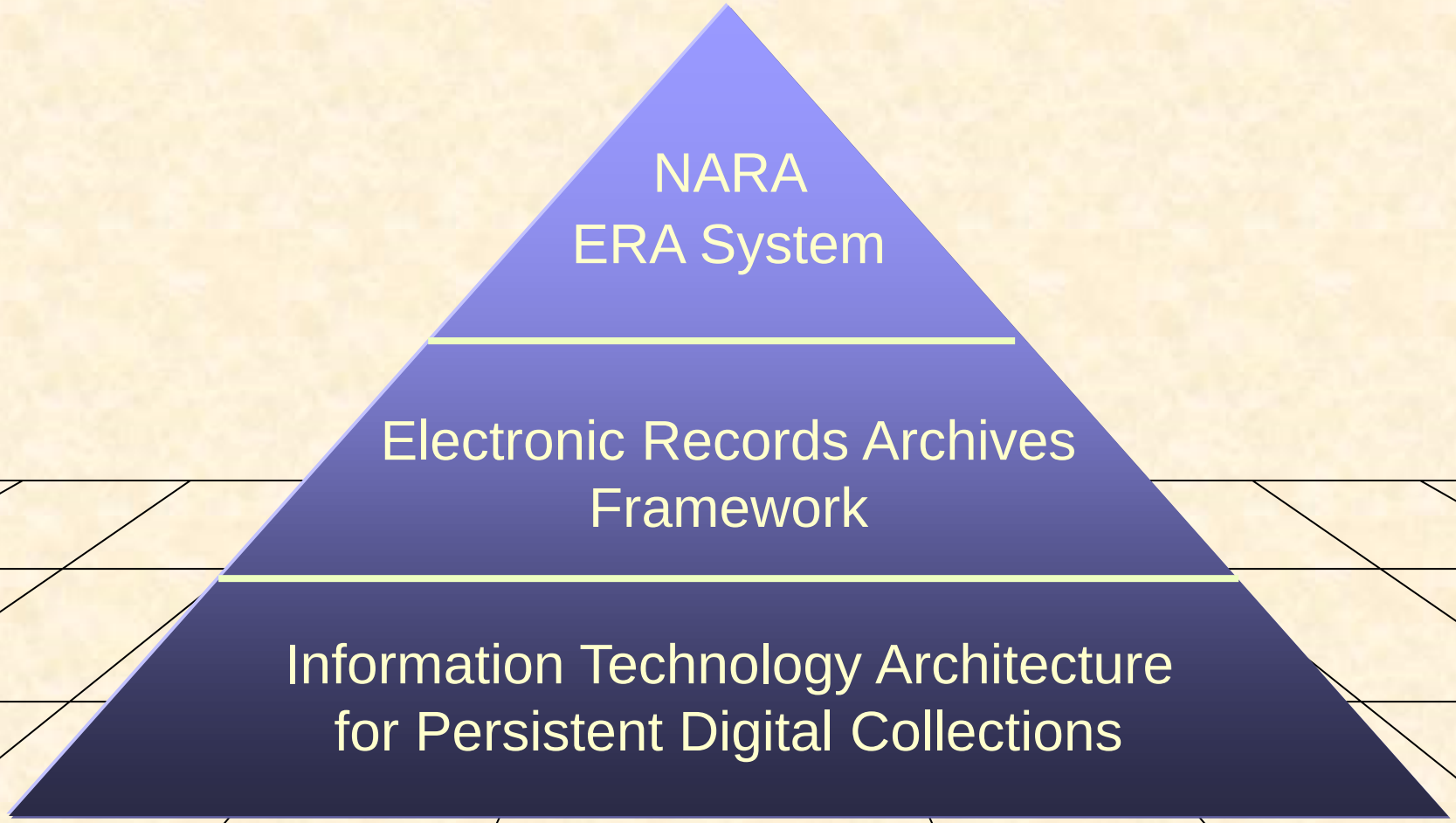
National Partnership for Advanced Computational Infrastructure (NPACI) “ is creating a continuous, ubiquitous, and pervasive national computational infrastructure. NPACI's focus includes providing computation and information resources to enable discovery at scales not previously achievable; developing and deploying integrated, easy-to-use computational environments to foster discovery in traditional and emerging disciplines; and promoting computational literacy to extend the excitement, benefits, and opportunities of computational science to all U.S. citizens.”

Supplement to the President's Budget, FY 2001

Interagency Working Group on Information Technology Research and Development

National Science and Technology Council

How will we develop the Electronic Records Archives?



NARA Partnerships

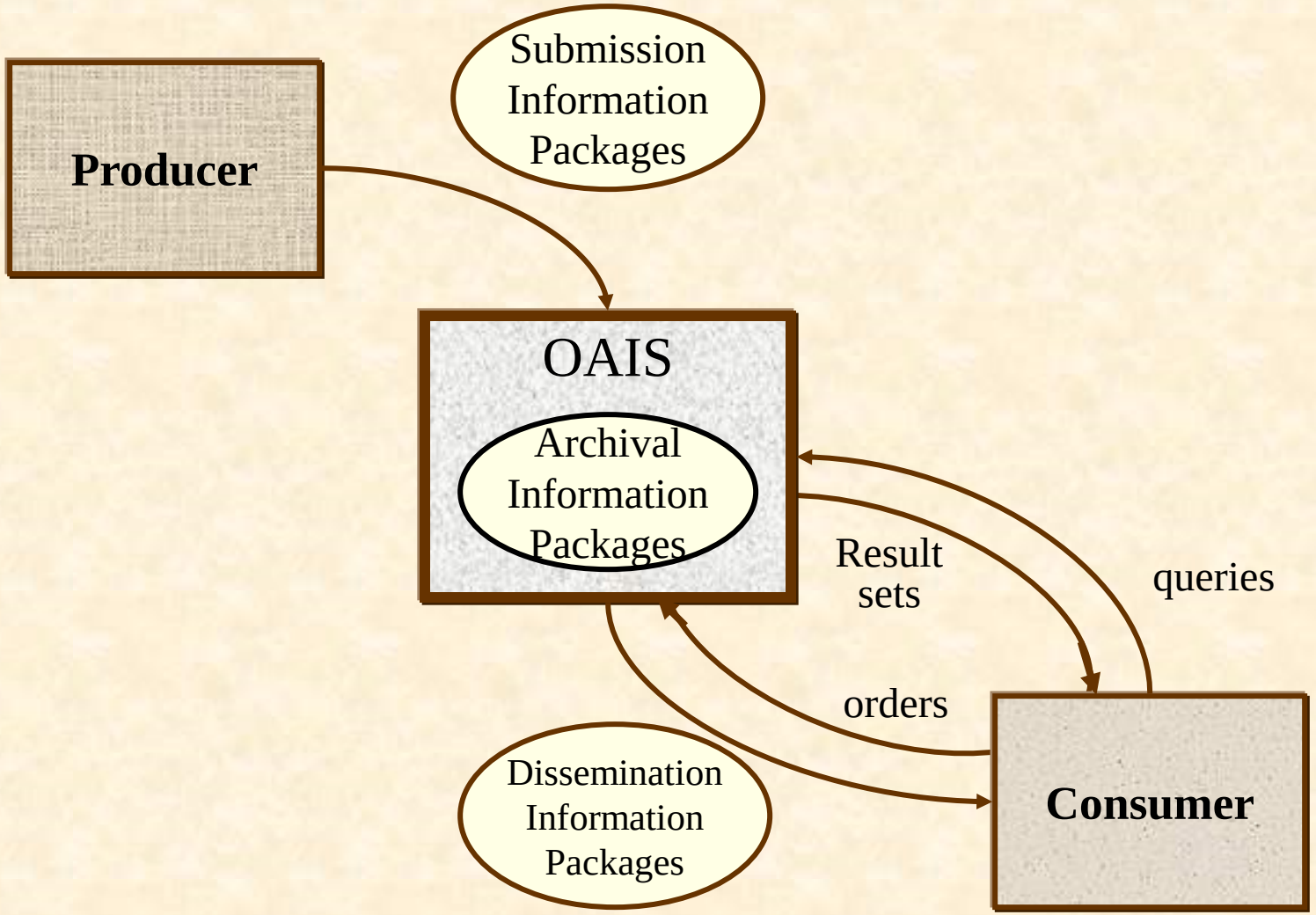
- **Open Archival Information System** (OAIS) Reference Model
  - NASA, Consultative Committtee on Space Data Systems
- **Distributed Object Computation Testbed** (DOCT)
  - Defense Advanced Research Projects Agency, U.S. Patent and Trademark Office
- **National Partnership for Advanced Computational Infrastructure** (NPACI)
  - National Science Foundation
- **Presidential Electronic Records Processing Operational System** (PERPOS)
  - Army Research Laboratory, Georgia Tech Research Institute
- **Archivist's Workbench**
  - NHPRC Grant to San Diego Supercomputer Center
- **International research on Permanent Authentic Records in Electronic Systems** (InterPARES)
  - 7 international research teams, 10 national archives

How do these activities fit together?

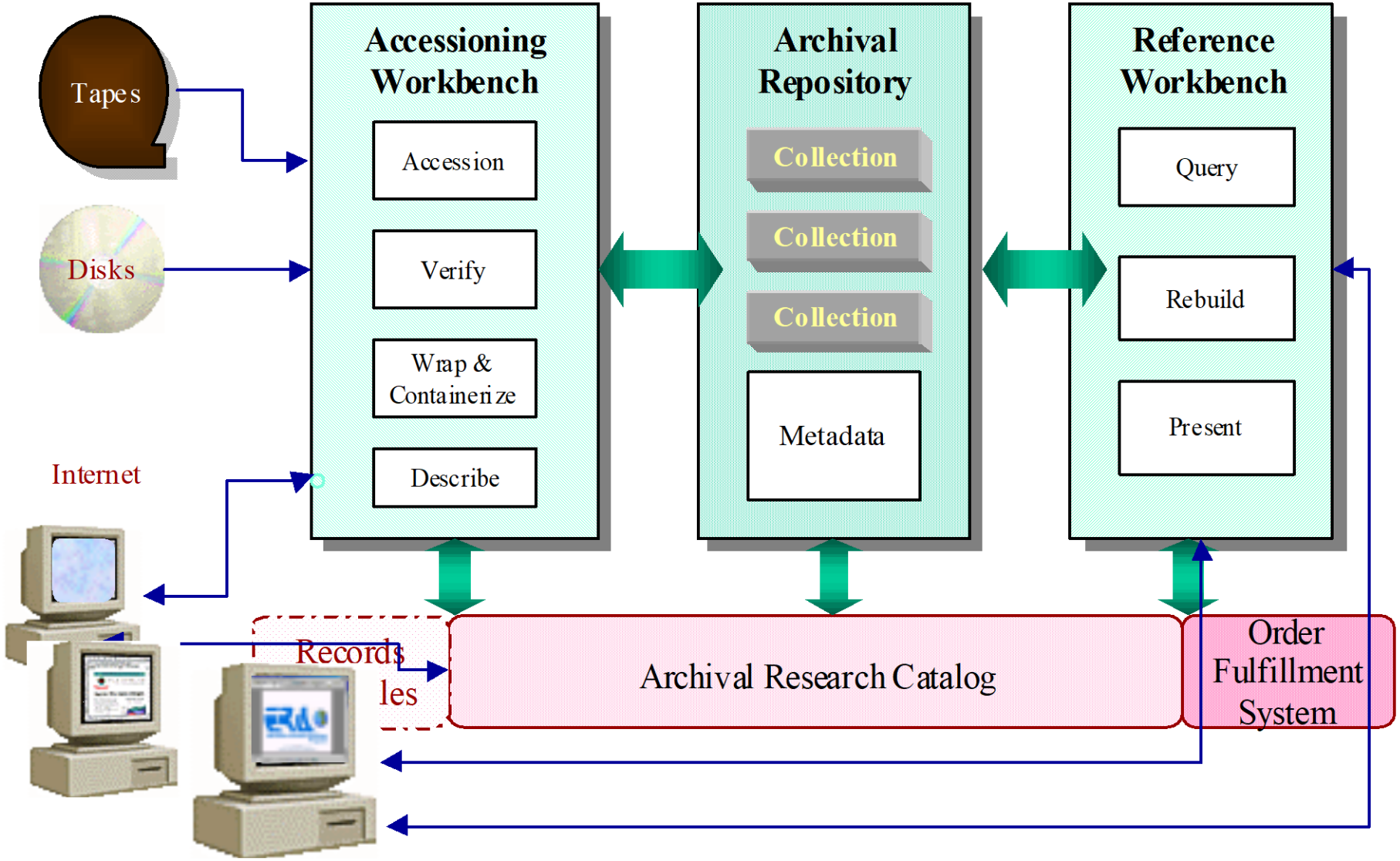
- **OAIS Model**
  - High level framework for administration, entities, functions, data flows
- **DOCT**
  - Persistent Object Preservation
- **NPACI**
  - Architecture and core technologies for a comprehensive solution based on Persistent Object Preservation
- **PERPOS**
  - Advanced tools for archival processes
- **Archivist's Workbench**
  - Scale ERA for smaller archives
- **InterPARES**
  - Archival requirements, electronic records typology, preservation model


IMA
ARTHUR ANDERSEN
The Canadian Institute of Chartered Accountants
Cohen Computer Consulting
International Accounting Standards Committee
EDGAR ONLINE
The Woodburn Group
www.woodburn.com
612-542-3800
Best Software, Inc.
CPA
AUSTRALIAN SOCIETY OF CPAs
the e-content company
ORACLE
SOFTWARE POWERS THE INTERNET
GREAT PLAINS
sage
SAP
THE INSTITUTE OF CHARTERED ACCOUNTANTS
IN ENGLAND & WALES
epartners
Deloitte & Touche
XBRL
NAVISION software
AICPA
IBM
Microsoft
CASEWARE INTERNATIONAL INC.
MORGAN STANLEY DEAN WITTER
KPMG
The Institute of Chartered Accountants in Australia
PRICEWATERHOUSECOOPERS
STANDARD & POOR'S
The McGraw-Hill Companies
LAWSON Software
ERNST & YOUNG
FROM THOUGHT TO FINISH
REUTERS
epicor
Software Corporation
ppa
Gesellschaft für Finanzanalyse und Benchmarks mbH
Hyperion
CROWE CHIZEK
Grant Thornton

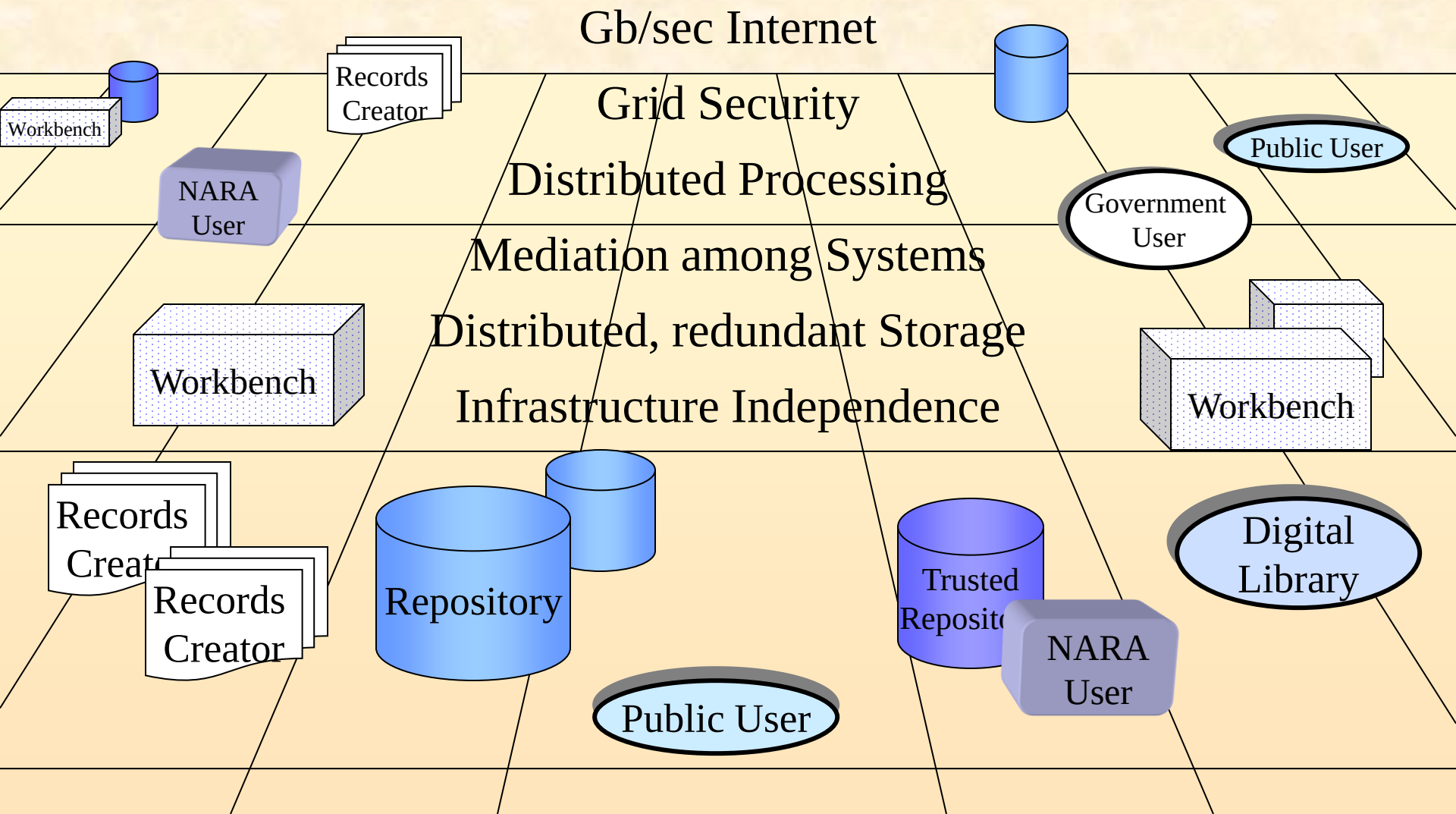
ERA Functional Model
An Open Archival Information System Implementation



ERA Virtual Workspace Model



ERA Infrastructure Concept



The Electronic Records Archives Framework

What are the Objectives of Digital Preservation?

- Preserve Technology

Computer Games

Virtual Reality Models

What is important?

Records

Photographs

Maps

- Preserve Things produced with technology

"Data Objects"

Transformation to Persistent Object Form

- Characterize all significant properties of the things that are to be preserved.
- Express these requirements in formal models
- Transform the objects by
  - Encapsulating them in metadata defined in the models
  - Eliminating other technical characteristics that are dependent on specific hardware or software, proprietary, or subject to obsolescence.
- Use software “mediators” to enable future technologies to interpret the models and metadata.

Persistent Object Preservation

+ Aims at independence of technological infrastructure
  - Do not preserve things in original technological materialization.
  + Reduce threats to integrity and authenticity by minimizing changes over time.
+ Embeds changes in a comprehensive information management architecture designed for preservation
+ Inherently extensible
+ Facilitates use of future, advanced technologies, without requiring change in what is preserved.
- Currently beyond state of the art of information technology.

Persistent Object Preservation of Records

+ Focuses specifically on satisfying archival requirements for preservation of records.
+ Preserves files, series, and archival fonds as well as individual records
+ Minimizing changes over time reduces threats to integrity and authenticity.
+ Requirements for integrity and authenticity are controlled explicitly and specifically.
+ Provides for the performance of archival functions as well as the persistence of the records.
+ Designed so that use of advanced technologies for search, retrieval & delivery does not compromise integrity.

Collection-Based Persistent Object Preservation

- Comprehensive
  - All types of computer applications
  - All types of electronic records
  - Collections as well as individual records
  - All required archival processes
- Highly Automated
- Scalable
  - >= X00,000,000 objects,
  - and down for small collections and repositories
- Implemented over the Internet

Thank you.

For more information:

www.nara.gov/era