

# **Open Source's Critical Role When Data Must Be Shared**

# Open Source's Critical Role When Data Must Be Shared

## Speakers:

Isaac Christoffersen (Moderator)

Contributor, InCommon Inc.

Quyen Nguyen

System Architect, Engineering Division, ERA Program Management  
Office, US National Archives and Records Administration

Jerry Johnston

Geospatial Information Officer, US Environmental Protection Agency

Tom MacWright

Engineer, Development Seed

# Data Center Feng Shui

Using Open Source Software to Align IT  
Demands with Business Objectives

Isaac Christoffersen  
*Contributor*

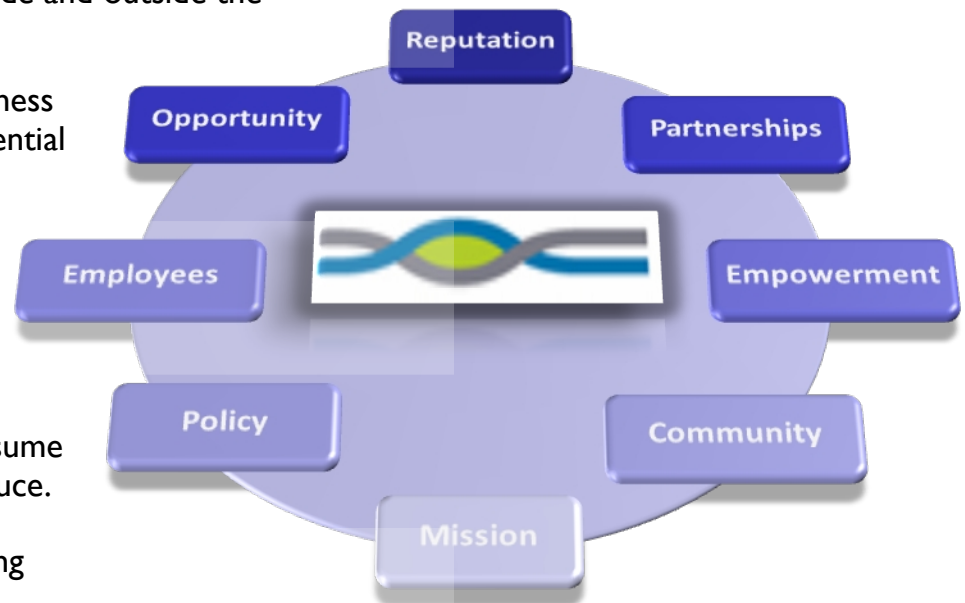
# INCOMMON – ADVOCATE. EDUCATE. ACCELERATE.

---

- ▶ As community engagement practitioners, we seek to build ongoing, permanent relationships for the purpose of applying a collective vision for the benefit of the community.
- ▶ inCommon, Inc was started to **advocate** the use of Open Source and Open Architecture solutions in the U.S. government market.
- ▶ As part of this mission, we seek to **educate** individuals on the value of Virtual, Grid, Data and SOA architectures based on Open Source and Open Architecture constructs.
- ▶ We work with clients to efficiently and effectively **accelerate** systems delivery while maximizing existing resources whenever possible.

# ▶ Data is essence of existence for Today's IT Systems

- Reputation** It takes a lifetime to build a good reputation and one system-wide outage to ruin it.
- Partnerships** Data is always a heterogeneous assembly of different sources. Partnerships help develop interoperability as well as enrich the content.
- Empowerment** Empowering employees and communities encourages growth and development of new capabilities.
- Community** The community extends both inside and outside the organization.
- Mission** The alignment organization's business objective and the technology potential should support the mission.
- Policy** Business objectives and/or the technology potential should help shape policy.
- Employees** Employees should have a vested interest in the data that they consume as well as the data that they produce.
- Opportunity** Focus on potential while identifying roadblocks and speed bumps.



# Principles of Open Data

- ▶ Complete
- ▶ Primary
- ▶ Timely
- ▶ Machine processable
- ▶ Non-discriminatory
- ▶ Non-proprietary
- ▶ License-free

*Source:*

*<http://wiki.opengovdata.org/index.php?title=OpenDataPrinciples>*



# Agile Implementation of e-Forms based on Open Standards and Open Software

Quyen Nguyen

National Archives and Records Administration

ERA System Engineering



# Agenda

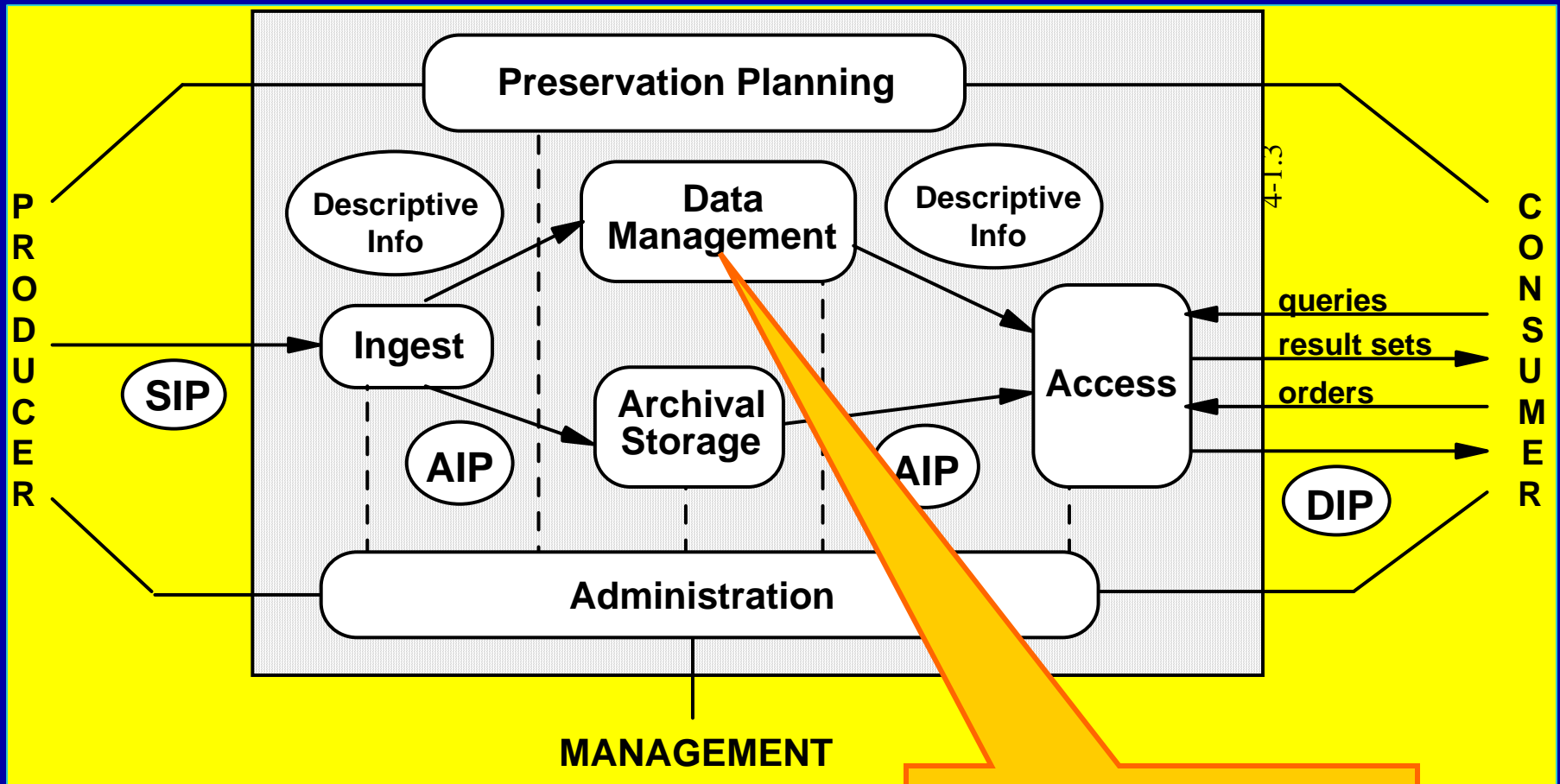
- Background: ERA System
- Problem Description
  - Archival Business Objects (ABO)
  - Authority Lists
- Agile Approach
  - XML-based Approach



# Agenda

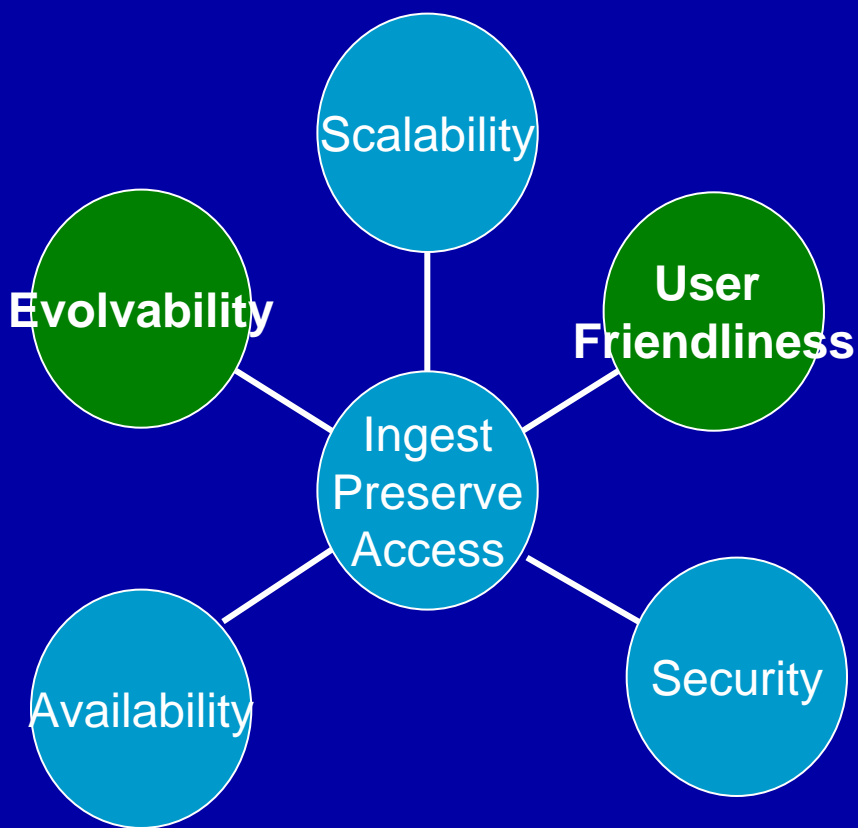
- Background: ERA System
- Problem Description
  - Archival Business Objects
  - Authority Lists
- Agile Approach
  - XML-based Approach

# OAIS Functional Model



- Record Scheduling
- Transfer Request

# ERA System Requirements



- ❑ **Evolvability**
  - New record types, data types, and services could be added;
  - New technologies in software and hardware could be inserted;
  - New rules could be easily adopted;
  - Use of Open Standards.
- ❑ **User Friendliness**
  - Browser interface;
  - 508 compliance;
  - Availability: key functions must be available more than 99%; no single point of failure.
- ❑ **Security**
  - Protection of system and its assets.
- ❑ **Scalability**
  - Adapt to record volume and user community growth.

# Agenda

- Background: ERA System
- Problem Description
  - Archival Business Objects (ABO)
  - Authority Lists
- Agile Approach
  - XML-based Approach

# Archival Business Object (ABO) Requirements

- ❑ Elimination of free text form
- ❑ Structured fields that conform to a pre-defined data model
- ❑ CRUDS (Create, Retrieve, Update, Delete, Search) operations
- ❑ Elaborate validation of input data based on embedded business rules during creation/update of business objects
  - Required vs Optional;
  - Inter-dependencies between fields: values, lists, calculation, etc.;
  - Editable fields based on state.
- ❑ Storage model: XML encoding
- ❑ Presentation of ABO
  - Browser View: W3C HTML standard browser;
  - Print View: XML --> XSL-FO --> PDF.

# Authority Lists

- ❑ Controlled vocabulary
  - Term standardization;
  - Enumerated values;
  - Drop down list in a form.
  
- ❑ Authority List Management
  - Maintainability;
  - Responsibility of an authoritative organization;
  - Facilitate data transfer.
  
- ❑ There can more than 20 authority lists in one ABO form

# Authority List Examples

Name	Value
NARA	National Archives and Records Administration
USPTO	United States Patent and Trademark Office
NOAA	National Oceanic and Atmospheric Administration

Paper Name	Size (mm)	Size (in.)	Size (pts.)
4A0	1582 x 2378	66.22 x 93.62	14768 x 6741
2A0	1189 x 1682	46.81 x 66.22	3370 x 4768
A0	841 x 1189	33 x 46.81	2384 x 3370

# In Search of an Approach for e-Forms

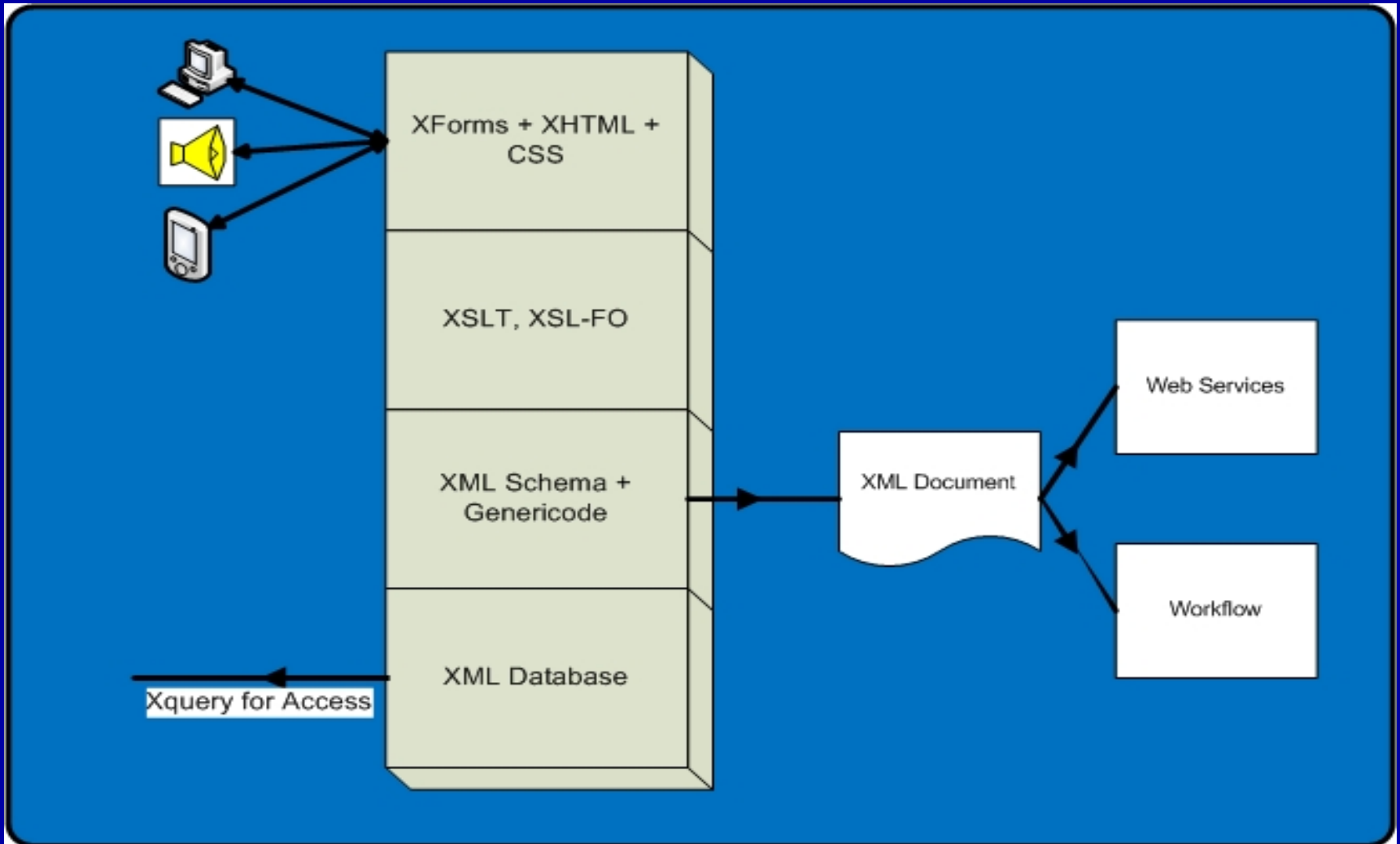
- ❑ How to ensure data integrity?
  - Validate input data;
  - Incorporate business rules easily.
- ❑ How to add or change a field?
  - Changes required throughout all layers of MVC;
  - Development cost and turn-around time.
- ❑ How to represent values in a code list in ABO data?
- ❑ User experience
  - Need to avoid screen refresh upon value selection in authority list;
  - Where to validate data input?
- ❑ Handling of XML data
  - Marshalling/Demarshalling of XML data;
  - Modification to schema requires code recompilation.
- ❑ Modularity, Reusability, Low-cost Maintenance



# Agenda

- Background: ERA System
- Problem Description
  - Archival Business Objects
  - Authority Lists
- Agile Approach
  - XML-based Approach

# XML-based Stack



# Advantages: Data Handling

## □ Data Integrity

- The Xforms are compliant with data constraints specified by the XML schema.
- No code is required to perform this kind of validation.
- Business rules governing the inter-dependency among the fields of a business object can be easily implemented using XForms constructs declaratively.
- The use of JavaScript is reduced.
- Some calculations, such as the total volume based on a container's dimensions, can be performed on the XForms itself.

## □ Data Exchange

- Conversions are markedly reduced.
- Output is an XML document, which can readily carry data between SOA components and services.

# Advantages: User Experience

## ❑ Performance

- Response time and, consequently, user experience are enhanced as latency is reduced thanks to Ajax-based implementations.

## ❑ Consistency

- XForms format specifies a construct to handle XML errors, facilitating uniform and consistent error handling and error messages.

# Advantages: Software Engineering

## ❑ Modularity and Reuse

- An XForms document composed of sections
  - Personal Contact Information;
  - Organization Information.
- Parallel development can be planned and organized.
- Build up a library of reusable XForms sections.
- Standard and pre-defined XForms constructs that control user input events will save development time and cost.

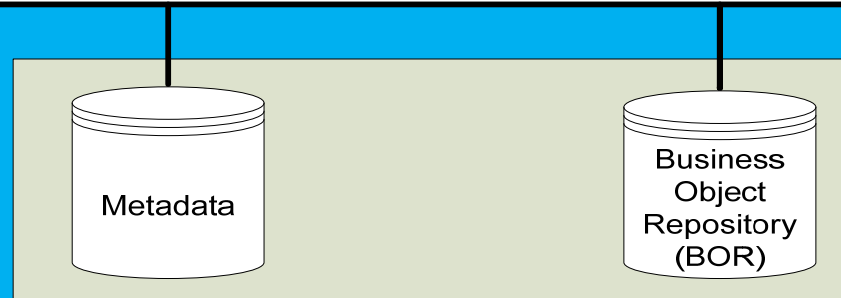
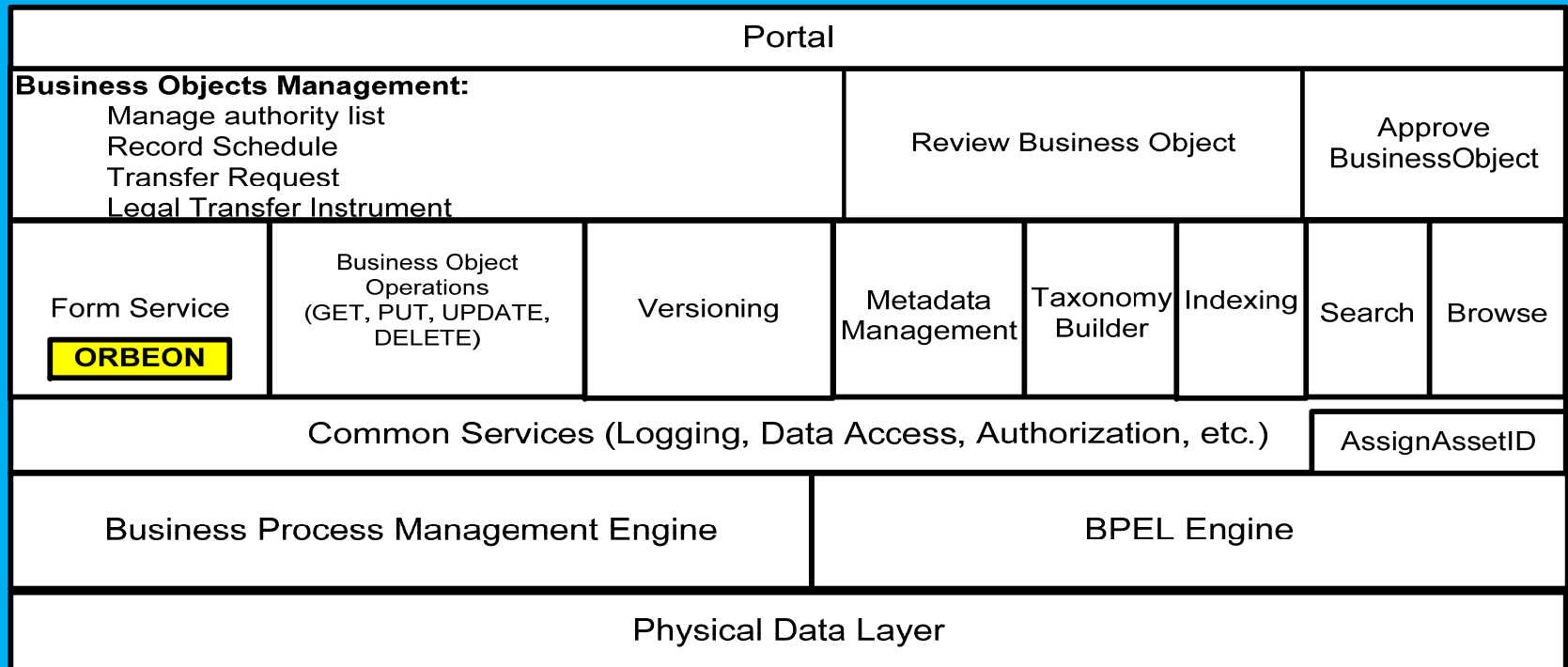
## ❑ No System Requirements imposed on End-User Browsers

- Server-side XForms processing does not impose any requirements on end-user browsers. It is important not to levy any configuration requirements for using the ERA applications on the NARA archivists and agencies' record managers.

# Synergy of Xforms and Genericcode

- ❑ Separate evolvability
- ❑ Separate modularity
- ❑ Separation of control and data
- ❑ XForms `<xforms:instance>` element can point to a file as:
  - a fully-qualified URL
  - pointer to the local file system

# ABO Application: Layered View



**XML DATABASE**

# Summary

- Importance of e-Forms in archival business processing
  - Issues and challenges
- Agility of approach based on XForms and Genericcode
  - Amount of custom code
  - Reusable code
  - Maintainable code
  - Parallel coding
- Show that XML-based approach satisfies system drivers of ERA
- Use of other Open Source Software



# Thank You



<http://www.archives.gov/era>  
[mailto: quyen.nguyen@nara.gov](mailto:quyen.nguyen@nara.gov)



# Facilitating Environmental Data Sharing The Role of Open Source

5 November 2009



# Introduction

- Data Distribution and Sharing at EPA
  - Current Overview
  - Challenges
  - Under Development...
- Open Source Case Studies
- Ideas for the Future



# EPA Data Sharing

- Mandate for data sharing / provision
  - Driven by multiple statutes
  - Regulatory nature of EPA, priority on data collection required to carry out statutory requirements
  - Implementation of many programs is delegated to states
  - However, data sharing continues to be an underlying key principle in our regulatory programs
    - Traditionally has been accomplished within programmatic stovepipes
    - Challenge is continuing to progress towards breaking these stovepipes and providing more comprehensive access to multi-media, multi-program data



# EPA Data Sharing

- Limitations on data sharing
  - Some data collected by EPA regulatory programs is protected under statute (FIFRA, TSCA) as Confidential Business Information
- Known user communities
  - National Dialogue on Information Access
  - EPA as a component in an ecosystem of environmental data producers and consumers:
    - Other Federal agencies, tribal/state/local government partners, environmental groups, communities and concerned citizens and industry
  - Challenge: What is the appropriate role for EPA as broker of some of this environmental data?



# EPA Data Sharing

- Current capabilities for data and services
  - National Environmental Information Exchange Network
    - <http://www.exchangenetwork.net>
  - GEOSS Registries
  - EPA's GeoData Gateway -> Geospatial One Stop (GOS)
  - Data.gov
- Near term future capabilities
  - “EPA Earth”
  - Data Publishing Program



# Case Studies

- Analytical Models
  - Pollutant transport, exposure models, watershed and surface and groundwater flow models, etc.
  - <http://www.epa.gov/CREM>
- BASINS
  - Move to Open Source (MapWindows) at v4.0
- Open Source Exchange Network Nodes
  - Multiple vendors have released open source implementations of EPA EN Node Specification v 2.0
  - Adoption rates are high... 3<sup>rd</sup> party addons starting to be developed



# Looking Forward

- Publishing Program
  - Open feeds and services, in some cases built on FOSS
- Code Repository
  - No matter what the underlying platform may be... share your source, share your intellectual investments
- Data.gov / eGov initiatives
  - Stimulating the open source development community?
  - Code development contests





## Contact:

Jerry Johnston  
Geospatial Information Officer  
[johnston.jerry@epa.gov](mailto:johnston.jerry@epa.gov)  
202-564-3175

## **Questions?**

**Tell us what you think:  
Complete the survey**