GPO's Federal Digital System (FDsys) Innovations, Collaborations, Reliability, & the Future

In 1993, Congress passed the U.S. Government Printing Office Electronic Information Access Enhancement Act (Public Law 103-40), which expanded GPO's mission to provide electronic access to Federal electronic information. In June 1994, GPO launched *GPO Access*, which provides online access to information from all three branches of the Federal Government. For more than 15 years, *GPO Access* was America's source for Government information online. In 2009, GPO unveiled the next generation of Government information online with GPO's Federal Digital System (FDsys). In December 2010, FDsys assumed the role as GPO's electronic system of record for Federal Government information. Later this spring, *GPO Access* will become an archive only, with no new content added, and later this year, *GPO Access* will be retired.

FDsys is not merely a replacement for *GPO Access*. FDsys is…
- A Content Management System ensuring content integrity and authenticity
- A Preservation Repository guaranteeing long-term preservation and access to digital Government content
- An Advanced Search Engine combining extensive metadata creation with modern search technology

FDsys: America's Authentic Government Information

INNOVATIONS

With FDsys, GPO has established one of the Federal Government's only systems for the management, access, and preservation of digital content.
- FDsys securely manages electronic Federal content to safeguard digital content integrity and authenticity.
- FDsys combines extensive metadata with modern search technology allowing the public to find the Government documents they need.
- FDsys follows archival system standards to ensure long-term preservation and access to digital content.
- FDsys provides access to millions of digitally-signed PDF documents proving to users that digital content has not been altered since it was made available by GPO.
- FDsys ensures continued congressional and public access in the event of a disaster through a fully functional Continuity of Operations (COOP) site.

- FDsys is an energy efficient, sustainable technology solution that reduces energy consumption by 40%.

COLLABORATIONS

FDsys is based on the concepts of interoperability and reuse, which puts GPO at the forefront of electronic information sharing between Federal Government agencies. GPO's technology supports the ability for other agencies to meet their missions and has fostered an open environment for information and technology exchange to increase efficiencies in Government. GPO maintains ongoing relationships with other Federal digital libraries and archives, including the Library of Congress, the National Library of Medicine, and the National Archives, allowing for the alignment of efforts between organizations with similar goals. GPO's Federal Government partners look to GPO to implement their mission critical initiatives and collaborate on shared technology solutions that enhance access to Federal Government information.

Office of the Federal Register:
- In support of the Office of the Federal Register (OFR), GPO converted the *Federal Register* and *Code of Federal Regulations* into XML format, provides access to the Public Papers of the Presidents of the United States in XML via FDsys, managed the development of the FR 2.0 website that uses this XML, and facilitated additional access to the XML through the Data.gov and FDsys web sites.
- In support of OFR, GPO developed and released on FDsys the new online-only publication, Daily Compilation of Presidential Documents.
- OFR has requested that GPO lead the development of the next generation of their mission critical eCFR and eDocs systems using the FDsys infrastructure.

Regulatory Information Service Center, Environmental Protection Agency, and Office of the Federal Register:
- GPO is collaborating with the Regulatory Information Service Center, the Environmental Protection Agency, and OFR to provide enhanced access to the Regulatory Plan and the Unified Agenda of Regulatory and Deregulatory Actions.

Library of Congress:
- GPO and the Library of Congress (LOC) are partnering on two major digitization projects, demonstrating converted content preservation and access in FDsys.
- Under the direction of the House Committee on Rules, GPO is working with LOC and the Clerk of the House to provide access to Constitutional Authority Statements through FDsys.

Library of Congress, Congressional Research Service:
- Under the direction of the Senate Committee on Rules and Administration, GPO and LOC's Congressional Research Service are collaborating on a project to enhance public access, through FDsys, to an authoritative version of the *Constitution of the United States of America: Analysis, and Interpretation*.

Office of the Law Revision Counsel & the U.S. House of Representatives:

- GPO worked closely with the Office of the Law Revision Counsel of the U.S. House of Representatives to bring the complete yearly editions of the United States Code collection into FDsys in XHTML and PDF formats, enhancing public access to current and historical content.

National Oceanic and Atmospheric Administration:
- GPO is collaborating with the National Oceanic and Atmospheric Administration to provide access to the Coastal Zone Information Center Collection through FDsys.

Administrative Office of the United States Courts:
- GPO and the Administrative Office of the United States Courts are partnering on a pilot project to provide access to historic and current Federal district, appellate, and bankruptcy court opinions through FDsys.

RELIABILITY

Not only is GPO bringing content from other Federal agencies into FDsys, but numerous other Federal Government sites now rely on FDsys to enrich their user experience including:

- THOMAS and the Legislative Information System (LIS)
    - Congressional sites such as THOMAS and LIS rely on electronic content generated by GPO and link directly to PDF and HTML content managed by FDsys.
- House Legislative Resource Center (Info-Link)
    - The House Legislative Resource Center receives a daily report of new Congressional publications generated through FDsys metadata to support their new publication inventory tracking system.
- Regulations.gov
    - The official site for public commenting on regulations uses FDsys metadata to power their public search and includes *Federal Register* content from FDsys in rulemaking dockets.
- Science.gov
    - The gateway to Government science information has integrated FDsys collections into their federated search function.
- Federalregister.gov
    - FDsys-created XML powers the Web 2.0 version of the *Federal Register*, developed by GPO on behalf of OFR.

THE FUTURE

As GPO looks to the future of FDsys, heavy focus will be placed on increasing the amount of content managed and preserved within FDsys, including at-risk, born digital content and increasing the ways the content within FDsys can be used and reused.

GPO is actively working with Congress and Federal agencies, and will continue to do so, to bring more and more content into FDsys, through partnerships and through various technological methods that are currently in the design phase.

GPO is working to develop the public Application Programming Interfaces (APIs) to facilitate robust system integration with FDsys from non-GPO systems. This activity is planned for FY 2011 and FY 2012 and is currently in the requirements phase.

Additionally, GPO is working on linking publications. This feature will enable search and retrieval of related publications from different collections. The activity is currently in the requirements phase.

These activities all aim at the goal of ensuring that Federal Government content is centrally located, accessible to the public, and enhanced so that it is as easy to use as possible. These activities will also allow GPO partners and others to utilize GPO's data and present it to their users in multiple and unique ways.

***Attend the session on FDsys Development Activities on Tuesday, April 5th at 2:00pm.
***Visit the FDsys demonstration table where you can explore FDsys and ask questions of GPO staff.

FDsys Council Session

San Antonio
April 5, 2011

Agenda

- Program Update – Lisa LaPlant
- Council Q&A – Lisa LaPlant
- Education and Training – Ashley Dahlen

1

1

Program Update

2

FDsys Background

- FDsys automates the collection and dissemination of electronic information from all three branches of government.
- Information is:
  - Submitted directly into FDsys
  - Permanently available in electronic format
  - Authenticated and versioned
  - Publicly accessible for searching and downloading

3

2

What is FDsys?

- FDsys is a *Content Management System*
  - FDsys securely controls digital content throughout its lifecycle to ensure content integrity and authenticity
- FDsys is a *Preservation Repository*
  - FDsys follows archival system standards to ensure long-term preservation and access of digital content
- FDsys is an *Advanced Search Engine*
  - FDsys combines extensive metadata creation with modern search technology to ensure the highest quality search experience

4

Release 1 Goals - COMPLETED

- Establish the foundational infrastructure

- Establish the preservation repository

- Replace the current public site

- Perform a large scale data migration

- Provide operational continuity for the system

5

Foundational Infrastructure

- FDsys search component went live in January 2009
- FDsys CMS component went live in March 2009
- FDsys is an energy efficient, sustainable technology solution that reduces server power consumption by 40% compared to comparable solutions

6

Preservation Repository

- World class preservation repository built on OAIS model went live with CMS launch in March 2009
- Completed an internal audit of FDsys and working to engage the services of an independent vendor to review FDsys documentation and conduct an independent external audit of FDsys

7

Public Access Site

- Worked closely with the library community, agency customers, and congressional customers to design the public site which went live in January 2009

- FDsys became the system of record in December 2010
  - 5.6 million searchable granules
  - 50,000 unique visitors per week

- Sun setting of GPO Access is near completion
  - GPO Access will become "archive-only" later this spring
  - Complete shut down planned later this year
  - Deployed code to handle GPO Access redirects and will turn on when GPO Access is shut down
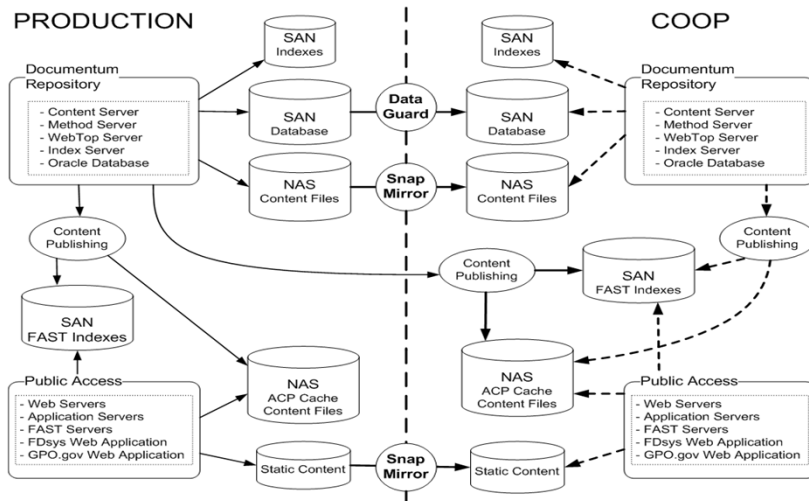
8

Data Migration

- GPO migrated content from GPO Access to FDsys
  - Content collections completed in June 2010
  - One-off publications completed in September 2010
  - Over 209,000 packages currently in FDsys

- GPO Introduced significant improvements during migration
  - Automated metadata extraction process for content collections developed to support search and preservation
  - Data inconsistencies handled by parser and metadata editor
  - Digital signatures added to millions of PDF files

- Remaining quarantined packages are being worked
  - Analysis of multiple storage servers identified full scope of GPO's data universe
  - Review uncovered issues with a subset of content
  - Less than ½ of 1% quarantined packages remaining

9

Continuity of Operations

- Functional Replica of Production
  - Replicates content and metadata
  - Replicates system components including Preservation Repository, CMS, and Public Access
  - Near-real-time data synchronization between Production and COOP
- Disaster Recovery
  - Standby instance for disaster recovery
  - Failover in less than five minutes with full content availability
  - COOP failover and failback successfully tested in February 2011

10

Data Synchronization Architecture

11

FDsys

Collaborations

- Office of the Federal Register

- Office of the Law Revision Counsel of the U.S. House of Representatives

- National Oceanic and Atmospheric Administration Coastal Zone Information Center

- Regulatory Information Service Center, Environmental Protection Agency, Office of the Federal Register

- Library of Congress

- Administrative Office of the United States Courts

12

FDsys

Reliability

- THOMAS and the Legislative Information System (LIS)
  – Rely on electronic content generated by GPO and link directly to PDF and HTML content managed by FDsys.

- House Legislative Resource Center (Info-Link)
  – Receives a daily report of new Congressional publications generated through FDsys metadata to support their tracking system.

- Regulations.gov
  – Official site for public commenting on regulations uses FDsys metadata to power their public search and includes Federal Register content from FDsys in rulemaking dockets.

- Science.gov
  – Gateway to Government science information has integrated FDsys collections into their federated search function.

- FederalRegister.gov
  – FDsys-created XML powers the Web 2.0 version of the Federal Register, developed by GPO on behalf of OFR.

13

Release 2 Goals

- Increase the amount of content managed and preserved within FDsys, including at-risk, born digital content

- Increase the ways content within FDsys can be used and reused

14

Implementation Approach

- Release 2 is being implemented over four planned major releases
  - Releases will include new features as well as identified system fixes
  - Release scope and schedule are determined at the beginning of a release
  - Minor or interim releases will also occur between major releases
  - New features will continue to be added to the FDsys roadmap as needs are identified by stakeholders
  - Priorities may change based on identified needs and will be communicated with stakeholders

15

Release 2: Group A

- System Continuity of Operations (COOP)
- WebLogic Application Server Upgrade
- XME Metadata Editor Enhancements
- Congressional Committee Browse Pages Enhancements
- Public Papers of the Presidents in XML
- Digitized Statues at Large

16

Demos

- Congressional Committee Browse Pages
- Public Papers of the Presidents in XML
- Digitized Statutes at Large

17

Release 2: Group B

- House Journal 2000-2006
- Constitutional Authority Statements
- Coastal Zone Information Center (CZIC) collection
- Unified Agenda in the Federal Register
- Strategic roadmap and design activities to support a basic collection for high-volume converted, harvested, and deposited content

18

Remaining Release 2 Features

Increase the Amount of Content within FDsys

- AOUSC Pilot
- Digitized Bound Congressional Record 1873-1998
- Format Validation using JHOVE
- Quarantined Packages
- Basic Collection for High-Volume Converted, Harvested, and Deposited Content

19

Additional Priority Features

Increase Content Use and Reuse

- API Development
- Linking Publications
- Government Manual in XML

20

Select Related Projects

- CONAN on FDsys
  – Enhance public access to a digitally signed authoritative version of the *Constitution of the United States of America: Analysis, and Interpretation*

- Authentication Working Group
  – Create a community of practice for content originators, stewards, and republishers in the Federal government to discuss standards, policies, and technologies to maintain and communicate content integrity and authenticity

- Composition System Replacement
  – Flexible, interoperable framework for composing and preparing essential documents of the Federal Government

21

Council Questions

Council Question 1

- What information would the FDsys team like from Council and the community that it is not already receiving?

Answer

- What are the key content collections to target for ingest into FDsys and how should new content be prioritized?

- How willing are libraries to partner with GPO on the creation of content packages based on standards?

- Is there a greater priority to get more content into FDsys or make it highly useable (e.g. granules, robust metadata, customized search)?

- What are some of the specific use cases for APIs?

24

Council Question 2

- How can Council and the community best communicate this information and work with the FDsys team?

25

Answer

- Continue to interface through scheduled conference calls
- Smaller working groups on specific issues (e.g. API use cases)
- Direct feedback about FDsys through askGPO

26

Council Question 3

- With respect to communication, what are the deadlines that affect Council so that the FDsys team can best proceed?

27

Answer

- Key activities will be carried out at the working group level, and these groups will have their own deliverables
- Definition of deliverables and the deadlines for completion will be shared with Council in advance to allow time for collaboration on identified tasks

28

Education and Training

29

Training

- Overview of training done
- FDsys training dos and don'ts
- Training currently available
- New initiative to develop training partnerships
- Requesting training

30