

FDsys Program Review

April 7, 2010

Government Printing Office:

Bob Tapella, Public Printer

Ric Davis, Director of Library Services and Content Management / Acting Superintendent of Documents

Mike Wash, Chief Information Officer

Scott Stovall, Director, Programs, Strategy, and Technology

Selene Dalecky, FDsys Program Manager

Federal Depository Library Council:

John Shuler, Associate Professor, Bibliographer for Urban Planning and Government Information / Documents Librarian, University of Illinois at Chicago

Suzanne Sears, Head, Government Documents Department, University of North Texas

Office of the Federal Register:

Mike White, Managing Editor of the Federal Register

Summary:

On Wednesday, April 7, 2010, Bob Tapella, Public Printer, United States Government Printing Office (GPO), convened a public meeting to review the status of GPO's Federal Digital System (FDsys) program. The objective of the meeting was to receive a program status update and to discuss program successes, issues, and opportunities with key stakeholders including GPO's Library Services and Content Management (LSCM) business unit, the Office of the Federal Register (OFR), and representatives from the Federal Depository Library Council. The meeting was also attended by observers from GPO, the House Administration Committee, and the House Committee on Appropriations.

FDsys Program Update:

Background

GPO's Federal Digital System (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, and publicly accessible for searching and downloading. To accomplish this FDsys relies on three key components, a content management system that securely controls digital content throughout its lifecycle to ensure content integrity and authenticity, a preservation repository that follows archival system standards to ensure long-term preservation and access to digital content, and an advanced search

engine that combines extensive metadata creation with modern search technology to ensure a high quality search experience.

Release 1 Goals and Status

The goals of Release 1 are to build a foundational infrastructure, establish a preservation repository, replace the current public site, perform a large scale data migration, and provide operational continuity for the system.

- The foundational infrastructure was put in place with the launch of the search component in January 2009 and the Content Management System (CMS) component in March 2009. The CMS component provides a place where text, locator, and PDF files can be stored in one location.
- GPO's world-class preservation repository went live in March 2009. The repository is built upon the Open Archival Information System (OAIS) model and provides sufficient control to ensure long-term preservation and access.
- The FDsys public interface launched with the search component in January 2009. GPO worked closely with its stakeholders from the library community, Federal agencies, and Congress to customize each collection based upon stakeholder requirements. Usability studies were conducted to ensure FDsys could meet user needs at all levels. GPO has held extensive Beta testing which yielded positive feedback from the user community and resulted in FDsys being named one of the "10 Great Government Websites" by *Government Computer News*.
- A large scale data migration effort began in January 2009, and to date, 29 out of 40 content collections have been completed. The remaining collections will be complete by June 30, 2010. Significant improvements were made as part of the migration process. Extensive analysis was performed on multiple GPO data storage servers to identify the full scope of GPO's content universe. An automated metadata extraction process was developed to support search and preservation, and inconsistencies in the data are handled manually via the metadata editor and programmatically by content parsers. The use of extracted metadata along with dynamically generated browse tables has significantly reduced the amount of manual work needed to provide access to publications on a daily basis. Furthermore, as new content collections are released a digital signature is applied to PDF files helping to assure the public that a document has not been changed since it was published by GPO.
- The system continuity and failover implementation is currently in process. The continuity of access site will be operational by August 2010, and the full system failover capability will be operational in December 2010.

In addition to working towards Release 1 goals, GPO engaged in other high profile projects to support internal and external stakeholders. In conjunction with the change of administration, the *Daily Compilation of Presidential Documents* was released in February 2009. In support of the Office of the Federal Register (OFR), GPO developed XML for regulatory publications available

through FDsys and Data.gov. The Federal Register in XML was released in September 2009 and the Code of Federal Regulations was released in December 2009. GPO worked closely with the Office of the Law Revision Counsel of the U.S. House of Representatives to bring the yearly virtual main editions of the United States Code collection into FDsys in XHTML and digitally signed PDF formats. In addition, GPO has been working with the Law Library of Congress on a project to provide permanent public access to the digitized 1951-2002 volumes of the Statutes at Large in accordance with a specification acceptable to staffs of the Library and GPO. These volumes comprise approximately 147,000 pages of content. The Law Library has performed the actual conversion of the document from print to digital content via scanning and has provided GPO with the content. We now propose to process it to allow for robust searching by users and content management capabilities, including creating archival as well as access copies that are authenticated (a digitally signed PDF with searchable text). The processed content will then be made available online by GPO via FDsys, and GPO will provide the processed content back to the Library to be made available through its THOMAS system.

FDsys puts GPO at the forefront of the President's Open Government Initiative. FDsys is based on the concepts of interoperability and reuse. FDsys content is available in all major search engines. Citizens are finding new ways to interact with key public content through XML data. Examples include sites like FedThread.org and GovPulse. Other government sites rely on FDsys to enrich their user experience. Regulations.gov uses FDsys metadata to power their public search, and Science.gov has integrated GPO collections into its federated search gateway to government science information. GPO, OFR, and the Environmental Protection Agency (EPA) are working with the Office of Management and Budget (OMB) to improve the interoperability of data between federal rule making systems. GPO is working with OFR on the next generation of the Federal Register (FR 2.0) to allow citizens to better participate in the development of the policies of their government. GPO is working with the Administrative Office of the U.S. Courts on a pilot to publish federal district and bankruptcy court opinions via FDsys, and stakeholders continue to approach GPO with new ideas and projects.

Key components of completing Release 1 include the transition from *GPO Access* to FDsys and providing for system continuity and failover. The transition from *GPO Access* to FDsys is in process. Key activities have been identified and a completion plan has been developed. Once completed, FDsys will assume the role as GPO's electronic system of record. Sunset of *GPO Access* is planned by the end of the fiscal year. A phased approach will be used to implement continuity and failover for FDsys. Phase 1 will provide continuity of public access to historic and new content, with the ability to recreate the full system from backups. Phase 2 will provide full system failover capabilities. The design for system continuity and failover has been developed in conjunction with subject matter experts. An independent design validation has been performed, and a formal design review was conducted internally after the independent design activity.

Release 2 Goals

FDsys Release 2, which includes a key new component of external submission, is underway. The Program Management Office (PMO) is currently working with GPO's Congressional Publishing Services on Congressional content workflows. A plan is being developed for prototype content submission functionality by the end of fiscal year 2010. The full content submission component is scheduled to be complete by October 2011 based on an \$8 million level of funding. Beyond submission, additional functionality will extend access functionality and data usability. With the

foundational system in place, there will be a strong emphasis on bringing more content into FDsys. GPO is working with stakeholders to review remaining features and requirements, and new collection development will be “productized”, allowing collections to be added independent of feature development. New features will continue to be added to the FDsys roadmap as needs are identified by stakeholders.

Cost Summary and Analysis

The FDsys program began in 2004 with market research, a concept of operations, and program requirements. The PMO formed in October 2005 to support FDsys acquisition and development. The acquisition process started in 2005, resulting in an award for Master Integrator to Harris Corporation in August 2006. GPO recognized early in the Master Integrator phase that the contractor was over spending and under performing. In late 2007, a program redirection was planned and executed, and progress on the program continued without a major disruption. With GPO as the prime integrator, system development was more cost effective and delivered performance was consistent with our requirements. The estimated cost to complete Release 1 was reduced from \$62 million to \$42 million, saving \$20 million and allowing GPO to establish a strong core program team.

Content migration activities have had a direct impact on program costs. GPO meticulously converted and migrated content into FDsys to deliver effective content management and robust search, but the cost and time required to migrate content from *GPO Access* to FDsys was under scoped. The cost for content migration was \$11 million, which represents over 25% of the total estimated cost to complete.

Going forward, the yearly support cost for FDsys is \$3.25 million which includes \$1.75 million for software licenses and labor and \$1.5 million for technology refresh. \$8 million has been requested for fiscal year 2011 for Release 2 to complete the core system. Future development costs for system enhancements are projected to be \$4.5 million per year.

Program Risks

Two key risks have been identified for the FDsys program. First, there is risk associated with a delayed completion of the core system. Mitigation steps include maintaining sufficient investment to complete the core system and preventing loss of key resources resulting in more cost and time. Second, there is a risk of confusion and frustration within the stakeholder community regarding *GPO Access* retirement/sunset. A mitigation step is to develop and implement plans to address policy, training, and communication issues prior to sunset.

Lessons Learned

There have been a number of lessons learned on the FDsys program to date. First, it is important to have the right resource mix for system design and development. The Master Integrator approach did not have the right mix of resources and skills were developed while on the program. Second, adequate testing is required to ensure robust system performance, but resources required for testing were underestimated, and GPO is working to improve this. Third,

good data migration is complex and difficult. Finally, GPO needs to be more precise on published schedules.

Stakeholder Views and Discussions:

Library Services and Content Management – Ric Davis

Beginning with the days when FDsys was simply an idea, the Library business unit has proudly worked closely with the Program Management Office (PMO) in the brainstorming, documentation, and creation of this system. Numerous former unit staff members were utilized from the early days of this effort in working on the Concept of Operations, the Requirements Document, and even being detailed to the PMO to support this endeavor. I want to begin by thanking staff who have worked together to successfully launch this system.

The Library unit and the Federal Depository Library Program (FDLP) I administer are deeply vested in the success of FDsys, as it answers critical needs of my business unit, the FDLP community and the public now and going forward in meeting critical business requirements. Most critically is being able to manage the data, from a content management perspective, that has often been manually based in the past.

I am here today to talk to you from a voice of customer perspective as a business unit, in terms of key successes, issues that I see with FDsys and opportunities for the future, in terms of several high-level key areas for each category.

Successes

Ease of Search

- The FDsys capability to search across all content in the system is stellar. With our current *GPO Access* capability, publications are stored in nearly 5,000 separate databases.
- The filtering capability that allows users to narrow their search is something that the FDLP stakeholders have been requesting for a long time.
- The precise nature of the search is also key. FDsys does an excellent job of pointing the user to the exact content for which they are looking.

Metadata Files

- The MODS and PREMIS files are a key component of what the FDLP community has been asking for in order to better serve their customers. Additionally, the ZIP files that provide all format and metadata files are convenient and beneficial features.

Authentication

- The continuation of the authentication efforts that were begun in the library unit is exactly the direction that the FDLP community wants to head. We often hear the call for “more authentication.” It is important that we continue to work with content originators as well, to ensure the chain of custody of electronic information made available when feasible.

Preservation

- Providing current and permanent public access to content through FDsys is of paramount importance. FDsys is being built to handle a variety of file formats for ingest, for deposited, harvested, and converted content. Born digital materials are at particular risk, and it is critical that agencies be able to make this content available for easy ingest into FDsys, and that GPO enable this capability.
- Of equal importance, is the recognition that any content made available today, will require a flexible platform to migrate and refresh this information in the future, as these formats continue to change.

XML

- The Federal Register and Code of Federal Regulations in XML, as part of the open government initiative, is also something that is extremely beneficial for us, in terms of enabling new applications to be developed in personalized and customized ways.

Issues

Schedules

- Understanding the challenges of setting up a world-class system such as FDsys, we should also keep in mind the importance of adhering to the deadlines for releases that we promise the user community. During various presentations at library conferences, through various PowerPoint presentations, and in issues of Techline, release dates were presented that were not always met. It is important that as timelines are put forward in the future, consideration be given as well to the likelihood they will be met according to specific dates. The library community has suggested that a series of priority steps should be communicated more often. Additionally, an update of the Systems Releases and Capabilities document or comparable project plan made publicly available would help show a clear plan of what is coming, with associate milestones and deliverables.

Costs

- It is important to have a detailed cost schedule for the remainder of FDsys implementation as well as an operating cost budget that can be planned for. The FDLP S&E appropriation will support the operation of FDsys in terms of providing in-scope FDLP content, just as we have done with GPO Access for the past 15 years, at approximately \$2 to \$2.5 million per year. Knowing what these costs are will help better plan future appropriations. It is our expectation that through the automation technologies FDsys will enable as a world class system, daily operating costs incurred should actually decrease.

Opportunities

Original Requirements

- It is important that the original requirements worked on years ago be considered going forward, and revalidated with the business units at GPO and the library community if necessary. There are approximately thousands of requirements identified for FDsys. Many, but not all, of these original requirements in the Requirements Document have been implemented. It is important to make sure that requirements in original releases are met before moving on to future releases, and for those not addressed, a statement as to

why they were not met or eliminated in coordination with the customer business units. Additionally, it is important to demonstrate to the library community and other stakeholders a checklist that maps the requirements to what has been actually implemented. In other words, what “features of FDsys” (e.g., access, metadata, persistent name associated with PURLS) have addressed which “requirements.” We should be able to “check-off” on each requirement and state it was addressed in FDsys implementation. Finally, it should be validated that all current *GPO Access* functionality continue in FDsys before this system is shut down as the system of record.

Continuity of Services

- There is concern among the FDLP community that in the transition from *GPO Access* to FDsys that there will be legacy system failure, loss of data, service interruptions, etc. Demonstrating a plan to the library community for mitigation of these perceived risks would gain goodwill and trust. If the primary servers operating FDsys today were to fail, there needs to be an assurance that there is both a backup of all data and that service can be restored seamlessly to customers through a real-time failover to backup servers. It is also important that this back-up be at a remote off-site location separate from the main GPO building in Washington, D.C.

Integration of Legacy Applications

- As we look ahead, it is important that legacy applications map to FDsys in the future. Of particular interest is establishing metadata crosswalks with the Integrated Library System (ILS) and the FDLP Desktop search capability, along with ensuring touch points with other internal systems as appropriate.

Digital Dissemination

- While having a repository of content available at GPO is critical, there are opportunities to facilitate the availability of digital collections in libraries. Some in the FDLP community have expressed strong interest in having Access and/or Preservation level files digitally deposited in FDLP libraries. This will further the model established for tangible collections of content by having dispersed collections of electronic content, and through partnerships better ensure access and preservation of content.

Depository Library Council – John Schuler

Successes

Mr. Schuler suggested that the work completed thus far on FDsys demonstrates GPO's ability to find innovative and creative ways to change, without changes being made to Title 44. It showed the ability of the agency to adapt within a dynamic and changing environment. Additionally, it furthered efforts related to standards, which are critical for electronic content.

Issues

Mr. Schuler noted that release dates and schedules for FDsys need to have meaning, and that it is important that when dates are set and publicly announced, they are met going forward. He requested an open dialogue at the upcoming library council meeting on exactly where FDsys is in terms of progress and timelines, particularly related to requirements and release dates that were established in past years. Additionally, he mentioned it would be beneficial to have a

complete listing of all systems related to the FDLP that are not in scope of the next FDsys release, because many in the depository library community assumed FDsys would be all encompassing.

Opportunities

Mr. Shuler mentioned that the ingest capabilities of FDsys offer many opportunities for more partnerships between GPO and Federal depository libraries, and he would like to see more content ingested into the system that the library community has digitized, along with at risk federal content held by other agencies.

Depository Library Council – Suzanne Sears

Successes

Ms. Sears feels that FDsys demonstrates a consensus and shared vision between GPO and the depository library community regarding the importance of permanent public access to Federal Government information. She also feels it is a success to no longer be dependent on the WAIS platform and to have better search functionality. The FDsys advanced search feature is similar to that used in commercial databases and is easier for students to use and understand.

Issues

Ms. Sears noted that a better method of communicating changes to timelines and the reprioritization of features within a release is needed. The depository community would like GPO to be more specific on when features of FDsys, such as ILS integration and persistent naming, will be available, and feels priority should be given to both. Additionally, the capturing and distributed preservation of born digital data is still an issue of concern for Ms. Sears as this material is at risk. Finally, she expressed some apprehension about the cost to add new collections to FDsys.

Opportunities

Ms. Sears suggested that providing further clarification to the depository library community on the FDsys stakeholders (i.e. Congress, the Executive agencies, the general public), would help them to better understand why the system is being developed as it is. She is also interested in what methods FDsys will make available for depository libraries to submit content they have digitized and in what ways digitized content on FDsys will be made available to depository libraries for distributed preservation.

Office of the Federal Register – Mike White

Successes

Mr. White emphasized the standpoint of OFR, as a customer of GPO and also as a partner with GPO. He feels that the face of the Federal Register will be noticed through FDsys. Mr. White praised FDsys features such as the search capabilities, great metadata, and browse tables. He noted that FDsys provides a foundational infrastructure for more versatile and personalized services that may be utilized for Federal Register 2.0. FDsys components could help build a new face of the Federal Register.

Issues

Mr. White conveyed some disappointment that the electronic Code of Federal Regulations, the e-CFR, would not be available in Release 1 of FDsys. The e-CFR, an innovative service as it is, has not yet reached its real promise. By working hard together, the e-CFR vision will be realized. The efforts put into FDsys are for a good end, possibly revolutionary. Mr. White looks forward to this functionality being included soon.

Opportunities

Mr. White expressed a desire to become more involved and in-touch with Federal depository libraries and their services in relation to OFR products. A better understanding of the usage of the Federal Register in libraries was specifically noted. Mr. White offered thoughts on the potential for shared content management responsibilities and noted possible options for collective contributions. Mr. White also highlighted the importance of continuation of access to the Federal Register with regard to the nature of the Federal Register as a fundamental tool to Government functioning.

Key Discussion and Questions:

1. How have additional projects impacted the Release 1 schedule?

In addition to working towards Release 1 goals, GPO has engaged in other high profile projects in parallel to support internal and external stakeholders. Examples include the Federal Register and Code of Federal Regulations in XML projects. Engaging in high profile projects has impacted the program schedule by slowing the rate of migration of content from *GPO Access* to FDsys.

2. Has the scope of Release 1 changed from the originally planned functionality?

The goals of Release 1 are to build a foundational infrastructure, establish a preservation repository, replace the current public site, perform a large scale data migration, and provide operational continuity for the system. To this end, the scope of Release 1 was modified so that Congressional and agency submission, ILS integration, and persistent naming functionality are now within scope of a subsequent release.

3. Will there be any impact to the ILS when *GPO Access* sunsets?

The sunset of *GPO Access* will not impact GPO's Integrated Library System (ILS) or its public interface, the Catalog of U.S. Government Publications (CGP). FDsys integration with the ILS, which will allow for metadata crosswalks between the two systems, is planned for future releases. There are no plans for FDsys to subsume the ILS or CGP.

4. Is there a continuity plan to recover data and rebuild the system in the event of a disaster?

Critical FDsys data including content, metadata, source code, and documentation from the FDsys Development, Test, and Production instances is currently backed up off site to the Alternate Computing Facility. These processes are detailed in the FDsys Backup

Process Documentation. In the event of a disaster, GPO would use this data to rebuild FDsys on new hardware. Furthermore, continuity and failover for FDsys is being implemented in a phased approach. Phase 1 will provide continuity of public access to historic and new content, with the ability to recreate the full system from backups. Phase 2 will provide full system failover capabilities. The continuity of access site will be operational by August 2010, and the full system failover capability will be operational in December 2010.

5. How would the system and schedule be impacted in the event of the loss of its current IT development staff?

GPO has system design, system build, and backup process documentation that would allow system administrator and subject matter expert development staff to review and become familiar with the daily processes and, if necessary, rebuilding or continue the development of FDsys. The loss of system administrators and subject matter expert development staff would have a negative impact to system maintenance and development schedule.

6. Is FDsys in competition with USA.gov and other government information websites?

FDsys is not in competition with USA.gov and other government information websites. USA.gov is described as a web portal that makes it easy for the public to get U.S. government information and services on the web. The purpose of FDsys is not to serve as a portal, but instead to provide access to official and authentic content from all three branches of the U.S. government on our site, and through links to official agency and partnering web sites. Our main system functions encompass publishing information, enabling searching for information, preserving the information, and providing version control.

7. Has a list of tasks and schedule been created for Release 1 completion activities?

A detailed task list and schedule have been created for Release 1 completion activities. Once completed, FDsys will assume the role as GPO's electronic system of record. Sunset of *GPO Access* is planned by the end of the fiscal year.

8. What impact will FDsys have on the availability of print?

FDsys, as a system, does not determine the future of print. The list of *Essential Titles for Public Use in Paper or Other Tangible Format*, or simply the *Essential Titles List*, is a list of publications developed by the FDLP community that are deemed essential for the public and to Federal Depository Libraries in tangible format. So long as they are published in paper by the originating agency, titles on this list, even if available on FDsys electronically, will still be printed and made available to depository libraries in tangible format. If a title ceased to be published in paper and was only available online, the originating agency will hopefully choose to consider the electronic version official.

GPO is responsible for ensuring that all publications are disseminated and/or distributed to depository libraries in the most effective manner possible. Specific guidelines are

followed to determine the most appropriate method of dissemination. After the determination is made to include a product in the FDLDP, GPO ascertains the best method for making the government information product available to depository libraries. Criteria evaluated includes: (1) how the product will be used; (2) specific characteristics of the online and/or tangible product; (3) issues relating to permanent public access; and (4) cost of providing the material. With multiple formats available for many publications, specific product characteristics may dictate one format is preferable to another for library use.

9. When will FDsys be complete?

FDsys has been developed as a flexible and extensible digital system that is able to adapt to emerging stakeholder needs without requiring a complete system redesign, and therefore will never be considered complete. While there is no planned final completion date for FDsys, \$8 million has been requested for fiscal year 2011 for Release 2 to finish core system functionality. Additionally, GPO plans to continue to add new Federal content as well as manage the existing data collections over time.

10. What is the impact of flat or reduced funding on Release 2?

The impact of flat or reduced funding on Release 2 will be reduced scope for planned Submission functionality. The subsequent impact on functionality is not completely known at this point, but may include reduced functionality in the following areas: web interfaces for file upload and submission, content and order tracking through FDsys, and tools for collaboration, delegation of approval authority, and workflow management for GPO's content submitters (e.g., Federal Agencies and Congress).