GPO's Federal Digital System (FDsys): Open and Transparent Government

Evaluating Federal e-Life Cycle Management – A Town Hall Meeting with OSTP, NARA, and GPO

Prepared Remarks from ROBERT C. TAPELLA, *Public Printer of the United States*

102nd American Association of Law Libraries Annual Meeting and Conference

Washington Convention Center in Washington, DC | July 27, 2009

GPO U.S. GOVERNMENT PRINTING OFFICE | KEEPING AMERICA INFORMED
www.gpo.gov

ROBERT C. TAPELLA, *Public Printer of the United States*

Good afternoon. My name is Bob Tapella. I am the Public Printer of the United States. It's a pleasure to be here at the AALL Annual Conference and participate in this town meeting on Federal E-life Cycle Management.

As I was sitting with Mary Alice Baish during the opening session keynote on Sunday, I was struck by Jonathan Zittrain's fascinating talk about how the runaway success of the Internet is on a path to a lockdown, ending its cycle of innovation.

He expressed some great concerns about the dot com world and their interests in controlling digital property and content.

And then he went on to offer some great ideas for developing new technologies that allow users to work creatively and collaboratively.

Our discussion today, in many ways, is about how those of us in the dot.gov world are doing just that.

So please, sit back and relax and see if we're meeting Professor Zittrain's challenge from a government perspective.

I'd like to begin my remarks by taking a step back in time, because I believe our past often provides a glimpse into our future.

Distribution of government documents in the United States was first authorized by Congress in 1813. This law called for one copy of the documents of our democracy issued by the Senate and House to be sent to each university and college, and to the official historical society in each state.

At first the Secretary of State assumed the duty of distributing these documents. Later, the Clerk of the House, Librarian of Congress and even the Secretary of the Interior had the responsibility.

In 1895, when Congress combined various laws relating to printing into one comprehensive package, this responsibility transferred to the Public Printer. Before this new printing law, the distribution and storage of government documents was a haphazard operation. The number of copies of a document ordered did not usually relate to the amount of public or official interest in the topic.

Of the 420 official repositories, some were overwhelmed by mountains of government documents, while others received no distribution at all. In addition and perhaps more distressing, no standard system for titling government documents existed. Consequently, practical cataloging was virtually impossible.

The problems concerning the receipt, care, and distribution of government documents were perplexing, but F.A. Crandall, the first Superintendent of Documents, gave some indication of his grasp of conditions in an 1896 report to the Public Printer. Crandall said:

> There is probably not a man living who could tell off hand what is the subject matter of part 2, of part 2 of volume 4 of part 5 of No. 1 of part 2, volume 14, of the House Executive Documents of the Fifty-first Congress, second session. One might as well try to commit to memory lettering on a Chinese tea chest as charge one's mind with such rigmarole.

The disarray in the cataloging, distribution and storage of paper documents facing the Public Printer in 1895 is not much different than the disarray in the cataloging, distribution and storage of electronic documents we all face today. There is a solution for this electronic disarray – and we call it: e-lifecycle management.

Through the Federal Depository Library Program (FDLP), and its more than 1,200 library partners, the Government Printing Office has and continues to acquire, classify, catalog, and disseminate tangible content to libraries and manage electronic publications through GPO Access and the new Federal Digital System (FDsys).

Earlier this year I wrote President Obama and laid out how GPO could help support his vision for transparency and open government as well as digital access to much more information from the government.

Specifically, I proposed five goals and accompanying actions that GPO is prepared to undertake. The goals are:

First, position GPO's Federal Digital System as the official repository for Federal Government publications;

Second, enable and support Web 2.0 functionality through FDsys to support comments on pending legislation;

Third, establish a demonstration project to apply Web2.0 features to rulemaking documents;

Fourth, participate in and lead efforts to standardize electronic publishing formats; and

Finally, **fifth**, link the White House Web site to FDsys for public searches of Government documents.

From 1994 until this year, *GPO Access* was the information storage system used to house the electronic documents of our democracy. It served the nation well, but had a number of limitations.

There were limited file formats and summary information about documents. While there was full text search capability, there was limited access to supporting metadata[1]. The rigid query language used basic Boolean operators that provided only basic search results; but provided no filtering, sorting, or query refinement capabilities. For the skilled librarian who knew what they were looking for, *GPO Access* worked well. But for the rest of the world, its limitations made it appear obsolete. That's why over the past five years, GPO developed and earlier this year released the Federal Digital System.

GPO Access is evolving into GPO's Federal Digital System, a world-class information management system developed to authenticate, preserve, version, and provide permanent public access to Government information that is within scope of the FDLP. FDsys allows one to:

- Search across multiple collections with a single search box;
- Narrow search results through filtering options;
- Locate authentic documents with digital signatures[2];
- Search within search results; and
- Browse government publications.

In short, FDsys automates the collection, management and dissemination of electronic information from all three branches of government. With FDsys we have a created three great systems in one:

- A content management system that manages digital content in a secure environment throughout its lifecycle to ensure the content is not alter or damaged;
- 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 the highest quality search.

We have deposited just over 163,359 documents in FDsys and we expect that all of our existing GPO Access content will be deposited this fall, as well as new information products.

We have also just completed a Request for Proposals (RFP) to convert existing tangible documents for inclusion in FDsys. This mass digitization project is critical to significantly and effectively add vast amounts of information into FDsys. Our goal is to have a complete historical record, from the founding of the nation to the present, of all documents within scope of the Federal Depository Library Program.

This RFP is pending approval by the Joint Committee on Printing at this time so I cannot share the details, but I believe that we have hit a home run with this procurement.

---

[1] Metadata is information that facilitates the interpretation, management, and location by describing essential attributes and significant properties.

[2] Digital signature is a cryptographic code consisting of a hash, to indicate that the data has not changed, encrypted with the public key of the creator or the signer. A digital signature identifies the signer and verifies the integrity of the data.

GPO downloads or harvests online publications posted on Federal agency Web sites in order to preserve permanent access to these publications. We currently use manual and semi-automated methods to harvest.

Manual means human intervention – GPO staff working with Federal agencies, reviewing print order and notification forms, and responding to alerts from the library community; semi-automated meaning GPO staff uses a tool called Tele-port Pro which crawls agency web sites. Future releases of FDsys in the coming years will include tools to allow GPO to perform automated harvesting.

In addition to the publications from the Executive and Legislative branches, GPO is also working with the courts to fully support their content. Today GPO hosts the Supreme Court Web site, but most of their content isn't available in GPO Access or FDsys. We are in discussions with the Supreme Court to see how FDsys could help them provide authenticated content.

As you know, I'm a printer and in the print world, the authenticity of printed documents is relatively easy to preserve. For nearly 150 years the words "United States Government Printing Office" in printed U.S. Government documents identified GPO as the official disseminator of those documents and assured users of their authenticity.

In a digital world, authentication becomes much more difficult as technology makes electronic documents easy to alter or copy, leading to multiple, non-identical versions that could be used in unauthorized or illegitimate ways.

To help meet the challenge of the digital age, GPO has begun implementing digital signatures to select GPO electronic documents. The visible digital signatures on online PDF documents serve the same purpose as handwritten signatures or traditional wax seals on printed documents. A digital signature, viewed through the GPO Seal of Authenticity, verifies document integrity and authenticity on GPO online Federal documents, at no cost to the customer. These signatures not only establish GPO as the trusted information disseminator, but also provide the assurance that an electronic document has not been altered since GPO disseminated it. This is especially important when we are required to preserve these documents for future generations.

Congressional Bills, Public and Private Laws, and the *Budget of the United States Government* have been digitally signed and are currently available on FDsys. Next to come with digital signatures are the following titles:

- Bound *Congressional Record*;
- *Statutes at Large*;
- *Congressional Directory*; and
- *US Government Manual*.

While there are a few collections currently in FDsys that are not yet signed, we continue to move forward. Next to be signed are:

- *Federal Register*,
- *Congressional Record*,
- *Congressional Reports*, and
- *Daily Compilation of Presidential Documents*.

Just locking a PDF isn't good enough though. We are also working to provide more granular authentication[3] to enhance the usability of this content. We don't have a solution yet, but we are actively pursuing it. With legal and regulatory materials it is also very important to maintain the chain of custody showing exactly how each law or regulation was enacted or came into force. This provides the basis for the electronic version to be as official as a printed edition.

One of the most critical differences between *GPO Access* and FDsys is the use of metadata. As you know, metadata is data that describes other data. In an electronic repository like FDsys, bibliographic information about a publication is in the metadata. Structured metadata based primarily on cataloging processes vastly improves the accuracy and efficiency of a search.

The knowledge and skills the library community has in cataloging has added incredible value to structuring metadata for electronic access. That's why we spent so much time working with the library community to develop the metadata standards within FDsys.

GPO is perfectly positioned to be central in developing metadata for Federal publications given our skills in publication cataloging that goes back more than a century and our investment in understanding world-class content management systems.

Next month GPO will release the XML version[4] of the Federal Register. This is one of the items I wrote to President Obama about and we are delivering on our promise. Furthermore, when the Code of Federal Regulations is part of FDsys, it will also be available in an XML format. This is a major step forward to enable efficient re-use of this important information.

GPO is also starting to talk with groups about Application Programming Interfaces[5] or "APIs", with the goal of developing a specification that will help enable other systems to access our data. In the future we will be announcing a pilot program in this area to facilitate Web 2.0 applications using FDsys content.

---

[3] Granular authentication is verification that the content is authentic or official at the section, part, or chapter level of a document. Currently GPO is able to authenticate only the entire publications. In the future, GPO will also be able to authenticate a subset of the publication, such as the President's inaugural address which is contained in the Daily Compilation of Presidential Documents.

[4] XML is a mark-up language similar to HTML but uses tags that indicate logical structure in addition to display specifications..

[5] An API is a programming interface that calls upon functions in one system to access features or data from another system. This allows, for example, the exchange of resources or information between and among systems.

It seems as if I began my remarks more than a century ago, and then I quickly moved to the present and now I'd like to talk about where we should be going in the future and invite your participation.

I believe that the Federal Government has an obligation to provide complete legal and regulatory information online in an electronic format that is fully usable by the American people free of charge.

We should start with the Constitution Annotated showing how Supreme Court Decisions have affected Federal Statutes all the way through *Public Laws*, *Slip Decisions*, *Code of Federal Regulations*, and proposed laws and regulations. The advice and input from the AALL and its members, along with other stakeholders will help GPO refine this vision into an actionable plan. We need a citation standard[6] that is media neutral. This citation standard needs to be in the public domain and support permanent links so that the fidelity and integrity of documents will be maintained indefinitely.

We also need to create smarter systems that realize the value that XML and Web 2.0 can provide. As one example, the *Federal Register* is a daily publication and a regulation published in the *Federal Register* is incorporated on an annual basis into the appropriate *Code of Federal Regulations* volume.

GPO, working with the Office of the Federal Register, is looking at how we can create a so-called "point in time system." With such a system you could dial in a specific date and see what the regulations in effect were on that date. What if we could create a point in time capability for the entire legal and regulatory framework of the United States? It's an ambitious goal, but one that is worthwhile and one that we are pursuing at GPO.

To that end, this fall we will be holding a symposium at GPO entitled: *Content Creation, Management, Authentication, Dissemination and Permanence of the Complete Record of Federal Information in a Web 2.0 World*.

It's a mouthful, but an appropriate next step. We'll be inviting the AALL and its members to participate.

So, there you have it – my view of how GPO is living e-lifecycle management of Government information and hopefully how we, as a dot gov, are living up to Professor Zittrain's challenge.

Now, before I give up the podium, a commercial message:

GPO is in the Exhibit Hall Booth 830 if you'd like to see FDsys first-hand or talk to one of our very knowledgeable staff members please stop by!

Thank you.

---

[6] A citation standard is an accepted method for recording information about a publication that allows users to identify and locate that publication.