

Information Brief

April 2007

Digitization and the FDLP

BACKGROUND:

In "A Strategic Vision for the 21st Century", released in December 2004, the U.S. Government Printing Office put forth a plan to digitize all Federal publications back to the earliest days of the republic. The Vision included the development of a Digital Content System, as put forth in plans for GPO's Future Digital System (FDsys). One of the strategic goals of the Vision is to digitize a complete legacy collection of tangible U.S. Government publications to ensure no fee public access to these materials that remain available, in the public domain, for permanent public access. Initially GPO was charged by the Public Printer with digitizing the entire corpus of this material.

INITIAL STEPS

GPO began the digitization effort in 2004 by convening two meetings of experts, the first on Digital Preservation Masters and the second on Preservation Metadata. In 2005 GPO conducted a survey of the depository community to assist GPO in determining digitization priorities. At the same time GPO began developing the digitization specifications for converted content. The reports of the experts' meetings, digitization priorities, and the current version of the Specifications for Converted Content are available from *GPO Access* at: <http://origin.www.gpoaccess.gov/legacy/>.

DIGITIZATION DEMONSTRATION PROJECT

The conversion of tangible materials began in July 2006 as a six-month demonstration project, as directed by the Joint Committee on Printing (JCP). The focus of the project was the continuous improvement and validation of GPO's digitization specifications. The project was completed in December 2006.

In January 2007, GPO arranged a meeting of representatives from Federal government, academic libraries, and others in the information community. The goal of the session was to review and provide feedback to GPO on the converted content produced by Digital Conversion Services during the project. Though some suggestions were made to include additional elements, the group consensus validated that the digitization specifications for preservation level scanning were acceptable. One of the suggestions put forth by the reviewers was to have GPO focus on a special role or niche for digitization as it relates to Federal publications. Issues raised by the group include:

- Visual comparison of the documents, and an express wish to be able to view the documents at their desktops;
- Scope of material for GPO to scan;
- Metadata schema utilized and the role of metadata in the existing and future cataloging processes; and
- Technical approaches as related to paper degradation, restoration, grayscale vs. color, lack of policy/planning to deal with these types of issues.

OTHER EFFORTS:

GPO, the Library of Congress and the National Archives and Records Administration, as well as Federal depository libraries, are coordinating and developing partnerships on digitization efforts to maximize the benefits each partner brings to the digitization of the legacy collection.

GPO is one of several agencies participating in the National Digital Strategy Advisory Board's Digitization Standards Working Group. The objective of the Group is to develop common Federal digitization standards for still images that are comprehensive and objectives-based. Common standards will provide consistency across the Federal government, promote collaborative digitization projects, reduce the possibility of duplicative efforts by agencies, and reduce digitization expenditures. One of the first tasks before the group is to identify and prioritize the core standards that will be addressed.

CURRENT STATUS:

- Preparing to report on the results of the demonstration project to the JCP;
- Reviewing and refining specifications based on the results of the demonstration project;
- Reviewing the scope of material for GPO to scan;
- Working with the Project Management Office on the ingest of material into FDsys; and
- Awaiting final direction from the JCP on GPO's role for digitization of legacy materials.

PRELIMINARY ASSUMPTIONS AND QUESTIONS PENDING FINAL DIRECTION FROM JCP

GENERAL ASSUMPTIONS

1. Digitization is an accepted reformatting method for preservation and access to a range of materials, including publications of the Federal government.
2. GPO will coordinate digitization efforts with library and other partners to establish priorities, reduce duplication of effort, and ensure the use of broadly acceptable digitization standards.
3. GPO will employ preservation level standards and best practices to ensure authenticity and integrity of this information.
4. GPO will set up near free partnerships with a variety of sources, including Federal depository libraries, Federal agencies, and private organizations for the purpose of digitizing the legacy collection. "Near free" is defined as the partners assuming the costs for postage and handling of tangible documents and any internal costs occurring within the domain of the partner's site.
5. All converted content for the legacy collection will ultimately be digitized at preservation level specifications.
6. Access level converted content may be included until preservation level copies are created.
7. As the legacy documents are digitized, access copies will be made available in a variety of formats to facilitate search and retrieval, dissemination, or repurposing for print-on-demand and other services.

QUESTIONS FOR DISCUSSION

1. What is an appropriate role for GPO to pursue for digitization in terms of the hands-on, operational digitization we do ourselves, in addition to the coordination role described in the assumptions?
2. What is an appropriate look for preservation of legacy digitized documents (i.e., yellowed paper color made white, depository library stamp removed)?
3. Are there particular format issues unique to Government publications where different methods of capture or representation are appropriate (e.g., length, font sizes, numeric data, maps)?
4. When access derivatives are available, should re-digitization for preservation be a lower priority than for content not available in an on-line format?
5. How does the GPO's initiative fit with other digitization efforts already underway in depository libraries? Other Federal agencies?
6. To what extent does the diversity of our user community influence the ways content should be represented?



U.S. GOVERNMENT PRINTING OFFICE | KEEPING AMERICA INFORMED

Digital Distribution to Depository Libraries:

Exploring the Issues

April 17, 2007

Introduction

- During FY 2006, 93% of all new titles made available through the FDLP were available in electronic form.
- Libraries which exercise local initiative to acquire digital files by pulling them from GPO or from other sources are under no FDLP obligations for access or retention.
- With the development of GPO's digital content system (FDsys), GPO will continue to offer depository libraries and other users the option of downloading digital files for local use.
- GPO is also considering an affirmative distribution of authenticated and official published digital content to Federal depository libraries.
- Digital distribution would involve GPO "pushing" electronic publications to depositories based on their selection criteria.
- Title 44 of the U.S. Code requires selective depository libraries to retain Federal publications disseminated through the FDLP for at least five years and with few exceptions depository materials must be maintained permanently by regional depository libraries.

2005 Biennial Survey:

- 18% of depositories responded that they currently download, store and make online publications available via local servers
- 33% of the respondents indicated that they are willing to receive digital files from GPO to store on local systems
- Most libraries expressed interest in receiving only between 1 and 25 files

October 2006 Depository Library Council:

- many in the depository community were interested in having active digital distribution available as an option
- issue of who owns the electronic files
- issue of what requirements will be placed on libraries receiving the files

Definitions

Pull

Downloading of content on an as-needed basis. Content is made available for users to select and retrieve (“pull”) to local servers or computers. For example, currently users may be said to pull documents from GPO Access. There are no retention or access requirements associated with files pulled from GPO.

Definitions

Push

Intentionally and specifically serving out information to target recipients. Content is automatically sent (“pushed”) from GPO to a list of interested users. The push mechanism is analogous to shipping a box of depository documents, only with electronic content instead of tangible copy. Push technology on the Internet refers to a communications protocol where the request for a given transaction originates with the publisher, or central server.

General Assumptions

1. Electronic files of Federal publications will continue to be available for downloading by Federal depository libraries under FDsys.

General Assumptions

2. Under an FDLP distribution option, distributed digital publications would be sent to depository libraries via a “push” mechanism.

General Assumptions

3. Based on the most recent biennial survey, a high percentage of libraries do not want to receive electronic files at all. Those libraries that do wish to receive them want only a very low volume.

General Assumptions

4. Redundancy is needed to ensure future public access, and FDsys will provide this capability, either by storage at multiple sites, by relying on preservation partners, or both mechanisms.

General Assumptions

5. GPO will distribute digital files optimized for public access. These files will typically be smaller than the archival copies preserved by GPO.

General Assumptions

6. If GPO distributes digital files to libraries under the aegis of the FDLIP, then the various requirements and obligations of Title 44 apply.

Issues Related to Digital Distribution

1. Authenticated files accessed from a GPO server may display differently than the same files located on individual library servers.

Issues Related to Digital Distribution

2. Costs associated with management and storage of digital files are significant and can be a constraint on some libraries.

Issues Related to Digital Distribution

3. Depository libraries may have different system requirements and configurations for the servers where digital distributed files will be located.

Issues Related to Digital Distribution

4. As a practical matter, selective and regional libraries will only be required to retain access derivative files so long as they work in their electronic delivery environment. This is analogous to current requirements for maintenance and care of tangible documents, where due care is to be exercised, but preservation and conservation techniques are not an FDL P requirement for selectives.

Questions for Discussion

1. Partners are concerned that files contained in FDsys may become different from files stored at depository libraries. How can data synchronization best be addressed?

Questions for Discussion

2. Should regionals have to accept and retain all digital files pushed to depository libraries? Are subsets or subject specialties going to be allowed for regional libraries?

Questions for Discussion

3. When GPO makes a new or changed version of a digital file available, what responsibility do the libraries have to inform their users about other versions?

Questions for Discussion

4. Is there a role for libraries wishing to receive the preservation level files to ensure redundancy for permanent public access? Will the distribution of files via FDsys provide the opportunity for permanent public access partnerships through the provision of access derivative files and long-term partnerships for preservation level files?