

# Summary Report



U.S. GOVERNMENT  
PRINTING OFFICE  
KEEPING AMERICA INFORMED

October 1, 2007

## Digitization Demonstration Project

### BACKGROUND

In "A Strategic Vision for the 21st Century", released in December 2004, the U.S. Government Printing Office (GPO) put forth a strategic goal to digitize all Federal publications back to the earliest days of the republic. And further, GPO is to ensure these materials remain in the public domain and available in perpetuity to the American public for no fee.

GPO began planning 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 developed digitization specifications for converted content. The reports of the experts meetings, digitization priorities, and Specifications for Converted Content are available on *GPO Access* at: <http://origin.www.gpoaccess.gov/legacy/>.

In March 2006 a six-month digitization demonstration project was authorized by the Joint Committee on Printing (JCP). The project began July 2006 and provided GPO the opportunity to test equipment capabilities, develop workflow processes, analyze costs, and evaluate methods for ingest, storage, and access to the digitized files.

### ABOUT THE PROJECT

The Project was under the purview of the Customer Services Business Unit, Digital Conversion Services (DCS). DCS used GPO's *Operational Specification for Converted Content* (Version 3.3) to produce preservation master files. *Specification for Quality Control* (Version 1.1) was used to determine the quality of the preservation master files. The access derivatives were created as Adobe Acrobat Portable Document Format (PDF) files.

Library Services and Content management (LSCM) provided DCS with the publications for digitization, based upon previously identified priorities. The Chief Technical Officer (CTO) reaffirmed that the Specifications meet the requirements of the Future Digital System (FDsys). The resulting primary focus of the Project was the continuous improvement and validation of GPO's digitization specifications for converted content. The Project was completed in December 2006.

The digitization demonstration confirmed that GPO should not be used for high volume production of digital content. This conclusion should in no way indicate that GPO does not have a role or should not pursue digitization of the legacy collection. Rather, GPO ought to identify special materials to digitize, e.g., over sized or fragile publications. These would be materials that others are not likely to digitize. Additionally, as the Government Accountability Office (GAO) has urged, GPO should explore and participate in digitization alternatives that eliminate duplicative efforts of Federal agencies.

## VOICE OF CUSTOMER

In January 2007 GPO arranged a **meeting of specialists** that represented US Government agencies, including the Library of Congress (LC) and the National Archives and Records Administration (NARA); Federal and academic depository libraries; and others in the information community. The goal of the session was to review and provide feedback to GPO on the access derivatives of the converted content produced by the DCS during the Project.

The twenty-one attendees were surveyed and there was general consensus that the digitization specifications for preservation level scanning produced high quality acceptable derivatives that support access and search. Survey results indicated a score of 4.16 on a scale of 5, which is an “excellent” overall rating. Excellent (4-5)

is defined as the converted content being visually appealing, achieving end user needs for functionality, well-managed and easily searchable. The rating of “moderately effective” converted content is given to scores of 3–3.9 and is defined as needing improvement in visual appearance, design or presentation in order to achieve better results. The chart shows the group scoring for various publication types included in the Project.

Type of Publication	Average Rating
Planetary Scanned Pubs	4.75
Color Publications	4.0
Public Laws	4.16
United States Code	4.10
Code of Federal Regulations	3.87
Bound Congressional Record	4.0
Federal Register	4.29
Congressional Hearings	4.14
<b>Overall Rating</b>	
Excellent	4.16

There also was general consensus that GPO’s role in the digitization arena is not necessarily to digitize the entire collection of Federal publications. Rather, the group indicated that GPO should play a role in the cooperative environment of Federal publications digitization projects. They also pointed out that GPO should carve a niche for itself by digitizing special materials such as maps, fragile materials, microfiche, and publications with fold-outs.

The **Depository Library Council**, at its April 2007 meeting, recommended that GPO partner with libraries and other institutions on digitization projects. Council further recommended that GPO focus its efforts on standardizing partnership agreements and coordinating the dissemination of specifications for digitization.

## PROPOSED PATH FORWARD

Given the conclusions of the digitization demonstration project and given the similarities in the approach to digitization from the specialists and the Council, GPO proposes the following path to create a comprehensive digitized collection of Federal publications:

- ♦ GPO will set up free or near-free partnerships with a variety of sources including, but not limited to, Federal depository libraries, Federal agencies, and private organizations for the purpose of digitizing the legacy collection;
- ♦ GPO will identify special materials to digitize, e.g., microfiche, over sized, and fragile publications;
- ♦ GPO will coordinate digitization efforts with library and other partners to establish digitization priorities and to reduce duplication of efforts (especially between NARA, LC, and other Federal agencies);
- ♦ GPO will continue to work with the National Digital Standards Advisory Board (NDSAB) of the National Digital Information Infrastructure Preservation Program (NDIIPP) and other standards-creating bodies (e.g., National Information Standards Organization) to develop standards, and to ensure that broadly acceptable standards are used;
- ♦ GPO will use preservation level standards and best practices for digitization and will encourage partners to do the same;

- ♦ GPO will play a leading role in authenticating the digitized legacy collection;
- ♦ All converted content for the legacy collection will ultimately be digitized at preservation level specification;
- ♦ GPO will determine specifications for and manage issues relating to quality control of legacy collection digitization;
- ♦ Access level converted content may be included in the collection until preservation level copies are created;
- ♦ 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; and
- ♦ Priorities for digitization will be revisited and, with the proposed cooperative approach, adjusted as necessary.