

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.

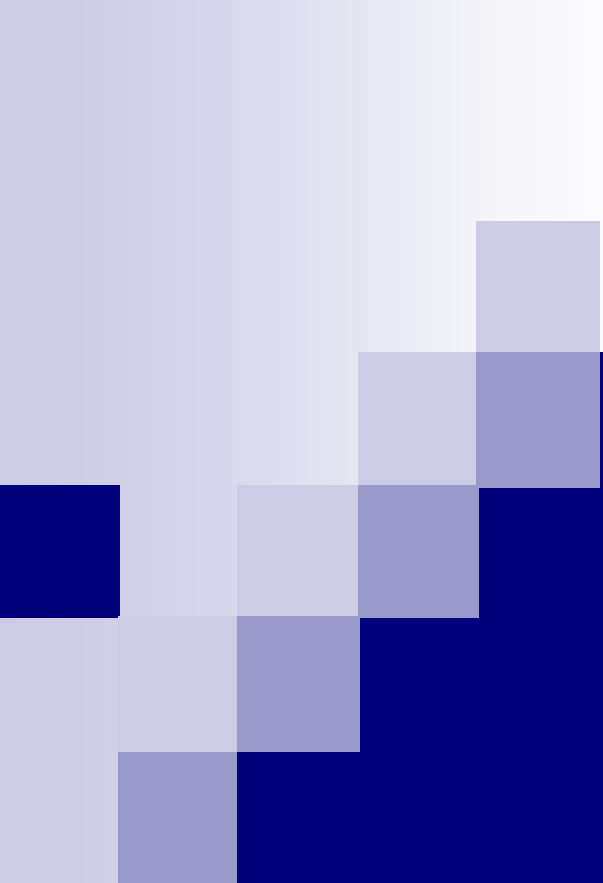
PRELIMINARY ASSUMPTIONS AND QUESTIONS PENDING FINAL DIRECTION FROM JCP

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?



GWLA Federal Technical Reports Digitization Project

Depository Library Council Meeting

Denver, Colorado

April 18, 2007

Tim Byrne, University of Colorado at Boulder

Technical Reports

“A report concerning the results of a scientific investigation or a technical development, text or evaluation, presented in a form suitable of dissemination to the technological community. The technical report is usually more detailed than an article or paper appearing in a journal or presented at a meeting. It will normally contain sufficient data to enable the qualified reader to evaluate the investigative process of the original research or development.”

U.S. Department of Defense, *Glossary of Information Handling*, 1964



Technical Report Production

- First U.S. Government Technical Report Series
 - USGS *Professional Papers* (1902-)
 - NBS *Technological Papers* (1910-)
- Most federal technical reports were not included in the FDLP or the *Monthly Catalog*



Technical Report Production

- Second World War saw a huge increase in research activity
- Federal agencies begin sending technical reports to research libraries around the country



Technical Report Collections

- Most libraries which began receiving technical reports in the late 1940s/early 1950s still have the reports in their collections
- Uncataloged
- Little used
- Taking up a lot of shelf space



Technical Reports Collections

- Space is a major issue for libraries today.
- Low use collections are receiving new scrutiny
- Replacing low use print and microform collections with digital collections is more and more popular.
- Digital collections often get more use than tangible collections.



University of Arizona

- University of Arizona holds extensive technical report collections like many research libraries
- Maliaca Oxnam, Associate Librarian, Science-Engineering Library, began a technical reports digitization list serve



University of Arizona

- Surveys and discussion on list serve indicated that librarians across the country were interested in digitizing technical reports.
- After consulting with Dean of Libraries, Carla Stoffle, it was decided to submit a proposal to GWLA.



Greater Western Library Alliance

- A consortium of 31 research libraries located in the Midwest & Western U. S. with common interests in programs related to scholarly communication, interlibrary loan, shared electronic resources, cooperative collection development, digital libraries, staff development and continuing education.



GWLA Strategic Plan

"Create a program and infrastructure for developing targeted, broad-based digital collections that support instruction and research on member campuses."



GWLA Proposal

- Open Access to Federal Technical Reports
 - “Develop a collaborative project with the Center for Research Libraries to identify, digitize, archive, and provide persistent and unrestricted access to federal technical reports issued prior to 1975.”



Taskforce Members

Maliaca Oxnam, Chair	University of Arizona
Tim Byrne	University of Colorado at Boulder
Mel Desart	University of Washington
Patricia Kirkwood	University of Arkansas
Daureen Neddill	University of Utah
John Saylor	Cornell University (non-GWLA, CRL)
Bob Schwarzwald	University of Hawaii, Manoa
Donna Swischer	Linda Hall Library
Alice Trussell	Kansas State University
Melissa Trevvett	Center for Research Libraries (Liaison)



Charge to the Taskforce

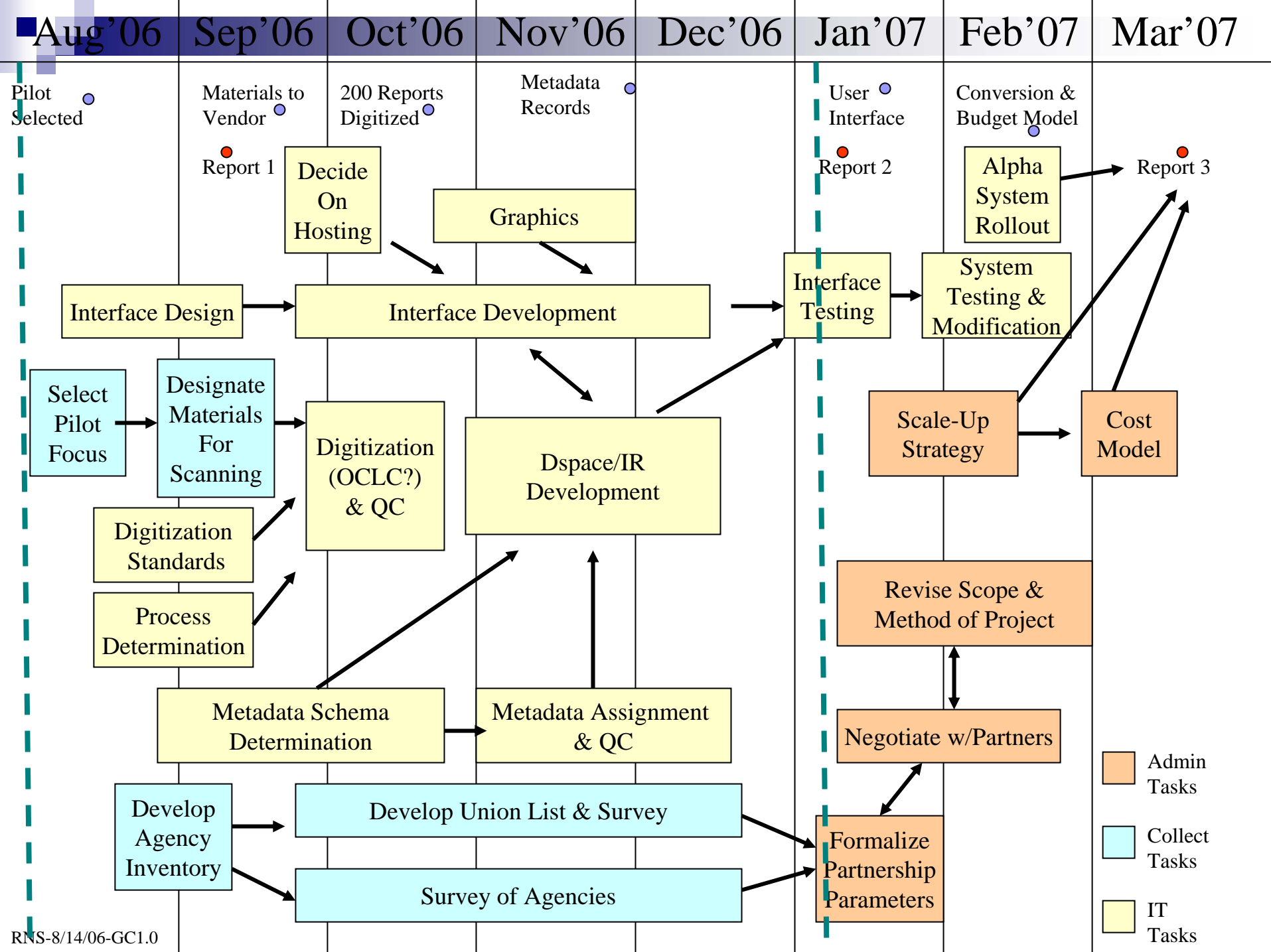
Develop a pilot project in which a sample set of federal technical reports is digitized and made freely available via the web.

Use the data gathered in the pilot to propose a cost model for digitizing and making accessible extensive collections of technical reports.

Final report of the pilot project would be made to the GWLA Directors in March 2007.

Taskforce Objectives

- Prototype digital collection (>200 reports)
 - >200 reports
 - GPO digitization specifications
 - Open access/Persistent access
- Cost-model and implementation plan
 - multi-year digital conversion
 - hosting and sustainability of the resultant digital resource
- Stakeholder approval
 - Project implementation
 - Cost models/projections
- Budget & Timeline = \$65,000 & 8 months



Prototype - Collection

- National Bureau of Standards (NBS)
 - Monograph Series
 - Discrete set of reports
 - Highly ranked in pre-project survey
- Atomic Energy Commission (AEC)
 - Division of Biology & Medicine sponsored reports
 - Selected 4 small sections of reports series
 - Heterogeneity for collection testing

Prototype - Technology

- Hosted at University of Hawaii
 - Open Source Software – Streetprint
 - Has interface and backend management functions
 - Allowed us to develop/test:
 - File sizes
 - Metadata schema (Dublin Core)
 - Workflows – collection assembly and distribution
 - Shortcomings were discovered and identified
 - i.e. File report number sequence (software)
 - Limitations on advanced search options (software)
 - Page counting estimates (workflow)

Proposed Structure

■ Preservation focus

□ 2 E-Archival Institutions

- Store archives of TIFF images, no hosting
- Uses a RAID 5 Storage System with spare to ensure no data loss

□ Print Archives

- Institutions volunteer to be a print archive
- Provided with complete set of documents for preservation
- Metadata (Dublin-core) could be converted to MARC for brief catalog record

Proposed Structure

■ Access Focus

□ 2 PDF Hosting Institutions

- Mirror hosting sites providing open access to reports
- If PDF is corrupted, can request from TIFF archive site

Proposed Structure

- **Coordination**

- **1 Institution serves as the Collection Coordinator**

- **Assembles/prepares collections or works with institutions to prepare collections for vendors**
 - **Coordinates digitization schedule**
 - **Quality Control checks**

Implementation

- Occurs in phases:

- Years 1-3 Development Phase

- Years 4-6 Building Phase

- Years 7-? Continuation of building phase?*

- Project could continue to preserve additional report series & collections as long as there is funding and value and interest to the user community*

- Years X-Z Sustainability Phase

Digitization

Speed & cost of digitization are variables in the cost models

Phase	Digitization Goal (in pages/phase)
Development (Years 1-3)	1 Million
Building (Years 3-6)	1 Million (slow) ~or~ 3 Million (recommended)
Sustaining (Years X-Z)	No digitization

Cost Model Summary

Phase	Digitiz. Goal (pgs)	Total 3 Year Cost	# Partic. Inst.	Suggested Yearly Cost
Development	Minimum (50K)	\$308,500.00	16	\$7,000
			31	\$4,000
	1 Million	\$958,499.20	16	\$22,000
			31	\$12,000
Building	1 Million	\$1,001,999.20	16	\$22,000
			31	\$12,000
	3 Million	\$2,614,000.00	16	\$55,500
			31	\$29,000
Sustaining	None	\$56,000.00	31	\$1,806.45
			97 (CRL) + 11 GWLA	\$470.59



Prototype Demonstration

<http://digicoll.manoa.hawaii.edu/techreports/>



The Future of the Project

- GWLA Directors agreed to have each member library contribute \$4000 to the taskforce for the next year.
- The taskforce will send a representative to a future meeting of CENDI Group of Federal STI Managers
- Will pursue grant funding



Next collection to be digitized

- Atomic Energy Commission

- MDDC Series

- Distributed to libraries in 1948

- Indexed in *Abstracts of Declassified Documents*

- Not included in *Energy Citations*



Tim Byrne


University of Colorado

Government Publications Library

Boulder, CO 80309-0184

tim.byrne@colorado.edu

DARTS: A NTIS/GPO Pilot Project



Cynthia Etkin
Program Planning Specialist
Office of the Superintendent of Documents

Depository Library Council Meeting
Denver, CO
April 18, 2007

Goals of Pilot Project

- ✈ Provide depositories with access to technical and scientific reports in the NTIS collection
- ✈ Compare bibliographic records to identify fugitive publications

What is DARTS?

Depository Access to Reports, Technical & Scientific

- ✈ Technical and scientific reports:
 - ✈ 240,000
 - ✈ 1964 – 2000
 - ✈ Full text online content
 - ✈ PDF files or TIFF wrapped files
 - ✈ Wide variety of subjects covered
 - ✈ New unique content for FDLP

Beta Testing

- ✈ 30 depository library volunteers
- ✈ February 7 - 27, 2007
- ✈ Testing usability, navigation
- ✈ NTIS was very pleased with testing
- ✈ Received good feedback
- ✈ Suggestions are being phased-in

Record 1 out of 7

Title	The NTIS/GPO FDLP Depository Access to Reports, Technical & Scientific (DARTS)
Author	United States. National Technical Information Service .
Publisher Info.	Springfield, Va. : U.S. Dept. of Commerce, National Technical Information Service, [2007-
Internet Access	http://purl.access.gpo.gov/GPO/LPS79112
SuDoc Number	C 51.22/2:D 44
Item Number	0188-A-18 (online)
Variation of Title	DARTS
Description	HTML file
Current Frequency	Updated irregularly, 2007-
General Note	Title from home page (viewed on Mar. 1, 2007).
Limited Use	Restricted to federal depository libraries with valid user IDs and passwords.
Limited Use	
Abstract	Provides full-text access to technical publications from the NTIS data storage and retrieval system.
System Details	Mode of access: Internet from the NTIS web site. Address as of 3/1/07: http://www.ntis.gov/fdip/ ; current access is available via PURL.
Subject - LC	Research, Industrial -- United States. Research -- United States.
Holdings	All items
holdings (2)	Item (no year)
Locate in a Library	(online) http://catalog-web2.gpo.gov/LocateLibraries/locate.jsp?ItemNumber=0188-A-18&SYS=000598142
OCLC Number	(OCoLC)85480046
System Number	000598142

About the Pilot Project

- ✈️ Open to all depository libraries (2/28)
- ✈️ Validation form signed by director required before access granted
- ✈️ www.access.gpo.gov/su_docs/fdlp/col2-dev/ntisform.pdf
- ✈️ 245 registered as of noon Thursday
- ✈️ NTIS cannot lose revenue with pilot

Obligations of Depositories

- ✈ Password and user ID are to be shared only with those who proved reference service, not with patrons
- ✈ DARTS may not be used to supply ILL requests unless requesting library is a depository library
- ✈ Depositories are not to download and distribute reports outside of your library users and the FDLP community

Obligations of Depositories

- ✈ May have up to 30 simultaneous users
- ✈ DARTS may be used for document delivery to the library's primary clientele
- ✈ Depository coordinator is the point of contact for the project
- ✈ Password may be shared with branch librarians

Frequently Asked Questions

- ✈ Is IP authorization available?
- ✈ Can my library use a proxy server for authorization of primary clientele?
- ✈ How are we to ensure there are only 30 simultaneous users?
- ✈ Can I make reports from DARTS available through ILL?
- ✈ How was the coverage of reports determined?

Summer Special

- ✈ Reports added to NTIS collection between 1999 - 2006
- ✈ On CD-ROM
- ✈ \$1.00 per title
- ✈ Can create a subject profile
- ✈ Reports return to regular price 09/01



Questions?

Comments?

Cindy Etkin
cetkin@gpo.gov



U.S. GOVERNMENT PRINTING OFFICE | KEEPING AMERICA INFORMED

Digitization and the FDLP

April 18, 2007

Background

- In 2004, “A Strategic Vision for the 21st Century“ included a plan to digitize all Federal publications back to the earliest days of the republic
- Initially GPO was charged by the Public Printer with digitizing the entire corpus of this material

Background

- Also included the development of a Digital Content System, the Future Digital System (FDsys)
- Efforts would help ensure no fee public access to these materials, in that they remain available, in the public domain, for permanent public access

Initial Steps

- Two meetings of experts:
 - Digital Preservation Masters
 - Preservation Metadata
- Survey of the depository community:
 - determining digitization priorities
- Digitization specifications:
 - converted content
- These reports are available from *GPO Access* at:
<http://origin.www.gpoaccess.gov/legacy/>

Digitization Demonstration Project

July 2006

- Six-month demonstration project began
- Directed by the Joint Committee on Printing (JCP)
- Focus on continuous improvement and validation of GPO's digitization specifications
- Completed in December 2006

Digitization Demonstration Project

January 2007

- Meeting of representatives from Federal government, academic libraries, and others in the information community
- Goal of the session was to review and provide feedback on the converted content produced by Digital Conversion Services
- Consensus validated that the digitization specifications for preservation level scanning were acceptable

Digitization Demonstration Project

Suggestions put forth by the reviewers:

- Include additional elements
- 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
- Technical approaches as related to paper degradation, restoration, grayscale vs. color, lack of policy/planning to deal with these types of issues

Other Efforts

- Coordination and development of partnerships: GPO, Library of Congress, National Archives and Records Administration and Federal depository libraries
- Participation in the National Digital Strategy Advisory Board's Digitization Standards Working Group
- Group Objective: develop common Federal digitization standards for still images that are comprehensive and objectives-based
- Purpose: provide consistency across the Federal government, promote collaborative digitization projects, reduce the possibility of duplicative efforts by agencies, and reduce digitization expenditures
- First tasks: 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
- Awaiting final direction from the JCP on GPO's role for digitization of legacy materials

General Assumptions

1. Digitization is an accepted reformatting method for preservation and access to a range of materials, including publications of the Federal government.

* pending final direction from JCP

General Assumptions

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.

* pending final direction from JCP

General Assumptions

3. GPO will employ preservation level standards and best practices to ensure authenticity and integrity of this information.

* pending final direction from JCP

General Assumptions

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.

* pending final direction from JCP

General Assumptions

5. All converted content for the legacy collection will ultimately be digitized at preservation level specifications.

* pending final direction from JCP

General Assumptions

6. Access level converted content may be included until preservation level copies are created.

* pending final direction from JCP

General Assumptions

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.

* pending final direction from JCP

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?

Questions for Discussions

2. What is an appropriate look for preservation of legacy digitized documents (i.e., yellowed paper color made white, depository library stamp removed)?

Questions for Discussions

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)?

Questions for Discussions

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?

Questions for Discussions

5. How does the GPO's initiative fit with other digitization efforts already underway in depository libraries? Other Federal agencies?

Questions for Discussions

6. To what extent does the diversity of our user community influence the ways content should be represented?