INFORMATION DISSEMINATION IMPLEMENTATION PLAN:

PRIORITIES FOR DIGITIZATION OF LEGACY COLLECTION

U.S. Government Printing Office
Office of Information Dissemination

Washington, D.C.
September 15, 2005

1.2.1.1. PRIORITIES FOR DIGITIZATION OF LEGACY COLLECTION

1.2.1.1.1. BACKGROUND

GPO is working with the library community on a national digitization project with the goal of digitizing all retrospective Federal publications back to the earliest days of the Federal government. The project includes the digitization of the complete legacy collection of tangible U.S. Government publications held in libraries participating in the Federal Depository Library Program (FDLP). The objective is to ensure that the digital collection is available, in the public domain, for no-fee permanent public access through the FDLP. The project will ensure that the collection is digitally reformatted for preservation and access purposes. The conversion of tangible materials will begin with print publications, but will eventually include microfiche and other tangible formats. Information will be digitized based on agreed upon priorities or local needs. The digital preservation masters and the associated metadata will be preserved in the GPO electronic archive (in addition to any other places that the materials might be held) and there will be no-fee public access to the content through derivative files on *GPO Access*. Information specifically about prioritization of material for the digitization of the historical collection is addressed in this paper. Information about other issues related to digitization project, such as *Access to Official Publications* and *Preservation Services* are included in the *Information Dissemination Implementation Plan, 2005-2006*, located at http://www.access.gpo.gov/su_docs/fdlp/pubs/id_plan/toc.html. Digitization efforts are part of the GPO planned Future Digital System. The current Future Digital (FDsys) Concept of Operations can be viewed at: http://www.gpo.gov/projects/fdsys.htm.

1.2.1.1.2. HISTORY

Digitization of the legacy collection will be a very large undertaking with significant staff and financial resource implications. Prioritization of materials to be converted will be essential to effectively manage resources, while serving a variety of other internal and community/user needs. For these reasons, GPO undertook a multi-step approach to determine a priority plan for digitization that included collaboration with other interested parties in the information community. This included a survey of materials to digitize and the incorporation of input from organizations such as the Association of Research Libraries (ARL) and the Government Document Roundtable's (GODORT) Rare and Endangered Government Publications Committee.

GPO Survey to Create a List of Priority Titles

In May 2004, GPO surveyed the depository library community and other libraries to develop a list of priority government document titles or series that should be among the early items to be digitized.

The survey began with a list of candidates for digitization that had already been proposed from two sources. The first source was a 2003 ARL survey of members to identify priority titles for digitization. The second source of titles came from breakout sessions held during the fall 2003 Depository Library Council meeting/conference where participants gathered to discuss and identify priorities.

Survey participants were asked to review the composite list as presented and then add other titles that they felt should be included. The results of the survey were then compiled, consolidated and released for the next part of the survey.

Survey Ranking

Following the analysis and consolidation of the recommendations, GPO asked the community to rank suggested titles and series for digitization. This ranking period for the survey was open from June 23, 2004 through August 6, 2004. Survey participants were advised the resulting ranking was non-binding, and for information only, but survey results would be significant and used to help GPO and individual libraries make decisions about titles for digitization.

The survey process allowed for ranking by depository and non-depository libraries. Each participant was allowed to select and rank up to ten titles for digitization, based on personal experience and perceived local or national need. Real-time ranking results were made available for both depository and non-depository libraries. A total of 310 depository libraries, and 134 non-depository libraries participated in the survey.

Survey Results

The final results of the GPO survey to identify priority titles for digitization from the historical government documents collections located in Federal depository libraries were compiled and the information was made available at: http://www.gpoaccess.gov/legacy/priorities/index.html. The results can be displayed as an overall consolidated list or as customized lists for one ore more library type(s). This made it possible to identify both the overall priorities of the community as well as to identify the titles that are of greatest interest to specific types of libraries, such as public, law, state, academic, and medical. This method of presenting the lists serves to focus attention on high interest titles and provide suggestions for institutions that are planning digitization projects.

Of particular interest is the identification of the Congressional Record, U.S. Congressional Serial Set, Statistical Abstract of the United States, and U.S. Statutes at Large in the first four positions for both depository libraries and non-depository libraries. Of the 25 titles identified for each list, only three titles were not included in the top 25 titles for both lists.

Other Activities and Input Affecting Priorities

In February 2005, four members of the GODORT Rare and Endangered Government Publications Committee provided recommendations to GPO on the twenty most important

titles to be digitized from the results of the "Digitization Ranking Survey" and identification of those titles that are optimal candidates for initial scanning. Three reviewers recommended one title, the Annual Report of the War Department, as a priority.

Finalization of Priority List and Plan

The initial priority list as shown below has been developed based on review of a number of recommendations and resources conveyed to GPO by librarians in the Federal Depository Library Program and greater information community. In developing the priority list, Information Dissemination also worked with other areas of GPO to ensure sufficient material was being identified and obtained for scanning to meet the goal of the Strategic Vision to have 70% of the legacy collection digitized by December 2007. Resources reviewed include *Report of the Meeting of Experts on Digital Preservation: Digital Preservation Masters*, *Report of the Meeting of Experts on Digital Preservation: Metadata Specifications*, the results of the Digitization Ranking Survey, specific recommendations based on GODORT's Rare and Endangered Publications Committee, a review of use statistics for the databases on *GPO Access*, and feedback from the Fall 2004 breakout sessions for different types of libraries on digitization priorities.

1.2.1.1.3. KEY ASSUMPTIONS FOR THIS CHAPTER

- Information Dissemination's (ID) digitization of the legacy collection will be executed in stages.
- This paper presents the first four priority groups of publications that will be digitized.
- The same years of related publications, such as the Federal Register, Code of Federal Regulations and Unified Agenda, should be available for simultaneous use. To enable this, related publications from a specified time period, such as 1990-1995, will be digitized concurrently.
- Upon significant completion of the digitization of the publications in the first three groups, ID will collaborate with the Federal Depository Library community to establish the subsequent order for processing.
- Publications will be digitized and created files will include: TIFF files, text files, and press optimized fully searchable PDF files, resulting from the OCR processing.
    - TIFF image files will be of preservation level quality to ensure future access to digital files and will be retained for preservation purposes only.
    - PDF files will be made available for web access at the screen-optimized level.

1.2.1.1.4. SCOPE OF THE PROGRAM OR PROJECT

- For the first three sets of priorities, GPO will digitize publications dating back to 1990 working up through the present day or the point in time where acceptable content is available through GPO Access. If pre-1990 years are required to have a complete edition, for example the U.S. Code 1988 Edition and its supplements, then the pre-1990 years will be included.

- The first four sets of priorities are:

1. Back files for online versions of the following publications:

   NOTE: The publications have been grouped and listed in a priority order based on review of information from the Digitization Ranking Survey, GPO usage statistics for the different databases, and input from the depository community. Each application is followed by congresses, years or time frame that will need to be digitized.

   o **Code of Federal Regulations:** 1990 - 1996
     - **List of CFR Sections Affected:** 1990 - 1996

   o **Federal Register:** volumes 55 (1990) - 59 (1994)
     - **Unified Agenda:** 1990 - 1994

   o **U.S. Code,** 1988 edition forward**:** Currently, *GPO Access* only has text files for this application from 1994 forward. PDFs will need to be created for all editions and supplements.

   o **Congressional Record:**
     - **Congressional Record, daily edition**: 101st Congress, volume 135 (1989) through 103rd Congress, volume 140 (1994)
     - **Bound Congressional Record:** 101st Congress, volume 139 (1989) - 105th Congress, Volume 144 (1998)
     - **Congressional Record Index:** All years, since GPO only has text files from 1983 - present, with the exception of HTML files for the year 1998.

   o **Monthly Catalog of United States Government Publications:** Working from June 1976 back through the earliest volumes in order to facilitate access to the largely un-cataloged pre-1976 material. July 1976 – 2004 will be digitized at a later date.

   o **Congressional Bills:** 101st Congress - 102nd Congress

   o **Public and Private Laws:** 101st Congress - 103rd Congress

   o **Statutes at Large:** 101st Congress - 107th Congress (currently working on making the 108th available through the typesetting process).

   o **U.S. Reports:** 1990 - present (currently working on this database).

2. Digitized files for the remaining applications on *GPO Access*. Such applications include but are not limited to *GAO Reports*, Congressional Reports, Documents and Hearings and the *Budget of the United States*.

3. Electronic files for those remaining *Essential Titles for Public Use in Paper Format*, http://www.access.gpo.gov/su_docs/fdlp/pubs/estitles.html, that are not addressed in the priority groups listed above. Such applications include but are not limited to *Census of Population and Housing*, *Social Security Bulletin*, and *Condition of Education*.

4. While the above 3 priority groups are being undertaken, GPO will proceed on a more limited basis, with simultaneous work on rare and endangered documents. Representatives of GODORT's Rare and Endangered Publications Committee have provided GPO with a priority list of titles to be considered for digitization. The first title to be digitized from this group will the Department of War Annual Reports (1789-1947), working from the most recent back through the earlier volumes.

- After the completion of retroactive digitization back to 1990, GPO will reassess digitization priorities. Regardless of the outcome of that assessment, digitization of publications included in the first three sets of priorities will continue in 10 year increments until all preceding publications have been digitized back to publication inception.

1.2.1.1.5. PLANNING ISSUES FOR INVESTIGATION

- Priorities may conflict with reality of available documents and the need to maintain sufficient material in house to keep workflow moving.
- Disclaimer statements should be developed to reflect the situation when documents are not available and therefore sets may be incomplete.

1.2.1.1.6. RECOMMENDED ACTIONS
- Share list of priorities as identified above with the information community to obtain comments and suggestions.
- Reassess priority plan after rollout of live digitization process is put into place in September 2005 and schedule work for the upcoming fiscal year.