

# Digitization of a Legacy Collection: the Collaborative Process When You Don't Have the Collection In-House

Robin Haun-Mohamed, Virginia Wiese; U.S. Government Printing Office; Washington, DC, USA

## Abstract

*The United States Government Printing Office (GPO) is like so many other organizations facing the challenges and opportunities required for the transition from an ink on paper to a digital society. We are addressing technology issues and are implementing plans for the creation, management, and preservation of digital collections. And yet GPO has an additional challenge in that we do not have a tangible collection at hand from which to digitize publications. It is this, our unique work experience with selecting content and creating a collection that we would like to share with other conference attendees, participants who may be facing similar circumstances in situations where they are responsible for digitizing what they do not have in hand.*

## GPO and Federal Depository Library Program Challenge

The American public's primary source of free access to authentic Federal Government information is the Federal Depository Library Program (FDLP). Administered by the GPO, the FDLP provides for the distribution of U.S. Government publications to a network of more than 1,250 public, academic, law, federal, and other libraries in virtually every congressional district, where they can be used by the public free of charge.

As this program is affected by the World Wide Web and increasingly advancing technologies, GPO understands the criticality of providing the official government documents through the online environment. In addition our partners and the public are beginning to expect it. GPO needs to provide online access to digital files of the legacy collection of materials held by our partners in Federal Depository Libraries.

In many ways GPO is like other organizations facing the challenges and opportunities required for the transition from an ink on paper to a digital society. We are addressing technology issues and are implementing plans for the creation, management, and preservation of digital collections. And yet GPO has an additional challenge. GPO does not have a tangible collection at hand from which to digitize publications. We are a Federal Library Program without a library collection! GPO relies on our partners to make historical Federal tangible documents available to the public. And so in a similar manner GPO is now relying on and working with the library community and others to make historical Federal tangible documents available to the public though the online environment.

## Project Goal

The goal of the project is to digitize the complete legacy collection of tangible U.S. Government publications and ensure that the digital collection is available, in the public domain, for no-fee permanent public access through the FDLP. This includes

ensuring that the collection is digitally reformatted for preservation and access purposes and that digital preservation masters and associated metadata will be preserved in the GPO electronic archive or other places. Publications to be digitized will include print, map, microfiche and other tangible formats.

## Project Plan

Successful completion of such a large scale digitization project involves many phases. GPO started by talking with experts, collaborating with our partner libraries, setting priorities, and conducting a small Digital Demonstration Project. Our next steps include an evaluation of the Demonstration Project, keeping abreast of what others are doing in the digital conversion environment, developing partnerships, and where necessary, learning from our experience, and modifying or tweaking our approach to the initial plan. We know that we must maximize our efforts, time, and money. We will continue to collaborate with others in the information community who are proceeding with their own digitization efforts.

## Cooperation to Succeed

The focus of this presentation is twofold. The first portrays our efforts to prioritize digitization for a large national government collection. The second focus of this presentation relays the methods employed to gather and obtain an actual physical collection of the legacy U.S. Government publications for digitization. We will tell about the actions we have already taken in these areas, with the acknowledgement that this is an ongoing process that will be monitored and polished as the project proceeds. Both of the efforts that we are presenting were undertaken in conjunction with several primary critical stakeholders. Our network consists of Federal agencies, and several types of libraries including Federal, Federal depository, and other. These are partners with whom the GPO has established unique critical historical relationships. Coordinating a broad range of digitization transition activities with these parties is critical to program success and execution. And as we will demonstrate, our success to date and in the future could not occur without their support and active involvement.

## Priorities for Digitization of Legacy Collection

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 where participants gathered to discuss and identify priorities. These participants were asked to review the composite list as presented and then add other titles that they felt should be included. The results of these efforts were then compiled, consolidated and released for the next part of the survey process.

### ***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 for approximately 6 weeks in 2004. Survey participants were advised that 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. The results were available as an overall consolidated list or as customized lists for one or 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 court. This method of presenting the lists served to focus attention on high interest titles and provide suggestions for institutions that are planning digitization projects. This information is still available at: <http://www.gpoaccess.gov/legacy/priorities/index.html> and may be helpful to you.

### ***Other Activities and Input Affecting Priorities***

In February 2005, members of the GODORT Rare and Endangered Government Publications Committee provided recommendations to GPO on the twenty most important fragile titles to be digitized from the results of the "Digitization Ranking Survey" and identification of those titles that are optimal candidates for initial scanning.

### ***Finalization of Priority List and Plan***

The initial priority list as shown below has been developed based on the 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, other areas of GPO and other resources were also consulted to ensure sufficient material was being identified and obtained for scanning. Resources reviewed included:

- the Report of the Meeting of Experts on Digital Preservation: Digital Preservation Masters,
- the 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
- feed back from the Fall 2004 Federal Depository Library Conference breakout sessions identifying different types of libraries' digitization priorities.

The top priorities would serve as the publications to be focused on for the Digitization Demonstration Project.

It is interesting to briefly note what our national constituency identified as the top needs.

- Back files for online versions of the legislative and congressional publications such as the: Code of Federal Regulations, Federal Register, U.S. Code, Congressional Record, Monthly Catalog of U.S. Government Publications, Congressional Bills, Public and Private Laws, Statutes at Large, and U.S. Reports.
- Digitized files for the remaining applications on GPO Access such as GAO Reports, Congressional Reports, Documents and Hearings and the Budget of the United States.
- Electronic files for those remaining Essential Titles for Public Use in Paper Format, ([http://www.access.gpo.gov/su\\_docs/fdlp/pubs/estitles.html](http://www.access.gpo.gov/su_docs/fdlp/pubs/estitles.html)), that are not addressed in the priority groups listed above.
- Simultaneous work on rare and endangered documents such as the Department of War Annual Reports (1789-1947).

The actual Digitization Ranking Survey results can be found at: <http://www.gpoaccess.gov/legacy/priorities/index.html>. And the final report for the priority plan is at: [http://www.gpoaccess.gov/about/reports/prioritydigitization\\_plan.pdf](http://www.gpoaccess.gov/about/reports/prioritydigitization_plan.pdf). Upon completion the final list of priorities was shared with the information community to obtain comments and suggestions and it was determined that there would be periodic reassessments of the priority plan to assure that community needs were being met.

### ***Acquiring publications***

At the beginning of this project GPO's integrated library system online catalog was not yet available and the acquisitions module had not yet been implemented. This required us to rely on an in-house home-grown bibliographic and acquisition system. We had an idea of the ranges of the publications we were in need

of, but we had no idea of the specific documents that constituted the collection, and what would be needed to fill in the gaps. In order to identify all individual pieces published with each publication we conducted a through review of the older Monthly Catalog of U.S. Government Publications, the historical shelf list and card catalog, that was sitting in a remote reference area, and other available reference resources to determine what pieces would need to be digitized.

Then, in order to digitize publications internally for the Demonstration Project, GPO announced to the library community that it was acquiring publications for digitization. GPO issued a call for material through the GPO FDLP Desktop website <http://www.gpoaccess.gov/legacy/index.html>. The list of needed publications was updated weekly and posted throughout the demonstration project or until publications were acquired. GPO preferred to receive materials from regional-approved de-accessions or collections of duplicates that were no longer needed. Donations were preferred, and loans of material were considered when publications were not available for donation. While GPO focused its acquisition of publications on the Priorities for Digitization of the Legacy Collection, it also acquired numerous other publications, especially long series or groups of publications, when it was deemed they would be difficult to locate or costly to obtain at a later date.

GPO personnel ordered the publications from offering libraries and arranged and paid for shipment and delivery to the main GPO facility. For the Digital Demonstration Project alone we obtained over 3,800 cartons of publications, weighing over 725,600 lbs. Once acquired, each publication needed to be inventoried, have a brief bib record created, processed, tracked and moved to warehouse storage locations prior to submission for digitization.

## Registry

To support regular review of the priorities paper and to identify potential digitization partnerships with institutions, GPO established an online resource entitled the Registry of U.S. Government Publication Digitization Projects. This is currently available on GPO Access at: <http://www.gpoaccess.gov/legacy/registry/>. The resource was developed based on best practices established by the Association of Research Libraries and other organizations for registries. The purpose of the Registry is to minimize duplication of effort through identification of planned or publicly accessible collections of digitized U.S. Government publications. Contributors are requested to identify how the publications have been digitized indicating preservation or access level. The Registry will be marketed regularly with the goal that it will provide a comprehensive coverage of all in scope digitization projects.

## Project Lessons Learned

Our experiences have highlighted the challenges apparent in such a project and these opportunities for growth will continue as we work with our partners to move forward, stay abreast in changes in the digitization and preservation communities, and identify ways to improve the efficiency and effectiveness of this important project. For example, GPO will continue to work with others to further review and update standards associated with digitization. We are currently working with the Library of Congress and National Archives and Records Administration and other agencies, to develop image and metadata standards as part of the Federal Digitization Standards Working Group. And on another front GPO continues development of the Future Digital System (FDsys), a life-cycle management system that is being designed to accept, organize, manage, and authenticate official Federal information content for user access and long term preservation. Digitization projects such as the legacy collection will need to be in sync with the FDsys ingest requirements.

## In summary

**The plan to digitize the legacy collection of federal documents will not proceed without the help and assistance of the Depository library and other library communities. Without their participation in the overall goals and priorities for this project we will have neither buy-in nor publications from which to digitize. By soliciting input and determining the community needs we are able to cooperate, work together and begin to deliver to the public this vast range of Government information to be made available online. Cooperation is necessary to succeed. As has been apparent throughout this project, working in a cooperative and collaborative environment is the most effective approach to minimizing duplication of effort, digitize the largest amount of material, and maximize the use of available funds to ensure future access to these important resources.**

**For more information on the GPO's Digitization and Preservation Initiatives you can go to: <http://www.gpoaccess.gov/legacy/index.html#priorities>. To gain access to a broad range of Federal publications that are already available in electronic format go to the GPO Access Web site at: [www.gpo.gov/gpoaccess](http://www.gpo.gov/gpoaccess). Additional support can be obtained through the FDLP libraries which are fully equipped and staffed to assist the public in locating and using online Federal Government information.**

## Author Biography

*Robin Haun-Mohamed received her MLS, and BA in political science and history, from the University of Washington, Seattle. Since then her work has emphasized public access and service for libraries, hospitals, and retail industries. Currently she is Director, Collection Management and Preservation at GPO. Her responsibilities include coordination, implementation, and evaluation of: joint efforts between depository libraries, GPO and other Federal agencies, and projects and processes impacting long term public access to Federal information products and services.*