

# Preservation of Federal Government Publications in Multiple Formats Proposal

The GODORT Preservation Working Group urges the Government Documents Round Table (GODORT) to promote a national conversation about the value of preserving **historic Government publications** in multiple formats in order to serve a diverse public and to publicize the need for Government publications librarians to help the public access those publications. GODORT should urge ALA to ask the US Congress to appropriate funds for preservation of Federal Depository Library Program government publications. This money should be used for direct support of depository libraries who want to preserve their paper and digital government publications.

The Preservation Working Group recommends the following:

1. ALA should urge Congress to support the Superintendent of Documents at the Government Printing Office (GPO) **Federal Information Preservation Network (FIPNet)**. FIPNet is leading the effort to develop **The National Preservation Plan**—a collection of guidelines, strategies, partnerships and best practices for the preservation of both legacy/tangible and digital government publications. FIPNet was developed as a response to the comprehensive results of the GPO's preservation survey of FDLP members. This survey asked specific questions about plans for digitization of tangible collections (at both the local and state/regional levels), hosting of all-digital collections, and other important government document preservation/access concepts/concerns.

Some of these questions include the following:

Library Forecast Questionnaire:

- **Question 13:** If your library digitizes FDLP material (in-house or outsourced), where do you store the master digital files? Please check all that apply.
- **Question 14:** Does your library plan, within the next five years, to digitize publications from the FDLP/government documents collection?
- **Question 15:** Would it be useful for GPO to provide advice and guidance for libraries that want to plan projects to digitize publications from the tangible collection?
- **Question 16:** As government information is increasingly produced and distributed in digital-only formats, what barriers to access, if any, do you anticipate in the next five years?

State Forecast Questionnaire:

- **Question 2:** If FDLP libraries within your state digitize FDLP materials (in-house or outsourced), where do they store the master digital files? Please mark all that apply.
- **Question 3:** Do FDLP libraries in your state plan to digitize publications from the FDLP/Government documents collection within the next five years?
- **Question 4:** Would it be useful for the GPO to provide advice and guidance for libraries that want to digitize publications from the tangible collection?
- **Question 5:** As government information is increasingly produced and

distributed in digital-only formats, what barriers to access, if any, do libraries in your state anticipate in the next five years?\*

These and other findings were detailed in "Preservation: An FDLP Forecast Study Working Paper." FIPNet includes as partners in the network depository libraries, the Library of Congress, other national libraries, the National Archives and Records Administration, and other bodies interested in preservation of government information.

2. The National Preservation Plan should include government publications/resource assessment criteria for a participating depository library to use to designate a particular government publication/resource as a worthy candidate for preservation. (Please see appendix A for proposed details.)
3. The National Preservation Plan should include an inventory of historic government publications held by all the depository libraries and an analysis of the physical condition of those publications. All government publications available through the FDLP should have a cataloging record in the GPO national catalog a.k.a. the Catalog of U.S. Government Publications (CGP). (Please see appendix B for proposed details.)
4. The National Preservation Plan should include a strategy for depository

\* Government Publishing Office, "Preservation: An FDLP Forecast Study Working Paper," 2013, [www.fdlp.gov/file-repository/about-the-fdlp/gpo-projects/fdlp-state-forecast/2370-preservation-an-fdlp-forecast-study-working-paper/file](http://www.fdlp.gov/file-repository/about-the-fdlp/gpo-projects/fdlp-state-forecast/2370-preservation-an-fdlp-forecast-study-working-paper/file).

libraries to cooperate in the collecting, housing, and cataloging of Government publications. Individual FDLP member libraries and FDLP regionals that, according to the results of the survey profiled in “Preservation: An FDLP Forecast Study Working Paper,” indicated an interest in playing leadership roles in preserving legacy documents and/or digitizing/housing digital collections should be contacted for further discussions. The plan should allow for depository libraries to commit to the collection, cataloging, and preservation of subsets of government publications with the approval of the regional library in each state and the superintendent of documents. There should be appropriate cross-references to the publication’s digital equivalent. (Please see appendix C for proposed details.)

## Appendix A

### *Candidate Designation Criteria for National Preservation Plan*

According to *Rebooting the Government Printing Office: Keeping America Informed in the Digital Age*, the following conditions still exist in 2016:

#### ***Preservation of the Legacy (Tangible) Government Collection (Finding III-3):***

***No comprehensive plan or program exists for preserving the legacy collection of government documents. While preservation of the legacy collection is not a GPO responsibility, this issue should be addressed as the FDLP becomes an increasingly digital program.***

Regional depository libraries are responsible for maintaining the tangible documents they receive through the FDLP. It is estimated that there are approximately 2.3 million items in the FDLP, but about

one-third of the collection has never been cataloged. In addition, individual library collections vary due to a number of factors, including when they entered the program, loss or destruction of printed documents, acquisitions of government documents that were not distributed as part of the FDLP, and so forth. As a result, no definition of a full government collection or the location of specific items currently exists.

Many depository libraries, faced with space constraints, are turning to digitization as one method of preserving the print collection. One goal of digitization is to provide flexibility for depository libraries to dispose of print copies of documents that have been digitized. Regional depository libraries may not substitute a digital surrogate for a tangible FDLP title, while selective libraries may substitute under certain conditions. However, many depository libraries have obtained government documents that were not distributed through the FDLP, and these items are not subject to the same rules as FDLP titles.

Digitization contributes to preservation by providing online access while reducing handling of the print counterpart. However, digitization is not in itself a comprehensive preservation plan for the print collection because digital content is less stable and has a shorter lifespan than print, and there is not yet a consensus on its long-term preservation. In fact, the LC currently recognizes only print and microfilm as preservation standards. A comprehensive plan for preservation of the print collection will require supplementing digital documents with a yet-to-be-determined number of full print collections, in controlled environments and in geographically dispersed locations. There is a danger of permanent loss of information if a significant number of print documents are disposed of before a comprehensive preservation plan is developed.

How digitization is carried out and the digitized products are made accessible deserve careful planning. Digitization is more complicated and costly than simply scanning documents. The digitized content needs to be searchable, discoverable, and authenticated, and there are quality control issues.

There are several digitization efforts that can be built upon and coordinated, including depository and other library networks, LC, and executive branch agencies. In addition, a new OMB/NARA directive instructs executive branch agencies to consider digitizing their collections.

The ingestion of digitized collections into FDsys improves preservation and accessibility. FDsys has this capability and collections digitized by LC and executive branch agencies have been ingested by the system. GPO currently does not allow ingestion of documents digitized by depository libraries into FDsys due to strict standards regarding authentication. Instead, GPO publicizes and supports collaborative digitization projects and digitized collections through its online Registry of U.S. Government Publication Digitization Projects.<sup>†</sup>

The Digitization Projects Registry can be found at <http://registry.fdlp.gov>.

While self-directed efforts of both individual and regional libraries continue (in particular, the Lots of Copies Keep Stuff Safe (LOCKSS), the need for FIPNet to work formally with these institutions to develop the National Preservation Plan becomes more critical as time progresses.

<sup>†</sup> National Academy of Public Administration, *Rebooting the Government Printing Office: Keeping America Informed in the Digital Age*. January 2013, [www.gpo.gov/pdfs/about/GPO\\_NAPA\\_Report\\_FINAL.pdf](http://www.gpo.gov/pdfs/about/GPO_NAPA_Report_FINAL.pdf), 32.

Since the scope of the National Preservation Plan is to preserve government publications, the inclusion of references to the historical value of these items should be emphasized.

Therefore a thoughtful, comprehensive National Preservation Plan Candidate Designation Criteria will need to include the following elements:

- identification of the types of Government publications considered to be historical in nature
- National Preservation Plan Candidate Designation Criteria, which may include the following categories and conditions:
  1. rarity as a depository item
  2. age
  3. historical value (primary source document; important policy; part of larger collection)
  4. ease of use as an historic document
  5. lack of digital equivalent
  6. use-potential rating (possible scale: 1–5: 1 = rarely used, 5 = heavily used)
  7. ease of physical preservation (evidence of stable material; binding, if applicable [possible scale: 1–5: 1 = difficult to preserve, 5 = easy to preserve])

The example of National Preservation Plan Candidate Designation Criteria items listed above are to be used at this point as exhibits only. The GPO and the individual/state members of the Federal Information Preservation Network will develop and approve the criteria, as well as develop an action plan (which may include legacy government publication inventory and collection maintenance details outlined in appendixes B and C) to be included in the National Preservation Plan.

## Appendix B *National Preservation Plan Inventory of Government Publications*

FDLP member institutions participating in the National Preservation Plan (either directly upon receiving funding as part of the National Preservation Plan or indirectly as a partner with another institution or the GPO) would be required to conduct an inventory of government publications in their collection which have been deemed eligible for preservation.

All records created during the transcription of the historic Shelflist are available from the Catalog of U.S. Government Publications (CGP). As of May 23, 2016, more than 170,000 shelflist records are available through the CGP.

Information on the GPO's transcription of the Shelflist and other efforts to catalog the legacy collection is available on the National Bibliographic Records Inventory Initiative (NBRII) page on the FDLP website ([www.fdlp.gov/project-list/national-bibliographic-records-inventory-initiative-nbrii](http://www.fdlp.gov/project-list/national-bibliographic-records-inventory-initiative-nbrii)).

### Cataloging and Indexing Program (Finding III-6)

GPO cataloging and indexing insures federal government information is discoverable. Significant cataloging and indexing of government documents are needed for ease of access and inventory management. In 1996, the GPO estimated that approximately 50 percent of government documents were not cataloged, indexed, or distributed to depository libraries. With the vast majority of government documents now born digital and posted on agency websites, the

current percentage of government publications that are fugitive is unknown, but can be assumed to be higher than the GPO's 1996 estimate. Unfortunately, posting information on a website does not mean citizens can find it. Given the federal government's enormous web presence and the tendency for URLs to change, finding government documents on agency websites can be very challenging, even for web-savvy users. Cataloging and indexing makes government publications discoverable. Cataloging the legacy collection is also the first step in preserving that collection; there is a need to define the collection to identify what needs to be preserved. Cataloging the full collection will need to be a collaborative effort because library collections vary depending on when they entered the program and other factors. The GPO's goal is to expand the online Catalog of Government Publications to make it more comprehensive, including historical and electronic documents. Activities to expand the catalog include increased harvesting of born-digital federal documents and expanding cataloging record services to depository libraries.

## Appendix C *Collection, Housing, and Cataloging of Government Publications*

The need to have a comprehensive plan to collect, house, and catalog government publications is essential to the success of the National Preservation Plan. Therefore it is important to identify the details of this portion of the National Preservation Plan.

First, the key stakeholders associated with the ongoing maintenance of government publications should include the following:

- designated FDLP libraries with large collections
- designated FDLP special libraries with historic documents collections
- GPO and government information centers

Second, the plan should identify government publication housing criteria including the following:

- secure locations with favorable general overall climate conditions
- provisions for items requiring special handling because of age and/or uniqueness (as identified via the inventory component of the plan as well as criteria determined by the National Preservation Plan Advisory Committee)
- provisions for items deemed to have significant financial value (using criteria determined by the National Preservation Plan Advisory Committee)

Third, the plan should include a cataloging production workflow to create and maintain catalog records and metadata associated with government publications acquired, processed and housed because of the National Preservation Plan. This portion of the plan should include the following:

- roles and responsibilities
- cataloging record examples
- estimated costs (short- and long-term—possible scenario below)

### Estimated Costs for Provision of Records Related to the National Preservation Plan

Marcive's expertise is in the selection and manipulation of sets of MARC records from an existing larger set of records. Marcive holds GPO cataloging

dating back to the 1970s as well as GPO's Historic Shelf List records; both of these files are updated monthly. Marcive would anticipate that libraries engaging in National Preservation Plan projects would be requesting records for particular agencies and time periods from either of these sources. Costs for a backfile from these files would typically include a profiling/setup fee (\$80) and a GPO records cost (\$0.07/record, \$2,000 minimum).

Specific details of a project may incur other costs as dictated by project scope, number of volumes, etc.

**Example 1:** A library participating in the ASERL (Association of Southeastern Research Libraries) project trying to ensure comprehensive coverage of records in their selected agency asked Marcive to provide all of the GPO records we had at the time for the agency. Approximately 20,000 records were provided at a cost of \$2,080.

**Example 2:** A library with current GPO cataloging wished to acquire records for materials acquired before their cataloging subscription had begun. The librarian in charge edited a list of SuDoc stems provided by Marcive to include the range of publication dates found on her shelves. Marcive staff then extracted GPO records matching the SuDoc stems on the list that for titles falling within the desired date range and prepared the records according to the already-established requirements for the library's catalog, including barcode labels for the print monographic titles. Approximately 31,400 records and 23,000 barcode labels were provided at a cost of approximately \$3,200.

## Appendix D

### *National Preservation Plan Model Use Case: Dartmouth Library US Congressional Serial Set Digitization Project*

For a government information library to successfully identify, plan, and implement a legacy government document preservation project (following the candidate designation guidelines associated with The National Preservation Plan and other details), having an example of a successful project (one that was well planned, preserved the integrity of the legacy versions while creating access to digitized iterations, had adequate funding, received assistance or sponsorship from supporting institutions or private-sector benefactors, etc.) is often helpful to for the government information library planning the project to use for project validation, strategic, and other purposes.

One such use case that is of particular interest is the Dartmouth Library US Congressional Serial Set Digitization Project.<sup>‡</sup>

Not only did the project provide access to a completely digitized version of the US Congressional Serial Set, but it also provided additional benefits associated with the legacy print documents, including

- conservation of existing volumes (repair of existing damage as well as any damage incurred during digitization);
- increased use of legacy volumes; and

---

<sup>‡</sup>R. Langdell, "US Congressional Serial Set—Readex Digitization Project," Dartmouth College Library, Preservation Services, April 2009, <https://www.dartmouth.edu/~library/preservation/ssreadex>.

- enhanced findability of content via increased indexing of digitized version (which benefitted legacy users as well)

Barbara Sagraves, Dartmouth Library's head of preservation, who led the project (which partnered with the Readex Corporation) shares the following overview of their successful National Preservation Plan Model use case:

### Case Study

The US Congressional Serial Set Digitization Project: a collaboration of Readex Corporation, a division of NewsBank, Inc., and Dartmouth College Library, 2003–13

In 2003, librarians at Dartmouth College Library in Hanover, New Hampshire, were contacted by staff at Readex Corporation in Chester, Vermont, for the loan of selected volumes of the US Congressional Serial Set containing color illustrations for a digitized version they were producing. The initial agreement was to borrow fifteen items per month; in exchange, Dartmouth would receive a discount of certain Readex products and a credit that could be applied to purchase for each colored illustration that was used. The digital product would be black-and-white scans of the text with maps and illustrations in full color.

The Library agreed to the offer and Preservation Services was responsible for its implementation. Existing staff retrieved the requested volumes, verified the needed illustration existed, inspected the physical condition, and packed the volumes for pickup by Readex staff. When the books returned they were inspected and conservation repair was performed if needed. There was no compensation for this work beyond the aforementioned product discounts. The item requests were low in number and a single

staff member who normally performed serials binding preparation absorbed the work.

Soon after the project commenced, Readex began to inquire if it would be possible to have access to the entire collection, an estimated 13,000 volumes, for digitization over a period of four years. This quantum jump would require retrieving and processing more than sixty volumes a week, and Preservation Services would be unable to absorb the workload. Both parties were interested in building on what was thus far a successful relationship, so a variety of scenarios were discussed. The core values were access, service, preservation, and communication. Readex wanted access to the volumes at a rate that would support their production schedule and Dartmouth wanted access within twenty-four hours to volumes that were at Readex. Service was key both in Dartmouth meeting weekly production benchmarks and supporting Readex by locating materials that were requested outside the schedule sequence. Preservation was of utmost importance and Readex staff were sensitive to treating the materials carefully. Communication was the stuff that finally greased the wheels. Each institution had staff assigned to the project and they met regularly for project updates and troubleshooting; the two teams met at least once a year at Readex, and the project managers of the two organizations kept in contact by phone, email, and face-to-face meetings.

A variety of scenarios were discussed during negotiations, including Readex staff working on-site at Dartmouth retrieving and scanning the volumes. This plan was abandoned for technical reasons: the amount of data that would be generated

during digitization could not be easily managed working offsite. The idea of dedicated staff persisted, so Dartmouth proposed that two conservation technicians be hired to work in Preservation Services with salary reimbursement provided by Readex. This number was arrived at by doing a sample time study to determine how much time it would take to retrieve and process the materials. Having 2.0 FTE dedicated to the project would ensure that benchmarks could be met and Readex would never have to wait on materials. Reimbursement was also provided for other project members,<sup>§</sup> but it was eventually eliminated during reevaluation of the agreement. Product discounts were also negotiated as part of the agreement.

By 2005 an agreement was finalized to digitize 13,800 volumes of the US Congressional Serial Set from 1789 to 1980. The project was to take four years and each year the principal partners would meet to discuss efficiencies and ways to improve the process. These meetings happened more often than that but built into the contract was the notion that the principal decision makers would meet face to face to build the relationship.

Two conservation technicians were hired to work in Preservation Services and were responsible for the day-to-day project

---

<sup>§</sup>The initial agreement provided for reimbursement for the time of the head of Preservation Services to manage the project, the government documents librarian to serve as content specialist and assist in the locating the materials, the Collections conservator who would train the conservation technicians as well as the machine operators at Readex, and 2.0 FTE conservation technicians.



tasks. They pulled in (almost) chronological-order volumes from the stacks and prepped them for shipment including vacuuming to remove dust, attaching a barcode to each volume and creating an item record in the library catalog, charging them out in the circulation system to pseudo-patrons to keep track of the volume, assessing them for treatment prior to digitization, packing them for shipment, and creating a packing list.

Once at Readex, the volumes were kept in a secure, climate-controlled room until needed for scanning. A Kirtax machine was used for automated digitization, with a technician monitoring the image capture. Numerous quality assurance steps were used to verify that all pages and images were captured at the same high quality. Once imaged the data was reviewed and the editorial unit at Readex indexed the content to project specifications.

The usual turnaround time for shipments was about four to six weeks. Once returned books were evaluated for conservation treatment and repairs were performed. Experience taught staff to keep the books in the department (and checked out to the project) for several weeks because quality control issues at Readex might require the return of a recently scanned book. It was simpler to keep the book in Preservation Services where it was easily retrievable if needed by Readex or a patron. Once staff were certain the book would not be needed they were discharged and returned to the stacks.

Classification practices at Dartmouth resulted in the Serial Set being shelved in varying locations throughout the library system. Conservation technicians, often

with the advice of the government documents librarian, would have to hunt the missing volumes down and be creative in problem solving. By the end of the project staff throughout the library system, and all locations provided support for locating and sending the volumes for preparation.

The Serial Set at Dartmouth displayed some of the same physical deterioration found at other libraries. Red rot of the sheepskin bindings, detached covers, broken spines. A great deal of time was spent repairing maps that are folded and bound into the books. Occasionally technicians would remove a map and place it in its own box. This was done because the folded map was of such great thickness that it damaged the spine of the book that held it and was damaged by unfolding. Other problems found were books that were not sufficiently cut open and might be damaged during digitization. In those cases conservation staff prepped the items to allow better imaging along the gutter.

The primary project work was done by Preservation Services staff; however, Cataloging and Metadata Services provided cataloging support, cleaned up the records for separately cataloged titles, and added serial statements and numbering to reflect each volume's connection to the serial set. This work was not underwritten by Readex but was essential. Through the life of the project the physical item and its bibliographic and item records were reviewed and fixed as needed.

Our agreement with Readex was for four years and as we neared the end of that period it was decided to enlarge the scope to include up to 1995 of the Serial Set. Thus it was extended and ended in 2013.

At project completion, 15,739 volumes had been bar coded, digitized, and conserved. Titles included the American State Papers, the US Congressional Serial Set, 1789–1995, Senate Executive Journals, and the House and Senate Journals, for a total of 11,935,564 pages. A total of 74,495 maps had been conserved as well.

The project was extremely valuable to the Library as it allowed a focused repair of an extremely large and valuable collection. In addition to the conservation treatment, individual items from the collection were finally added to the library catalog through bar coding and item record creation thus bringing the collection into circulation control. It was estimated that that operation alone would have taken six months to complete.

The library project staff also developed experience with a large-scale digitization project. The tracking techniques, which were developed using spreadsheets and a wiki, have been carried forward and are currently used by staff in Preservation's Digital Production Unit. The most important aspect is the creation of a virtual Serial Set that is complete, something that exists in no single library.

Libraries are service organizations and we treated our partnership with Readex no differently. We knew we were at the beginning of the workflow so always kept ahead of the project by prepping several shipments ahead of time. We observed that throughout the project shifts in workflow could vary immensely depending on the physical condition of the volumes (thus requiring more conservation) or the state of the bibliographic record of the needed volume (thus more time needed to locate and verify the volume.) Our project fluctuated between ninety

percent treatment to ninety percent assembling of the collection.

There was often great difficulty in locating the individual volumes or verifying publications with in volumes. For that reason we found it useful to be flexible and alert our counterparts to difficulties. Communication was crucial throughout the life of the project. Take nothing for granted and over-communicate.

Occasionally a volume was too thick to be scanned on the Kirtas machine. When that would happen we would think of the greater project goals (creating a single virtual collection) and work with conservation technicians to temporarily disbind the volume for scanning.

Throughout the project organization was essential, be it through using the

circulation system to track volumes, spreadsheets to record publication information, or a wiki tool for shared access to documents. Technology is essential to communication with project members; our wiki tool was critical.

This grand work was completed in 2013, and both teams gathered to celebrate the conclusion.<sup>9</sup> It was bittersweet—we were proud of the work we all had accomplished and were sad to see it end. Our counterparts at Readex were top-notch professionals who valued and cared for the Serial Set as if it was their own. Our shared values of access,

---

¶ Carol Forsythe, "Preserving a National Treasure: A Partnership with the Dartmouth College Library," *Readex Blog*, January 6, 2014, [www.readex.com/blog/preserving-national-treasure-partnership-dartmouth-college-library](http://www.readex.com/blog/preserving-national-treasure-partnership-dartmouth-college-library)

service, preservation, and communication resulted in a high-quality product for Readex and an amazing amount of conservation work being completed for Dartmouth.

May 19, 2016

Barb Sagraves (sagraves@dartmouth.edu), Head, Preservation Services and The Book Arts Program, Dartmouth College Library, Hanover, New Hampshire

Report by the GODORT Preservation Working Group (Tom Adamich, Co-Chair; Bernadine Abbott Hoduski, Co-Chair; Sarah Erekson; Jim Noel, Marcive; Alar Elken, Newsbank.Readex; Andrew Laas, ProQuest), June 14, 2016.