



National Collection of U.S. Government Publications

Contents:

- I. Briefing Topic: National Collection of U.S. Government Publications2
 - I.1 Setting the Stage2
 - I.2 Conceptual Overview of the National Collection2
 - I.3 New Information3

- II. Revised Assumptions4

- III. Questions to Council, with Council Discussion5
 - III.1 Are the revised assumptions on the National Collection still appropriate,
or should something be changed or added?
 - III.2 Now that the plan for the National Collection has been revised in light of
discussions with Council and other public comments, are there elements of
the plan that need further refinement or need to be added?.....
 - III.3 In the event that a publishing agency indicates a desire to withdraw a publication
from the FDLP does the proposed revision of ID 72 adequately provide for
necessary actions?.....

- IV. Questions from Council Addressed at the Meeting6

- V. Audience Questions Addressed at the Meeting7

I. BRIEFING TOPIC: NATIONAL COLLECTION OF U.S. GOVERNMENT PUBLICATIONS

I.1 SETTING THE STAGE

The National Collection of U.S. Government Publications (National Collection) supports the Federal Depository Library Program (FDLP) in its mission to ensure no-fee permanent public access to the official publications of the United States Government regardless of format.

The National Collection includes preservation and access copies of digital publications and tangible publications. The National Collection will consist of geographically-dispersed collections of tangible and digital publications, located at multiple sites, and operated by various partners within and beyond the U.S. Government. This briefing topic is focused on the National Collection dark archives for preservation of tangible and digital publications.

I.2 CONCEPTUAL OVERVIEW OF THE NATIONAL COLLECTION

Contents	National Collection Preservation Sites	Access Collections for Public Use
Tangible publications	Dark archives with preservation copies	Light archives (minimal use, active preservation), or Depository library collections (normal preservation efforts) with access copies
Digital publications	Dark archives with digital preservation masters	Service archives at GPO or partner sites with access copies or derivatives

For tangible publications the dark archive(s) include preservation copies which are only used when all other avenues to obtain a copy have been exhausted. Public access is provided from depository library collection access copies, or from light archives, which have minimal use and active preservation.

Preservation copies of digital publications will be maintained in dark archive(s), and used to create access copies or derivatives for delivery from GPO service archives or partner sites.

GPO will provide online public access and other publications and services derived from the digital preservation masters and other items in the National Collection. Access copies, also referred to as “derivatives,” of the stored digital publications will be available for no-fee online use by the public as well as for print-on-demand and document delivery services. Publications in GPO portions of the National Collection will be bibliographically controlled using GPO’s Integrated Library System. The cataloging records will be part of the National Bibliography data base, provided to OCLC, and made available to requesting libraries and vendors. Ultimately, preservation metadata will be included in GPO’s Future Digital System.

I.3 NEW INFORMATION

National Collection Acquisitions

The National Collection receives two copies of every in-scope tangible publication printed by or through GPO, whether or not it is distributed through the FDLP in tangible form, as well as two copies of every publication produced by print-on-demand. GPO will preserve the digital master files and print-on-demand master files, as well as access files and other derivatives.

In addition, there have been a number of important acquisitions from depository and Federal libraries, including:

- Complete *U.S. Congressional Serial Set*
- Complete *Records of the War of the Rebellion*
- Bound *Congressional Globe*, 1866-1873
- *Foreign Relations of the United States*, 1907-1957
- 82 folios of the *Geologic Atlas of the United States*
- *Monthly Catalog*, 1959-1999
- *Federal Reserve Bulletin*, 1918-2002
- Over 50 years of reports from the U.S. Tax Court
- *Social Security Bulletin*, 1939-1989
- Numerous U.S.G.S. Circulars

ID 72

GPO is in the process of revising SOD 72 and the new policy, ID72, once approved, will govern GPO's response to agency requests to withdraw material from the FDLP. (Copies of both the proposed revision and the current policy statement are attached.)

Future Digital System

Preservation is one of the core functions of GPO's Future Digital System (FDSys). The FDSys planning team, under the leadership of the Office of Innovation and New Technology, has developed a detailed set of requirements for the System. Overall, the System design conforms to the OAIS (Open Archival Information System Reference Model, ISO 14721:2003) reference model for an archive. Preservation planning for the FDSys includes requirements for storage, access, preservation processes, metadata, security, system workflow and administration, and other functions.

Center for Research Libraries' Publication

GPO recently placed online *Baseline Requirements for Digital Reformatting and Delivery of Legacy Federal Documents Collections*, prepared for GPO by the Center for Research Libraries. The report describes CRL's recommendations for the minimum requirements governing file formats, interoperability, accessibility, asset management controls, and other aspects of digital reformatting and delivery for print and microform legacy materials. This report was used in developing the preservation requirements for the FDSys. The Baseline Requirements report is available from the FDLP Desktop at: http://www.access.gpo.gov/su_docs/fdlp/pubs/crl_digital_baseline.pdf and is included as an attachment.

NARA

GPO and NARA are discussing transforming the set of FDLP tangible publications that NARA currently holds for GPO into one of the proposed National Collection dark archives. That would allow NARA to move that material to storage, providing greater preservation for those materials. NARA will continue to refer users to FDLP collections for tangible documents and will use the digital copies in the EC for access. GPO is working with NARA to develop procedures for the addition of publications to the National Collection dark archive that were not previously distributed to depository libraries. This will

allow GPO to assemble comprehensive coverage of all content that should be in the FDLP, whether it was distributed at the time of publication or not.

II. REVISED ASSUMPTIONS

- II.1 The National Collection primarily supports the FDLP goal of no-fee permanent public access, but also supports other GPO information dissemination and preservation programs.
- II.2 Administration of the National Collection and management of the tangible and digital publication dark archives are inherently governmental responsibilities.
- II.3 Publications in the National Collection will be included in the National Bibliography of U.S. Government Publications.
- II.4 National Collection assets will be maintained in geographically-dispersed locations. A National Collection segment may be consolidated in a single location or as a distributed collection at multiple sites which together form a single collection.
- II.5 Management of the National Collection light and dark tangible archives will be benchmarked against the criteria for assurance developed by the Center for Research Libraries in *Federal Document Repositories: Decision Framework by Tangible Repository Type* or successor documentation of best practices.
- II.6 National Collection preservation activities for digital resources will be based on the agreement between GPO and the National Archives and Records Administration (NARA) designating GPO as an archives affiliate.
- II.7 The National Collection includes the FDLP Electronic Collection.
- II.8 National Collection assets in the “dark archives” are held for preservation rather than public access.
- II.9 Access copies of the National Collection FDLP assets are publicly accessible from light archives (tangible) or service archives (digital).
- II.10 GPO will acquire tangible copies from a variety of sources, including the transfer of portions of the legacy FDLP Collections from depository libraries to GPO.
- II.11 The print publications in the National Collection will be digitized for preservation and access. After digitization the original publication, even if disbound, will be retained and preserved in case the item must be digitized again.
- II.12 Tangible copies of “born digital” publications may be produced for the dark archive as backups for the digital publications when such copies may be produced without loss of content or functionality. Tangible copies so produced for the National Collection dark archive will, to the extent practicable, be produced on archival media; e.g., preservation microfilm.
- II.13 The National Collection will utilize a variety of preservation processes; e.g., deacidification and rebinding for print publications, and refreshing, migration, and emulation for digital content.
- II.14 GPO will continue to consult and participate with other Government information and preservation organizations to share best practices and inform GPO’s decisions on preservation.
- II.15 Preservation masters produced by GPO will conform to evolving standards; e.g., standards described in the Report on the Meeting of Experts on Digital Preservation or successor documentation.

III. QUESTIONS TO COUNCIL, WITH COUNCIL DISCUSSION

III.1 **QUESTION: Are the revised assumptions on the National Collection still appropriate, or should something be changed or added?**

DISCUSSION BY COUNCIL

After discussion, Council agreed with Assumptions 1, 3, 4, 6, 7, 8, and 10 as presented.

Assumption 2 was changed to reflect Council's distinction between an inherently governmental function and a responsibility.

Assumption 5 was clarified so that it pertains to both light and dark archives and to clarify exactly which Center for Research Libraries report is referred to.

Assumption 9 was clarified by adding the modifier FDLP to the phrase "National Collection FDLP assets".

Assumption 11 now specifies print publications are the candidates for digitization, rather than all tangible publications which would include digital media.

Assumption 12 was substantially modified after the Council discussion. Council was concerned about the cost implications of producing print copies of born digital content, whether those copies are on paper or film. There was also concern over the loss of functionality compared to the digital originals when making these backup copies. The modified Assumption still provides for the possibility of such conversions as deemed necessary for program needs.

Three new assumptions have been added based on the Council discussions, numbers 13-15.

Assumption 13 articulates that GPO will utilize a variety of preservation processes as appropriate for various formats or presentations of content. These may include deacidification and rebinding for print publications, and refreshing, migration, and emulation for digital content.

Assumption 14 states that GPO will continue to take an active role in Government information and preservation organizations to share best practices and inform GPO's decisions on preservation.

Finally, Assumption 15 articulates that GPO-produced preservation masters will conform to prevailing standards.

III.2 **QUESTION: Now that the plan for the National Collection has been revised in light of discussions with Council and other public comments, are there elements of the plan that need further refinement or need to be added?**

DISCUSSION BY COUNCIL

Council's discussions of the issues raised in the review of the Assumptions led to addition of the three new assumptions described above.

III.3 **QUESTION: In the event that a publishing agency indicates a desire to withdraw a publication from the FDLP, does the proposed revision of ID 72 adequately provide for necessary actions?**

DISCUSSION BY COUNCIL

Council raised numerous questions and issues regarding the draft ID 72, which is the internal policy document regarding the withdrawal of previous unrestricted publications from the FDLP.

To illustrate, three hypothetical situations were described that could occur with an agency coming to GPO and identifying a publication that had been inappropriately disseminated. One situation concerns the materials distributed to depository libraries, which remain Federal property, and therefore form the legal basis for any recall.

In the case of the sales program, the ownership transfers with the sale. GPO has made it clear in the draft policy that there is no Government ownership or other right, as there is with the Federal Depository Library Program, that would inherently make it easy to recall those materials. Similarly, with the International Exchange copies, once GPO has sent those materials to a foreign Government on behalf of the Library of Congress for the International Exchange Service, they are no longer U.S. Government property; they are now property of a foreign sovereign nation.

In discussing potential recalls with the publishing agency, GPO would disclose to them all the places the publication had been distributed and what the implications of a recall decision are. GPO's position is that our first choice is not to recall something. There are many times when an agency approaches GPO and we talk through the whole process with them and the nature of the information, and the agency steps back and it never goes anywhere.

Council's next question involved actions concerning the bibliographic record staying in the ILS with an annotation of the action unless the record is deemed classified. It is GPO's position that the record not be expunged from the catalog, unless the existence of the document is itself classified.

Council asked the purpose of the provision for notifying the Depository Library Council and the Library Associations. GPO's objective is to convince the publishing agencies that they do not need to request a recall.

IV. QUESTIONS FROM COUNCIL ADDRESSED AT THE MEETING

Council raised a question about statutory authority for the National Collection and its management, and whether preservation in light or dark archives is an inherently Government function or one conferred upon the depository libraries. GPO responded that the legislative authority for digital preservation is derived from 44 U.S.C. Chapter 41, which is the GPO Access legislation. An area in which GPO recognizes the need for legislative authority is for the redevelopment and retention of income from the building. GPO has stated in the strategic vision that GPO will work with the oversight and authorizing committees to be sure that they concur, and to see if they believe any additional statutory authority is necessary.

Council asked how maps would be treated in GPO's legacy digitization effort. The response was that, at this point in time, GPO is not ready to digitize cartographic materials, but that it was a future possibility. In the near term, the tangible versions would be preserved in the National Collection. In addition, there may be cartographic materials that are solely available in digital format that GPO would try to capture for preservation.

Council asked if GPO's agreement with NARA envisions that GPO would assume the archival responsibility for Government information in lieu of NARA, or would NARA be duplicating that effort. GPO's current agreement with NARA covers the core information on GPO Access. GPO has assumed the archival responsibility for that, which is what being an archives affiliate means. If GPO is to extend that agreement in the future to cover a broader scope, a modification to the agreement will be needed. The potential exists for amending the agreement, but our preservation of this material under NARA's records authority actually documents GPO's fulfillment of its responsibilities as an agency, but transfers

from GPO do not fulfill the publishing agencies' responsibility to NARA. One of the other things GPO and NARA have discussed doing as an amendment at some point in time is to work on a trilateral agreement between the publishing agency, NARA, and GPO to alleviate the publishing agency responsibility to also deposit electronic copy with both GPO and NARA, and by doing so would fulfill responsibilities under the Electronic Records Act.

V. AUDIENCE QUESTIONS ADDRESSED AT THE MEETING

The facilitator of the Council sessions accepted questions from the audience written on GPO-supplied cards. Twelve of twelve questions were answered during the Council session. Those questions and their answers are summarized below.

V.1 QUESTION: Is the transfer of documents from a depository to the National Collection voluntary?

RESPONSE: Yes.

V.2 QUESTION: Is the set of FDLP publications at NARA the same collection content as that at many regional Depository Libraries that were formed from 1907 or under the Land Grant designations?

RESPONSE: GPO is in discussion with NARA on the scope and composition of its Government publications collection. The NARA collection includes the material that GPO, in the course of its responsibilities as a Government agency, has routinely turned over to NARA as evidence of our program activities.

V.3 QUESTION: What technical standards are going to be used to select the diverse locations of National Collection of U.S. Government Publications?

RESPONSE: The best practice for geographically-dispersed collections is a 600-mile separation, with the proviso that they both must be outside of the Washington, DC, metropolitan area. GPO has referred to Eastern and Western sites for the Collection, but it is premature to earmark locations at this time.

V.4 QUESTION: Will the entire NTIS inventory become part of the National Collection?

RESPONSE: Some of the NTIS holdings are clearly not in scope, because they are not all U.S. Government documents. But of the things that are in scope, they have the potential to be there, although there are legal and other issues to discuss.

V.5 QUESTION: Will GPO work with the library community to develop an improved mechanism to identify libraries holding tangible publications? The current item profile base system doesn't tell the whole story by reflecting only the current profile. Is there a better way, which would not be labor-intensive, but would allow for non-receipts, weeding, and library profile changes?

RESPONSE: GPO wants to work with the library community to have improved mechanisms, but the process is probably labor-intensive. The ILS could eventually have holdings, and GPO could put in holdings as we distribute things. But libraries would have to notify GPO when publications are removed from your collection, and that holdings maintenance is labor-intensive regardless of who performs it.

V.6 QUESTION: What areas of work on the National Collection do you envision being done by partners, cataloging, selection, or by machine, harvesting, etc.? What percentage of the work could be farmed out to institutions outside GPO? Will this speed up the process or will the process be long no matter what the methodology?

RESPONSE: Potentially most of those areas would lend themselves to working with partners, and GPO would be happy to have partners come and make proposals to take on part of this work. We have service partnerships underway already. GPO declined to estimate what percentage of work could be accomplished through other means, but forecast that the process of building the Collection is going to be long and complicated.

V.7 QUESTION: I am interested to hear about any cooperative efforts being discussed and initiated with the Library of Congress regarding the National Collection.

RESPONSE: The Public Printer has met with the Archivist of the United States and with the Librarian of Congress, and GPO has contacts at the staff level. GPO and LC agree in principle that the Library of Congress's primary focus is on materials that are under copyright, and that LC is comfortable with the idea that if GPO focuses on U.S. Government information, that frees LC's time and resources for other purposes. The Library of Congress has indicated that they are looking at more collaboration themselves.

V.8 QUESTION: In GPO's response to the ALA GODORT comments on the draft of ID 72, specifically referring to the final comment, GPO should not presume that communicating with the library community about withdrawal requests that are in process is unnecessary. Isn't our knowing what's kept and your reporting it to us an excellent chilling effect on Government agencies removing their content?

RESPONSE: GPO clarified that the communication with the library community concerns recall requests once the decision has been made by the agency. GODORT asked why GPO does not inform the community as soon as an inquiry comes from the agency. GPO does not believe that is the appropriate time for communication; the agencies involved need to discuss and work through the issues, and that discussion may lead to a decision not to recall a document. Another issue is if there is a recall because there is a legitimate national security concern letting that information out before GPO is ready to execute a recall could start a scramble for people trying to find that document and sequester it, and that could be dangerous.

V.9 QUESTION: One very important reason why the GPO's procedure for agencies' requests for removal of material from the FDLP works as well as it does is because the bulk of the material distributed by the FDLP is in tangible format. Tangible equals very inconvenient to withdraw. How can ID 72 deal with born digital materials that will be far easier for agencies to withdraw or alter? In addition, what procedures are there to deal with agencies that want to block GPO Web harvesting efforts?

RESPONSE: One of the reasons we updated the draft of ID 72 was to make sure it reflected more about what was online, to make sure that there was an equivalently rigorous process for removal of electronic content as there was for removal of print. When an agency has taken online content off their Web site and wanted us to take it out of our server, we are discussing it under the framework of this revised policy. At one level, it appears simpler to take something down from one or two locations than to get it back from several hundred to over a thousand depository libraries. But the reality is that all of these documents, once they've been on the Web

site, have been downloaded many places, such as the Way-Back Machine and the Internet Archive. In the CENDI group that Dr. Warnick and I both participate in, one of the topics is the need to communicate to other parts of their agencies, and potentially to the broader Governmental community, about the futility of taking things off Web sites once they are posted, because they believe some people don't even understand the extent you can't scrub it off the Internet. If they understood the extent to which taking off their Web site was a futile action, they would be less likely to bother to take it down.

GPO is working with these major technology agencies to communicate about this. I think it's very important that this policy address the electronic content. We intend to enhance our Memorandum of Understanding with NARA so that in the event of recall of an electronic document, that either GPO will hold it in our digital archive or NARA will hold it, so that it isn't lost and that it is available if in the future it is cleared for public access. We're making similar provisions on the print, that the print would be withdrawn, but we wouldn't destroy all copies. GPO believes that, at some point, these things will come back and probably be allowed for public access.

On the harvesting issue, we are going to have some conversations with agencies. One of our concerns was that there are a lot of agencies that have no idea that GPO is harvesting online publications and putting them in servers. We feel that part of that trust relationship in the acquisitions process is to let them know what we're doing. Certainly EPA is going to be very aware that we're crawling their Web site, because we're working together on a pilot project. But I think we have an affirmative obligation to let agencies know, and if they have a concern about harvesting, to then have that discussion about their responsibilities under Title 44. It's in our interest not to be invisible to them, and it is important that they know we're doing this and they cooperate with us, rather than doing something without their knowledge or consent. I do not expect that we'll be blocked. GPO is only crawling public Web sites. NARA harvested all of the public Web sites at the very end of President Bush's first term in 2004, and those snapshots are preserved. If content is taken down after January 20th, it is still available up on the NARA site on their harvested version of the Web site. NARA is not going to go back and scrub those either. Some of this is educating agency people to really understand the technology.

COUNCIL FOLLOWUP COMMENT: The threat to the Depository Library Program, the threat to GPO, is by far greater from fugitive documents than it is from taking documents down. The number of documents taken down is a handful. The number of the documents that are fugitive is in the tens of thousands. We are focusing on the wrong thing by worrying about the few documents that are taken down as opposed to the tens of thousands that never show up in the first place.

V.10 QUESTION: When will the creation of light archives begin, how will that happen, and who will make these decisions?

RESPONSE: GPO is already having preliminary discussions about the creation of a light archive, and this is a process where people in the library community are going to need to bring proposals to GPO, and then through an iterative process, we will try to work through all the issues and develop an agreement that suits the interests of all the partners and the public. Right now there is no road map for developing light archives partnership agreements, but GPO is ready to start working on one if people in the field are ready.

V.11 QUESTION: In 1989 or early 1990, I called NTIS to try to inquire about an early PB series technical report. I was told that all their rolled microfilm publications, which would be 1946 to

sometime in 1964, had been transferred to the Library of Congress and are now available through the LC Reading Room, Science and Technology Division. I think I read elsewhere that LC was storing its films at Wright-Patterson Air Force Base in Dayton, Ohio. LC has probably built a film vault by now. Maybe it's at their Fort Benning remote storage site. I would have no way of knowing if it is just in custody of LC or those films are now owned by LC. Do you want to add any clarity to that?

RESPONSE: GPO has forwarded this question to the Library of Congress.

V.12 QUESTION: The National Collection is a permanent collection that is intended to provide comprehensive, timely, permanent public access to the final published official versions of U.S. Gov pubs in all formats. Please elaborate what is meant by final published official versions. What about early editions? What about things like press releases, frequently updated material like overseas business reports?

RESPONSE: All of those examples qualify as final published versions, because they are ones that the agency has cleared for public release. There are things that would not be retained for the National Collection. In the context of the Future Digital System, we are thinking about the capability for agencies to come in and work in GPO's system space to create documents, so they might use versions that are works in progress that are being added to or edited, but haven't yet been finalized or cleared for release.