



National Collection of U.S. Government Publications—formerly known as Collection of Last Resort

Contents:

I. Briefing Topic: National Collection of U.S. Government Publications	2
I.1 Setting the Stage	2
I.2 New Information.....	3
I.3 Micro Recap	3
II. Revised Assumptions	4
III. Questions to Council, with Council Discussion	5
III.1 Is the “Last Resort” the right terminology for this collection?.....	5
III.2 The need for redundancy is recognized. How many collections and how many copies provide sufficient redundancy?.....	6
III.3 Should all National Collection components be geographically separated, i.e., should a digital NC and a tangible NC be collocated?	7
III.4 Must the NC be able to supply an exact replica of the preserved object? Or is a copy that presents all of the content but does not replicate the packaging acceptable? ..	7
III.5 Should all the information in the NC be official?.....	8
IV. Questions from Council Addressed at the Meeting	8
V. Audience Questions Addressed at the Meeting	9

I. BRIEFING TOPIC: National Collection of U.S. Government Publications— formerly known as Collection of Last Resort

I.1. SETTING THE STAGE

The U.S. Government Printing Office (GPO) Collection of Last Resort (CLR) supports the GPO mission to provide comprehensive, timely, permanent public access to U.S. Government publications in all formats. The primary purpose of the CLR is to support 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.

The CLR consists of multiple collections of tangible and digital publications, located at multiple sites, and operated by various partners within and beyond the U.S. Government.

The CLR is comprehensive and includes publications of the Federal government, which are of public interest and educational value, regardless of format. Publications classified for reasons of national security and those produced solely for administrative or operational use are excluded by law from depository distribution. However, whenever possible administrative and operational publications will be acquired for the CLR, identified by metadata and included in the National Bibliography. Since the legal scope of the GPO Cataloging and Indexing Program is broader than that of the FDLP, some products will be included in the CLR solely because they are represented in the National Bibliography. The CLR will also serve as the repository for products from future GPO business initiatives.

GPO will proactively acquire and preserve tangible and electronic copies of Government publications for inclusion in the CLR based on the requirements of all GPO information dissemination programs. In addition to publications acquired, harvested, or created for the information dissemination programs, the CLR will include agency source data files acquired pursuant to the OMB compact or other GPO services to publishing agencies. The CLR will support diverse GPO organizations and operations through access to stored digital objects. GPO will provide online public access and other information products and services derived from the digital preservation masters and other items in the CLR.

Access copies of the stored digital objects will be available for no-fee online use by the public and for print-on-demand and document delivery services. The CLR will enable Federal depository libraries to access digital copies or to acquire printed copies for their collections. In addition, Federal depository libraries will be able to consolidate or reduce their local tangible FDLP Collections secure in the knowledge that copies will be perpetually available from the GPO CLR.

Bibliographic access to all items in the CLR will be provided through GPO's National Bibliography and potentially by other metadata services. Cataloging records for online publications will include a persistent link to the publication. Digital objects will be accompanied by preservation metadata describing their content, file type, provenance, etc.

**Table 1. Conceptual Overview of the
Federal Depository Library Program Collections**

Contents	National Collection	Access Collections for Public Use
Digital Objects	Preservation masters in dark archive(s)	Access copies from <i>GPO Access</i> or partner sites
Tangible Publications	Preservation copies in dark archive(s)	Access copies in: <ul style="list-style-type: none"> • Light archives <ul style="list-style-type: none"> ○ Minimal use, active preservation • Depository library collections <ul style="list-style-type: none"> ○ Normal preservation efforts

Digital objects may be ingested or created for the FDLP Electronic Collection portion of the National Collection. Creation includes digitization activities conducted by GPO, depository libraries, or other partners. Ingested digital objects include “born digital” files from agency publishing activities as well as objects harvested from the Web. Digital objects in the National Collection will initially be text with accompanying graphics, and the most prevalent file types in the near term are expected to be TIFF, PDF, HTML, and ASCII. In the future the National Collection may include video, audio, and other non-text file types.

A publication that has been digitized by GPO or its partners will be represented in the National Collection in multiple formats, including the original format, the digital preservation master, and one or more access file formats.

As the legacy documents are digitized, access copies will be available for search and retrieval, dissemination, or repurposing for print-on-demand and other services. GPO will coordinate digitization efforts with the library and other interested communities to establish priorities, reduce duplication of effort, and ensure the use of broadly acceptable digitization standards.

Tangible copies of “born digital” products will be produced for the dark archive as backups for the digital objects in the National Collection. If an access or public use copy of a National Collection print title is required, it will generally be reproduced from a digitized version.

I.2. NEW INFORMATION

Collection of Last Resort was developed and presented for Council review in April 2004. The plan was extensively revised based on comments received between the April meeting and early June. The plan was released for public comment to a broader audience on June 18, 2004. The original deadline for comments, July 30, was extended to September 7, 2004.

I.3. MICRO RECAP

The Federal Depository Library Program Collections (FDLP Collections) include preservation and access copies of digital objects and tangible publications. These collection components are geographically dispersed, serve different functions, and are managed according to their specific roles in the overall program for public access to government information. As shown in Table 1 (below), the Collection of

Last Resort serves three roles in the conceptual overview, serving as the dark archive for preservation of tangible publications and digital objects as well as providing online access.

The CLR will become, over time, a comprehensive set of tangible and electronic titles that will back up the tangible collections in regional depository libraries or shared repositories into which regional library collections may be consolidated in the future.

II. REVISED ASSUMPTIONS:

- II.1 The NC is primarily created to support the FDLP goal of no-fee permanent public access, but also supports other GPO information dissemination and preservation programs, including print-on-demand for publications sales.
- II.2 General administration of the NC and management of the tangible and digital publication dark archives are inherently governmental functions.
- II.3 Publications in the NC will be included in the National Bibliography of U.S. Government Publications.
- II.4 NC assets will be maintained in geographically dispersed locations. An NC collection may be consolidated in a single location or as a distributed collection at multiple sites which together form a single collection.
- II.5 NC management will be benchmarked against the criteria for assurance developed by the Center for Research Libraries or successor documentation of best practices.
- II.6 NC preservation activities for digital resources will be based on the agreement¹ between GPO and the National Archives and Records Administration (NARA) designating GPO as an archives affiliate.
- II.7 The NC includes the existing FDLP Electronic Collection. The FDLP Electronic Collection consists of:
 - a. *GPO Access*, i.e. core legislative and regulatory documents such as the *Congressional Record*, *Federal Register*, and other government information.
 - b. Electronic publications published or made available by GPO, within specific agreements for services between GPO and the originating agency.
 - c. Electronic publications published and made available by their originating agencies, which GPO identifies, describes, and links to at the agency site or from an EC access site.
 - d. Tangible electronic Government publications, such as CD-ROM or DVD-ROM, which GPO distributes to libraries.
 - e. Digitized files created, typically by scanning with or without optical character recognition, by GPO's partners. GPO's partners may include publishing agencies and other partners such as depository libraries.

¹ Memorandum of Understanding (MOU) Between the Government Printing Office and the National Archives And Records Administration , August 2003, <http://www.gpoaccess.gov/about/naramemofinal.pdf>

- II.8 GPO will have a NC of digital materials, the FDLP Electronic Collection, including:
 - a. Objects born digital and acquired by discovery or harvest.
 - b. Digital preservation masters resulting from printing composition or related processes.
 - c. Digital preservation masters scanned or otherwise produced from tangible originals.
 - d. Access copies of digital objects derived from the preservation masters.
- II.9 The contents of the NC will be described by standard metadata schemes appropriate for various program needs, including:
 - a. Access metadata.
 - b. Preservation metadata.
 - c. Persistent links, such PURLs, Handles, or DOI (Digital Object Identifiers).
 - d. Digital and tangible assets in the “dark archives” of the NC are held for preservation rather than public use.
- II.10 Access copies of the electronic assets in the NC will be publicly accessible.
- II.11 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.12 The tangible products in the NC will be digitized for preservation and access. After digitization the original publication, even if unbound, will be retained and preserved in case the item must be digitized again in the future.
- II.13 Tangible copies of “born digital” products will be produced for the dark archive as backups for the digital objects in the NC. Tangible copies in the NC dark archive will, to the extent practicable, be produced on archival media.

III. QUESTIONS TO COUNCIL, WITH COUNCIL DISCUSSION

III.1 QUESTION: Is the “Last Resort” the right terminology for this collection?

DISCUSSION BY COUNCIL

There was general agreement that “Last Resort” would not function well as a name for the collection. Alternative suggestions included National Collection, FDLP Collection, GPO Archiver, or GPO Preservation Depository. There was concern that the name should relate to the name of the National Bibliography, that it should reflect its relationship to the other collection components, and should reflect its storage preservation element and digital aspect.

III.2 QUESTION: The need for redundancy is recognized. How many collections and how many copies provide sufficient redundancy?

DISCUSSION BY COUNCIL

GPO has agreed to have two dark archives, but the number of light archives is in question. Maybe eight would be a sufficient maximum.

There was some discussion about the Stanford LOCKSS System, Lots Of Copies Keeps Stuff Safe. LOCKSS involves a minimum of eight to ten simultaneous caching machines to perform the authentication and other functions that the software provides. It's not really the same as the dark archive-light archive approach. It deals more with the preservation and authentication of access copies.

The redundancy of LOCKSS is very important. GPO should ensure that the Collection of Last Resort dark archive has at least two copies. But that shouldn't be extrapolated to the broader need for many dark archives of similar content and the community's need for actually preserving the access copies independently.

One of the beautiful features of the LOCKSS program is its low cost and overhead for institutions in managing the preservation of the electronic information, whether it's born digital or converted to digital format.

GPO, Council, and members of the FDLP as a whole need to discuss this issue further.

III.3 QUESTION: Should all National Collection components be geographically separated, i.e., should a digital NC and a tangible NC be collocated?

DISCUSSION BY COUNCIL

Collocation might be convenient but is not necessary; there was no negative reaction.

III.4 QUESTION: Must the NC be able to supply an exact replica of the preserved object? Or is a copy that presents all of the content but does not replicate the packaging acceptable?

DISCUSSION BY COUNCIL

This question raises social and legal issues that we may not be able to answer here. We don't know what the legal requirements might be.

There are also historical issues. What the item looks like in its original form can be very important to researchers, beyond what it contains. So it should be available both ways.

The issue here is rendering and not content, because the content would be an exact replication. It's the renderability of that information that might change. Renderability is the look and feel of the original item. It might look the same, but that look and feel might be very important in a certain context.

In some cases it's important to see a full replica of the preserved object, so it can be put back together and printed out the way it was originally planned to be.

A distinction needs to be made between permanent access copy and archive copy. If this is part of the partnership with National Archives, then obviously in the dark archive we need an exact copy. A reproduction should be as reasonable a facsimile as it can be.

III.5 QUESTION: Should all the information in the NC be official?

DISCUSSION BY COUNCIL

“Official” should be defined: does it mean legally acceptable or legally admissible? The files should be authentic.

There is also the concept of official designation, where as much as possible it will be the official copy, but for some of the older things, there never was an official copy.

We might have an exact copy of the original, but be unable to verify that it's an exact copy. It becomes awkward when it is a copy from an unofficial source. This raises the concept of best available copy. A statement could be made along the lines of “every effort has been made to ensure the authenticity and official nature of the materials.” In some cases, there are going to be things that there's no way to say this is officially official. It's going to be the courts and the legal system that are going to be the most concerned about official materials.

With the PKI and other technologies, for the newer born-digital materials, this question will answer itself. The concern is with materials from the nineties.

IV. QUESTIONS FROM COUNCIL ADDRESSED AT THE MEETING

IV.1 QUESTION: We're interested in learning how the progress on the LOCKSS service setting has gone in the GPO.

RESPONSE: GPO is very much interested in being a publisher to the LOCKSS system for Government content, if the LOCKSS documents program goes forward. The current system is organized to support primarily journals, but as it develops it might pick up non-serial items. We do see it as something to work with as part of the overall effort to see how we might facilitate distribution of electronic content to depositories.

IV.2 QUESTION: We have the two darkives and we feel good that those will come to pass, but what will the light archive situation be?

RESPONSE: With JSTOR, various groups of libraries are coming together spontaneously and deciding to have those light archives because they feel the need to do it. If GPO facilitates and encourages that, the same kind of thing can happen with Government documents.

IV.3 QUESTION: But JSTOR is not a preservation archive, it's people deciding to store things.

RESPONSE: We recognize that storage alone is not enough, but light archives for JSTOR are being developed, and they're being developed out of the community of users of JSTOR.

IV.4 QUESTION: We can't answer the questions about what light archives are going to be, are they shared depositories, et cetera. Some framework is going to have to start evolving so that institutions can decide what the boundaries are as the light archives grow. They have to go back to the basic program goals, but they will evolve. They're already starting to evolve, but we need them to be part of the continuum, and to not evolve in the wrong direction.

RESPONSE: We need to develop a continuum and a checklist, so that as people begin talking about putting things like this together, we have a way to check off different aspects, such as funding and governance, and make sure they have been considered.

In terms of our administrative responsibility, as people come to us with proposals, a checklist would give us a way to say, explain your proposal to us in the context of these things so that we know what you're proposing and where that fits on the continuum. GPO's responsibility is to help the dialogue along, and to at least assure people that we are not blocking it by our own rules and procedures.

V. AUDIENCE QUESTIONS ADDRESSED AT THE MEETING

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

V.1 QUESTION: I am hoping, and assuming, that the key assumptions questions and comments that occur within sessions will be distributed or made available to interested parties at a future time, hopefully sooner, rather than later.

RESPONSE: These key assumptions are all statements extracted from public comments. They have been out for comment and they will all be reflected back in the next generation of these documents. The intent here is to get input.

V.2 QUESTION: A procedural possibility (that could be an assumption) is to establish a "digitized on-demand" service to assure rapid delivery of materials not yet digitized.

RESPONSE: Procedures need to be worked out, but a quick PDF could be made as needed and the official preservation copy could be done later. Or alternately, a publication could be taken out of the queue and the full digitization done on demand.

V.3 QUESTION: GPO's focus on sale of dark, light, et cetera, seems to undercut depositories who wish to maintain tangible collections. GPO should still acknowledge individual FDLPs as the ongoing backbone of the CLR.

RESPONSE: The continuing presence of collections, both light archives and the traditional depository collections in the community, were in the master report as well as in the summary. And the archive isn't meant to be the access copy. The assumption is that those access copies remain out in the community and that this is a backstop to a failure in those collections.

And we've said in a number of these sessions that we would only expect changes in the tangible collections if the libraries wish to make them. There is no expectation that we would be driving people to change what they do with respect to their tangible collections, but rather to provide options and flexibility.

V.4 QUESTION: Are the FDLP "content" permanent site materials considered to be part of the GPO electronic collection? If so, is that defined in the document? If not, why not?

RESPONSE: Yes, and yes: the FDLP "content" permanent site materials are part of the GPO electronic collection, this is stated in the document.

V.5 QUESTION: How can GPO take a digital product produced at GPO's expense and not provide no-fee access to all users, regardless of where that access might be; home, FDLP, other sources?

RESPONSE: This is a matter of a distinction between the sales program and the depository and cataloging and indexing programs which are funded by an appropriation. Right now, when we buy print copies for the sales program, we ride the agency print order much like we ride using the appropriation to buy copies for the depositories. We ride with money that's retained earnings from the sales program to buy copies which are then sold to recover their cost and refresh that fund.

And the same thing would be true in electronic form. We have statutory authority to use the funds belonging to the sales program to create a product, hold it, sell it and recover that expense.

V.6 QUESTION: The 5-year retention requirement does not apply to federal agency libraries. Can we be sure to have a way for them to offer materials not retained? Many do not have time or staff to create major lists, but a centralized contact would be helpful and expedite the process.

RESPONSE: The simple answer to that is yes. Many federal agency libraries are voluntarily participating in Needs and Offers in order to share their resources.

And many are also willing to make sure that publications from their agencies are part of what becomes contributed for digitization for these other collections. So there will definitely be participation by federal libraries.

V.7 QUESTION: Will this digitization project necessitate a change in procedure for libraries, selectives and regionals, to discard items? That is, Needs and Offers lists, et cetera.

RESPONSE: Yes, it will necessitate changes and we haven't fully identified what all those changes need to be. We will need to have policies and procedures in place in advance of when some of this begins to happen.

V.8 QUESTION: For retention of the collection materials and digitization, does the plan account for any potential interest or interest in the trails or artifacts? Will there be some way to retain items in such a way that aspects such as bindings, physical properties of originals, et cetera, may be preserved?

RESPONSE: We are expecting to preserve the artifacts in the darkive and in other collections, using nondestructive digitization for rare materials. When there are lots of copies available, a destructive process might be acceptable, but for the older materials, we expect to use nondestructive processes.

V.9 QUESTION: Multiple sites need more than two tangible archives. Recommend at least four. Seems possible if so many want to remove last digitized.

RESPONSE: It's certainly possible that over time we can collect four dark archive collections. We think it's an inherent Governmental responsibility to fund and maintain these two darkives. I think it's unlikely we're going to get money from the Federal Government to maintain four dark archives.

The National Archives has one set of the things they retain right now. So we're already doing more than is done under the Federal Records Act. While we don't doubt that there would be a value to doing it, we doubt that there would be resources to do it.

V.10 QUESTION: The National Bibliography as a foundation of the CLR must insure "that no document is left behind."

RESPONSE: When we've talked about wanting to be comprehensive, we have repeatedly said that we would continue to look for documents until we found every single document we could conceivably find.

V.11 QUESTION: Does the documentation discussion include SuDoc numbers on the CLR documents? If yes, what is the result; if not discussed, why not?

RESPONSE: We are expecting, as we're doing the legacy selection, to capture those SuDoc numbers because that's the primary labeling on all those documents. The National Bibliography will include the SuDoc numbers.

V.12 QUESTION: Are two collections of last resort sufficient? Will either be a dark archive to protect the materials from theft? How will those researchers who need to study physical aspects of legacy materials get access?

RESPONSE: Access is tiered: there are selectives, there are regionals, there are potentially light archives and there are these darkives, which are primarily for preservation, not access and therefore are protecting those copies against damage and theft.