

# COLLECTION OF LAST RESORT

## U.S. Government Printing Office

Discussion Draft, April 6, 2004

Introduction.....	1
Collection Overview .....	2
Key Assumptions .....	2
Scope.....	4
Funding .....	4
Collection of Digital Objects .....	5
Collection of Print and Other Tangible Items.....	5
Acquisitions Sources and Strategies .....	6
Bibliographic Control .....	7
Access .....	7
Maintenance of Tangible Collection.....	8
Preservation.....	8
Location and Space .....	8
Appendix 1 : Guiding Principles.....	10

### Introduction

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. This draft plan represents GPO's thinking as of April 2004. This plan will continue to evolve as public comments are received and evaluated, as technology and the theory and practice of digital information preservation develop and as new knowledge becomes available.

Although 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 will support diverse GPO organizations and operations through access to stored digital objects.

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. 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 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.

## Collection Overview

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.

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

<b>Contents</b>	<b>Collection of Last Resort</b>	<b>Access Collections for Public Use</b>
Digital Objects	Preservation masters in a dark archive	Access copies from <i>GPO Access</i> or partner sites
Tangible publications	Preservation copies in a dark archive	Access copies in: <ul style="list-style-type: none"> <li>• Light archives (minimal use, some active preservation).</li> <li>• Depository library collections (no special preservation)</li> </ul>

## Key Assumptions

1. The CLR 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.
2. GPO will have an enterprise-wide CLR of digital materials, including:
  - a. Objects born digital and acquired by discovery or harvest.
  - b. Digital preservation masters resulting from print processes.
  - c. Digital preservation masters scanned or otherwise produced from analog originals.
  - d. Access copies of digital objects derived from the preservation masters.
3. CLR assets will be maintained in geographically dispersed locations.
4. CLR management will be benchmarked against the criteria for assurance developed by the Center for Research Libraries.

5. CLR preservation activities will be based on the agreement between GPO and the National Archives and Records Administration (NARA) designating GPO as an archives affiliate.
6. The CLR includes the existing the 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.
  - d. Tangible electronic Government publications, such as CD-ROM or DVD-ROM, which GPO distributes to libraries.
7. The contents of the CLR will be described by standard metadata schemes appropriate for various program needs, including:
  - a. Access metadata, such as AACR2 cataloging records.
  - b. Preservation metadata.
  - c. ISBN/ISSN or other unique identifiers.
  - d. Persistent links, such PURLs, Handles, or DOI (Digital Object Identifiers).
8. Digital and tangible assets in the “dark archives” of the CLR are held for preservation rather than public use. There are no direct users of the print copies in the dark archive.
9. Access copies of the electronic assets in the CLR will be publicly accessible.
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.
11. It will take three to five years to assemble the tangible CLR and digitize the 2.2 million titles (60 million pages) for the electronic CLR.
12. It is estimated that the depository library community and others will make an initial investment of \$50 million to digitize legacy FDLP Collection of print materials.
13. GPO estimates the Government’s portion of establishing and managing the CLR at approximately \$1.5 million per year for the next five years. Once the final plan is complete, we will be able to more accurately estimate the out-year funding requirements for this project.
14. The tangible products in the CLR will exist as a source and a backup for the digital objects CLR. After digitization copies of the original publication, even if disbound, will be retained and preserved in case the item must be digitized again in the future.
15. Digital objects in the CLR, whether harvested or created by GPO, will be certified as authentic, official versions.

## Scope

The CLR will become, over time, a comprehensive set of tangible and electronic titles that will backstop the tangible collections in regional depository libraries or shared repositories into which regional library collections may be consolidated in the future. The legacy collection of print documents is currently estimated at 2.2 million titles (60 million pages). Over the next three to five years, a comprehensive collection of tangible documents will be gathered for preservation and digitized for both preservation and public access. Most of the already existing titles for the tangible CLR will be obtained through voluntary transfers from depository libraries. New titles will be acquired by GPO as they are issued. The digitization of the legacy print collection will be accomplished in partnership with the depository library community and others. The partners expect to invest an estimated \$50 million in the retrospective digitization of print materials.

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 FDLDP, 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.

## Funding

GPO has included \$1.5 million in its FY 2005 Salaries and Expenses Appropriation request to cover the initial startup costs for the CLR. A major part of our effort in FY 2005 will be planning for the ultimate location and management of the CLR. We will explore the potential for establishing contractual relationships with libraries and other organizations to house the tangible CLR versus maintaining and preserving the tangible and electronic collections ourselves. These decisions will be made in consultation with the library community. To assist us with writing a final plan for the Collection of Last Resort, we have contracted with the Center for Research Libraries (CRL) for a study on the characteristics of and levels of assurance for repositories for such a collection.

The funding requested for FY 2005 is for the interim step, which will allow GPO to begin to assemble the content for the CLR while the final plan is being prepared. Initial expenditures in FY 2005 include the costs of transporting and storing materials that are acquired for the tangible CLR, purchasing storage equipment and supplies, and investing in the necessary information technology to develop and house the digital CLR materials. Once the final plan is complete, we will be able to more accurately estimate the out-year funding requirements for this project, but it is anticipated that it will cost approximately \$1.5 million per year for the next five years. Once the tangible CLR is assembled and the

legacy digitization is complete, the costs will be reduced to cover incremental addition of new content and maintenance of the established tangible and digital CLR.

After receiving approval by GPO management, the final plan will be presented to Congress.

## **Collection of Digital Objects**

Digital objects may be ingested or created for the CLR. 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 CLR 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 CLR may include video, audio, and other non-text file types.

Every new textual publication in the current stream of processing will be digitized if a digital copy is not already available. 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.

## **Collection of Print and Other Tangible Items**

Tangible copies of “born digital” products will be produced for the dark archive. Print products in the CLR will be preserved for their own value and as permanently available backups for the digital objects in the CLR. If an access or public use copy of a CLR print title is required, it will be reproduced from a digitized version.

The CLR is intended to fulfill customer information needs, expand options for access, and assure that the documentary history of the United States is permanently available. Activities that support these ends include:

- Eliminating out of print publications by offering print-on-demand.
- Acquiring two copies of every print publication selected for the FDLP and/or the National Bibliography.
- Capturing or creating a digital copy of all new publications on a flow basis.
- Digitizing historic publications in collaboration with the library community and other partners.

Tangible products in the CLR include:

- The format in which the agency produced the publication, including agency-produced microfiche, maps, posters and publications of all types.

- Microfiche produced under contract for GPO, when the source document is not available.
- Tangible electronic products, such as CD-ROM and DVD-ROM titles.

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

### Acquisitions Sources and Strategies

Strategies for acquiring current and retrospective products for the CLR are illustrated in the tables below.

**Table 2. Current Acquisitions**

<b>Tangible Information Products</b>	<b>Digital Information Products</b>
<ul style="list-style-type: none"> <li>● Riding agency print orders for two copies.</li> <li>● Agency mailing lists.</li> <li>● Acquiring fugitives.</li> <li>● External user or publishing agency notification mechanisms.</li> <li>● Documents Expediting Project (LC)</li> <li>● Depository library discards.</li> </ul>	<ul style="list-style-type: none"> <li>● Automated Web harvesting for individual products.</li> <li>● Manual mining of agency Web sites for individual products.</li> <li>● External user or publishing agency notification mechanisms.</li> <li>● Printing source files from agencies.</li> <li>● Official partnerships</li> </ul>

**Table 3. Retrospective Acquisitions**

<b>Tangible Information Products</b>	<b>Digital Information Products</b>
<ul style="list-style-type: none"> <li>● Unaccessioned materials at NARA</li> <li>● GPO processing “CLR boxes”</li> <li>● External user or publishing agency notification mechanisms.</li> <li>● Withdrawals (offers) from Federal depository libraries</li> <li>● Withdrawals (offers) from non - depository Federal libraries</li> <li>● Documents Expediting Project (LC)</li> <li>● Collections from libraries leaving the FDLP</li> <li>● Agency bibliographies</li> <li>● Booksellers</li> </ul>	<ul style="list-style-type: none"> <li>● Authentic digital copy obtained from an official entity or partner.</li> </ul> <p>NOTE: Digital objects are not considered “retrospective” until they are no longer available from the publishing agency.</p>

## **Bibliographic Control**

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

Bibliographic control will be provided to the individual product level for all access copies of publications in the CLR. Applying metadata at this level will enhance the performance of metasearch tools and OpenURL linking technologies. GPO bibliographic records will conform to the practices and standards established for the National Bibliography. Digital objects intended for print-on-demand reproduction and sales will also have ONIX metadata. The metadata for digital objects should indicate the permitted access to that item if any restrictions apply. Other or additional metadata systems or elements may be applied to other portions of the CLR.

## **Access**

The access copies of digital publications in the CLR will be directly accessible via links from bibliographic or other metadata descriptions or by direct searching of the content. Access for tangible copies, as shown in Table 1, is through the Federal depository libraries. Users requiring access to tangible titles will rely first on local depository collections, then on collections in regional depository libraries and finally on light archives in shared repositories that may be established by the depository library community in the future. A user must exhaust all opportunities for access to a tangible resource from the collections maintained in and by Federal depository libraries before seeking access to a tangible product in the Collection of Last Resort, and such access must be sponsored by a regional depository library that certifies that no other option is available and that the user requirement for access cannot be met by access to a digital copy or a print-on-demand tangible facsimile copy. As with other rare book collections, access to the tangible items in the CLR will be on site at the preservation facility and under specific guidance to be developed.

The terms and conditions for depository libraries to obtain print-on-demand copies of titles in the CLR are yet to be determined. Options being considered include an authorized account for each depository library with a pre-established value that can be used to order print copies, as well as the possibility for depository libraries to purchase additional print-on-demand items at a discounted price.

## **Maintenance of Tangible Collection**

- Tangible products in the CLR may be arranged by bar code, radio-frequency identification (RFID), accession number sequence, or successive technology for robotic retrieval.
- The CLR must include provisions for growth space.
- The tangible and digital dark portions of the CLR will be maintained in closed, non-public locations, outside the Washington, D.C. area.
- CLR security will be provided.
- GPO will benchmark its long-term preservation, storage, and management of the copies in the dark archives against current NARA guidance and preservation standards for print, microfiche and electronic materials.

## **Preservation**

A preservation plan that encompasses all formats and media represented in the CLR will be formulated within the first six months of the existence of the CLR.

Acquired retrospective materials will be evaluated upon intake and given appropriate preservation treatment.

Accepted preservation guidelines and best practices will be employed, particularly when publications are digitized.

Selection of digitization format must be consistent with long-term preservation capabilities.

Under the affiliated archive relationship with NARA, GPO will retain physical custody of specified permanent records that are accessioned into NARA's legal custody. Publications cataloged by GPO for the National Bibliography Program may in the future be among the records covered by this agreement. GPO is responsible for providing expertise in interpretation, access, and service for the publicly accessible portions of the CLR. GPO's practices will be guided by NARA's policies for reference, arrangement, description, preservation, and security.

## **Location and Space**

Preservation copies of tangible items in the CLR will be stored in environmentally controlled, secure facilities outside the Washington, D.C. metropolitan area. An arrangement using compact shelving would entail an initial space requirement estimated at 7,500 square feet. Using a "bin" system for robotic retrieval may require less space, but higher initial infrastructure investment. Geographically separate redundant facilities for the access copies of tangible products will be developed by GPO or its partners.



The online CLR will be located in multiple facilities for redundancy and security. Initially the GPO secure data storage facilities are expected to be in Washington, D.C., a location outside the Washington area, and the Alternative Congressional Facility. Under contract or other binding agreement, portions of the CLR may be located in other Federal agency facilities, depository libraries, or other non-Governmental organizations. Such agreements will define the roles and responsibilities of each partner institution. At least initially, the agreements will be modeled after GPO's content partnership agreements. (GPO's content partnerships may be viewed at [http://www.access.gpo.gov/su\\_docs/fdlp/partners/index.html](http://www.access.gpo.gov/su_docs/fdlp/partners/index.html).)

## Appendix 1: Guiding Principles

GPO will adhere to several guiding principles regarding Federal government information dissemination, including the following:

- *GPO's Report to the Congress: Study to Identify Measures Necessary For A Successful Transition To A More Electronic Federal Depository Library Program*. Principles for Federal Government Information. U.S. Government Printing Office Publication 500.11, June 1996.  
[http://www.access.gpo.gov/su\\_docs/fdlp/pubs/study/studyhtm.html](http://www.access.gpo.gov/su_docs/fdlp/pubs/study/studyhtm.html)
- U.S. National Commission on Libraries and Information Science (NCLIS) Principles of Public Information. <http://www.nclis.gov/info/pripubin.html>

Of specific note are the following excerpts from the NCLIS Principles of Public Information:

- The public has the right of access to public information.
- The Federal Government should guarantee the integrity and preservation of public information, regardless of its format.
- The Federal Government should ensure a wide diversity of sources of access, private as well as governmental, to public information.
- The Federal Government should not allow cost to obstruct the people's access to public information.
- The Federal Government should guarantee the public's access to public information, regardless of where they live and work, through national networks and programs like the Federal Depository Library Program.