

Collection of Last Resort

U.S. Government Printing Office
Washington, D.C.

Revised June 18, 2004

This document is located on *GPO* Access at
www.gpoaccess.gov/about/reports/clr0604draft.pdf

Comments on this document may be sent to Judy Russell, Managing Director,
Information Dissemination (Superintendent of Documents) at jrussell@gpo.gov.

Comment period ends July 30, 2004

CONTENTS

I. PREFACE.....	2
II. COLLECTION OVERVIEW... ..	2
TABLE 1. CONCEPTUAL OVERVIEW OF THE FEDERAL DEPOSITORY LIBRARY PROGRAM COLLECTIONS.....	3
III. KEY ASSUMPTIONS	3
IV. SCOPE.....	5
V. FUNDING.....	5
VI. COLLECTION OF DIGITAL OBJECTS.....	6
VII. COLLECTION OF TANGIBLE PUBLICATIONS.....	6
VIII. ACQUISITIONS SOURCES.....	7
TABLE 2. SOURCES FOR CURRENT ACQUISITIONS	7
TABLE 3. SOURCES FOR RETROSPECTIVE ACQUISITIONS	8
IX. BIBLIOGRAPHIC CONTROL.....	8
X. ACCESS.....	8
XI. CLR MAINTENANCE.....	9
XII. PRESERVATION	9
XIII. LOCATION AND SPACE	9
XIV. RELATIONSHIP WITH NARA.....	10
APPENDIX I: DEFINITIONS.....	11
APPENDIX II: GUIDING PRINCIPLES.....	14
APPENDIX III: PLANNING DOCUMENTS REFERENCED IN THIS PAPER.....	16

I. PREFACE

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 June 2004, and has been extensively revised based on the comments received in the April – June 2004 period. 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.

At the macro level, the CLR envisions the Government managing a complete depository collection. The CLR will consist 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 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.

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.

While frequently alluded to in this document, GPO's plans for the preservation and access to digital information are more fully articulated in the companion plan, *Managing the FDLP Electronic Collection, 2nd Edition, June 2004*, available at www.gpoaccess.gov/about/reports/ecplan2004rev1.pdf.

II. 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

Contents	Collection of Last Resort	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 (minimal use, active preservation). • Depository library collections (normal preservation efforts)

III. 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 a CLR 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.
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 (see Appendix III).

5. CLR preservation activities will be based on the agreement¹ between GPO and the National Archives and Records Administration (NARA) designating GPO as an archives affiliate.
6. The CLR 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. Digital 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.
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. ISBNs, ISSNs, 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.
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.

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>

14. The tangible products in the CLR will exist as a source and a backup for the digital objects CLR. After digitization the original publication, even if disbound, will be retained and preserved in case the item must be digitized again in the future.
15. Tangible copies in the CLR dark archive will, to the extent practicable, be produced on archival media.

IV. SCOPE

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. 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.

V. 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.

VI. COLLECTION OF DIGITAL OBJECTS

Digital objects may be ingested or created for the FDLP Electronic Collection portion of 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. 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.

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.

VII. COLLECTION OF TANGIBLE PUBLICATIONS

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

The CLR is intended to fulfill user 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 digital copies of all new publications.
- Digitizing legacy publications in collaboration with the library community and other partners.

Tangible products in the CLR include:

- The format(s) in which the publication was produced, including microfiche, maps, posters, and other publications formats.
- Microfiche produced under contract for GPO, when the source document is not available.
- Tangible electronic products, such as CD-ROM and DVD-ROM titles.

VIII. ACQUISITIONS SOURCES

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

Table 2. Sources for Current Acquisitions

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

Table 3. Sources for Retrospective Acquisitions

Tangible Information Products	Digital Information Products
<ul style="list-style-type: none"> • GPO records (FDLP publications) at NARA. • External user or publishing agency notification mechanisms. • Copies offered by Federal depository libraries. • Copies offered by other libraries. • Copies or collections from libraries leaving the FDLP. • Agency bibliographies. • Booksellers. 	<ul style="list-style-type: none"> • Authentic digital copy obtained from an official entity or partner. • Digital objects created by an official FDLP partner, i.e. from legacy collection digitization projects.

IX. BIBLIOGRAPHIC CONTROL

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.

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 book industry standard 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.

X. ACCESS

The access copies of digital publications in the CLR will be directly accessible via links from the National Bibliography or other metadata descriptions. Access to 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. The CLR dark archives are not open to the public, and have no

reading rooms or other public facilities. Access to publications in the dark archives will be provided to a digital copy or a tangible facsimile copy.

The terms and conditions for depository libraries to obtain tangible 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.

XI. CLR MAINTENANCE

- 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.

XII. 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.

XIII. 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 FDLP Electronic Collection, the digital portion of the 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.)

XIV. RELATIONSHIP WITH NARA

Like all other Federal agencies, GPO has a responsibility to transfer to the National Archives those products that are scheduled as permanent records of GPO's operation. This has historically included a record set of the tangible agency publications distributed in the FDLP as well as record copies of GPO publications such as the Monthly Catalog of U.S. Government Publications. GPO will continue to work within applicable records schedules to ensure that its records management responsibilities are fulfilled in all media and formats.

Under the affiliated archive relationship with NARA, GPO will retain physical custody of specified permanent records that are accessioned into NARA's legal custody. 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.

GPO and NARA have begun a discussion concerning transforming the set of FDLP tangible publications that NARA currently holds for GPO into one of the proposed Collection of Last Resort 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 materials to the CLR dark archive that were not distributed to depository libraries at the time of publication because they were classified, cooperative, or fugitive. This will allow GPO to assemble comprehensive coverage of all content that should be in the FDLP, whether it was distributed at the time or not.

APPENDIX I: DEFINITIONS

Access (or service) copy is a digital object whose characteristics (for example a screen-optimized PDF file) are designed for ease or speed of access rather than preservation.

Accessibility is the degree to which the public is able to retrieve or obtain Government publications, either through the FDLP or directly through an electronic information service established and maintained by a Government agency or its authorized agent or other delivery channels, in a useful format or medium and in a time frame whereby the information has utility.

Authenticity means that a digital object's identity, source, ownership and/or other attributes are verified. Authentication also connotes that any change to the object may be identified and tracked.

Born digital: Relating to a document that was created and exists only in a digital format

Collection of Last Resort, or CLR, is a comprehensive collection of all in-scope products content that should be (or should have been) in the FDLP, regardless of form or format. Products in the dark archive will only be used when no other copy is available from Program sources.

Collection Plan, or Collection Management Plan, means the policies, procedures, and systems developed to manage and ensure current and permanent public access to remotely accessible electronic Government publications maintained in the Collection.

Dark archive – A collection of tangible materials preserved under optimal conditions, designed to safeguard the integrity and important artifactual characteristics of the archived materials for specific potential future use or uses. Eventual use of the archived materials (“lighting” the archives) is to be triggered by a specified event or condition. Such events might include failure or inadequacy of the “service” copy of the materials; lapse or expiration of restrictions imposed on use of the archives content; effect of the requirements of a contractual obligation regarding maintenance or use; or other events as determined under the charter of the dark archives.

Distribution means applying GPO processes and services to a tangible product and sending a tangible copy to depository libraries.

FDLP Electronic Collection, or EC, means the electronic Government publications that GPO holds in storage for permanent public access through the FDLP, or are held by libraries and/or other institutions operating in partnership with the FDLP. These electronic products may be remotely accessible online products, or tangible products such as CD-ROMs maintained in depository library collections.

FDLP partner means a depository library or other institution that stores and maintains for permanent access segments of the Collection.

Format means, in a general sense, the manner in which data, documents, or literature are organized, structured, named, classified, and arranged. For example: full narrative text in English language in the form of books or articles; abstracts of text; indexes and catalogs; maps; photographs; sound recordings, video tapes, statistical and other tabulations, etc. A screen format is the layout of text or fields on the computer screen; a record format is the layout of fields with a record; a file or database format is the layout of fields and records within a data file.

Light archive – A collection of tangible materials preserved under optimal conditions, designed to safeguard the integrity and important artifactual characteristics of the archived materials while supporting ongoing permitted use of those materials by the designated constituents of the archives. A light archive normally presupposes the existence of a dark archive, as a hedge against the risk of loss or damage to the light archives content through permitted uses. A light archive is also distinct from regular collections of like materials in that it systematically undertakes the active preservation of the materials as part of a cooperative or coordinated effort that may include other redundant or complementary light archives.

Government publication means a work of the United States Government, regardless of form or format, which is created or compiled in whole or in part at Government expense, or as required by law, except that which is required for official use only, is for strictly operational or administrative purposes having no public interest or educational value, or is classified for reasons of national security.

Metadata, literally data about data, refers to the content of a surrogate record that describes or characterizes an object.

Official content is FDL P EC content that is acquired from the publishing Federal agency or its business partner.

The **official source** for FDL P information is the publishing agency or other trusted source.

Online dissemination means applying GPO processes and services to an online product and making it available to depository libraries and the public.

Online means the product is published at a publicly accessible Internet site.

Permanent access means that Government publications within the scope of the FDL P remain available for continuous, no fee public access through the program. For emphasis, the phrase "permanent public access" is sometimes used with the same definition.

Preservation means the activities associated with maintaining publications for use, either in their original form or in some other usable way. Preservation also includes

substitution of the original product by a conversion process, wherein the intellectual content of the original is retained.

Preservation master: A copy which maintains all of the characteristics of the original digital object, from which true copies can be made.

Storage, or Storage facility, means the functions associated with saving electronic publications on physical media, including magnetic, optical, or other alternative technologies.

Trusted content means official content that is provided by or certified by a trusted source.

Trusted source means the publishing agency or a GPO partner that provides or certifies official FDL P content.

Appendix II: 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
- 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.

APPENDIX III: PLANNING DOCUMENTS REFERENCED IN THIS PAPER

Decision Framework for Federal Document Repositories, Discussion Draft, April 12, 2004
www.access.gpo.gov/su_docs/fdlp/pubs/decisionmatrix.pdf

Managing the FDL P Electronic Collection, 2nd Edition, June 18, 2004
www.gpoaccess.gov/about/reports/ecplan2004rev1.pdf

The National Bibliography of U.S. Government Information: Initial Planning Statement,
June 18, 2004 www.gpoaccess.gov/about/reports/natbib0604.pdf