

Managing the FDLP Electronic Collection A Policy and Planning Document

Second Edition

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

June 18, 2004

This document is on *GPO Access* at
www.gpoaccess.gov/about/reports/ecplan2004rev1.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. EXECUTIVE SUMMARY.....	2
III. COLLECTION OVERVIEW.....	3
Table 1. Conceptual Overview of the Federal Depository Library Program Collections.....	3
Table 2. Overview of the FDLP Electronic Collection.....	4

IV. KEY ASSUMPTIONS.....4

V. DEFINING THE FDLP ELECTRONIC COLLECTION.....6

VI. OFFICIAL STATUS OF EC CONTENT.....7

VII. DISCOVERY, ACQUISITION, AND BIBLIOGRAPHIC CONTROL.....8

VIII. USER COMMUNITY.....8

IX. ASSURING ONGOING ACCESS.....9

 A. SELECTION FOR THE ELECTRONIC COLLECTION.....9

 B. GPO’S ROLE IN DIGITAL ARCHIVING.....10

 C. SECURITY OF THE COLLECTION.....10

 D. OTHER AGENCIES’ ROLE IN DIGITAL ARCHIVING.....11

 E. NARA’S ROLE IN DIGITAL ARCHIVING.....11

APPENDIX I: DEFINITIONS.....13

APPENDIX II: PLANNING DOCUMENTS REFERENCED IN THIS PAPER16

I. PREFACE

In 1998 the U.S. Government Printing Office (GPO) published the first edition of *Managing the FDLP Electronic Collection: A Policy and Planning Document*, http://www.access.gpo.gov/su_docs/fdlp/pubs/ecplan.html and established the FDLP Electronic Collection in order to provide stable, ongoing access to Government publications in digital formats. The six years since the initial edition have been a period of continuous development and change, both within and beyond GPO as procedures and mechanisms have been developed to manage our digital assets.

This second edition incorporates advances in the theory, technology, and practice of managing digital collections. Much of the revision recognizes the reality of a program in which over 65% of titles are online, and that every title in FDLP will be available in digital format within five years. The FDLP Electronic Collection (EC) is part of GPO's Collection of Last Resort, described at http://www.access.gpo.gov/su_docs/fdlp/pubs/clr.pdf.

II. EXECUTIVE SUMMARY

GPO provides permanent public access to Government electronic publications as a continuation of its historic role in providing permanent access to tangible publications in conjunction with regional depository libraries. The dual roles of the FDLP Electronic Collection are reflected in two collections: the FDLP EC dark archive(s) for preservation (a component of the Collection of Last Resort), and the access collection(s) for public use.

The transition to electronic publishing and dissemination expanded the universe of Government information to which current and permanent access must be provided. While the traditional role of the depository library system to ensure permanent access to tangible products continues, a concomitant responsibility for electronic information has emerged for the Government. As the FDLP grows ever more reliant upon digital information, GPO, as administrator of the FDLP, must ensure the ongoing accessibility of the electronic products that comprise the FDLP EC.

The Government has an obligation to provide permanent public access to its information, and GPO carries out this responsibility for all FDLP information. The mandates of 44 U.S.C. Chapters 19 and 41 establish GPO's responsibility for providing permanent public access to a comprehensive collection of tangible and digital U.S. Government publications.

GPO manages digital objects in the FDLP EC, links users to Federal electronic publications through cataloging and persistent identifiers, ensures authenticity, provides appropriate instruction and support for Collection users, and ensures continued no-fee public access to the entire range of Government information available under the auspices of the FDLP.

The EC consists of preservation copies in dark archives and access copies maintained by GPO or its partners in light archives for the convenience of reference. GPO or its partners will initiate steps, whenever feasible and cost-effective, to migrate the content or refresh the operating software as necessary to make the content readily accessible to a broad spectrum of users.

This second-generation plan defines parameters and requirements for the FDLP EC, and refines the policy framework on which development and maintenance of the Collection are based. In managing the EC, the guiding principle that the public has a right of access to Government information prepared and published at Government expense is the same principle that has guided the FDLP throughout its history.

GPO's permanent public access initiatives support and complement the public information missions of the Congress, NARA, the Library of Congress and the other national libraries, and other Government agencies. Success depends on the participation and cooperation of these and other constituents at various stages of the information life cycle. GPO is leading efforts to include products in the EC, provide metadata and locator services, as well as to facilitate partnerships between agencies and other constituents for data storage, access, and preservation.

III. COLLECTION OVERVIEW

The Federal Depository Library Program (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 includes dark archives for preservation of tangible publications and digital objects as well as access collections for public use. For its first five years, prior to the 2003 agreement with the National Archives and Records Administration (NARA), the FDLP EC was primarily operated as an access collection.

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 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"> • Light archives (minimal use, some active preservation). • Depository library collections (no special preservation)

Table 2. Overview of the FDLP Electronic Collection

EC Contents	EC Copies in Collection of Last Resort	Permanent Public Access to:
<p>Born Digital Objects Created by US Government</p>	<ul style="list-style-type: none"> • Harvested digital preservation masters in a dark archive • CDs/DVDs produced by US Govt. agencies • Print or microfiche copies produced for dark archive backup 	<p>Official, authenticated copies from:</p> <ul style="list-style-type: none"> • <i>GPO Access</i> • Partner sites
<p>Digital Objects Created by GPO or Partners</p>	<p>Scanning process produces digital preservation masters for the dark archive</p>	<p>Copies from:</p> <ul style="list-style-type: none"> • Light archives or other repositories • Depository or other library collections

The mission of the FDLP is to assure current and permanent public access to the universe of information published by the U.S. Government. The purpose of this plan is to articulate the GPO’s responsibilities and practices for the provision of current and permanent public access to eligible electronic Government publications. The plan will:

- Define the components of the EC.
- Outline GPO’s role in providing access via cataloging and metadata.
- Describe criteria and methods for building the EC.
- Provide a functional definition of official information in the EC.
- Describe considerations for preserving the collection

IV. KEY ASSUMPTIONS

Within the mandate of 44 U.S.C., GPO takes responsibility for key aspects of the life cycle management of electronic Government publications for the FDLP. Developing the FDLP EC emphasizes building content, assuring permanent access, and capitalizing on the cooperative strengths of GPO and the FDLP to build the necessary infrastructure for preservation, authentication, identification, access, retrieval, and delivery.

This plan rests on several broad principles, core values, and subsequent assumptions about the FDLP:

1. No-fee access to Government information is a right of the people.

2. The Government has an obligation to provide permanent public access to its information.
3. The mandates of 44 U.S.C. Chapters 19 and 41 establish GPO's responsibility for providing permanent public access to tangible and digital U.S. Government publications.
4. The FDLP includes all Government publications, regardless of format or medium, which are of public interest or educational value, except for those products which are for strictly administrative or operational purposes, classified for reasons of national security, or to which access is limited by legal constraints, such as considerations of individual privacy.
5. Information included in the FDLP EC is public information published by an official source, i.e. the publishing agency or other trusted source.
6. GPO will certify EC digital content with varying levels of authentication dependant upon provenance, chain of custody, and level of quality assurance in the digitization process.
7. A central coordinating authority provides the most complete and cost effective dissemination and locator services.
8. A system of shared responsibility for preserving and providing access to Government information will produce the greatest benefit in return for resources invested.
9. The cost of managing and maintaining the archive infrastructure to provide permanent public access to FDLP electronic Government publications will be borne by the Government and its official partners and not the end user.
10. The *GPO Access* online system, including GPO's content partners, is the principal electronic delivery vehicle for the FDLP.
11. The mix of institutions and users with interests in the Collection is diverse and complex and includes Federal depository libraries and their users, other information consumers, Congress, agency producers of information, information intermediaries of various kinds within and beyond GPO and the Government.
12. Products for the Collection are selected and added according to criteria and priorities derived from various constituencies.
13. GPO's National Bibliography services are the gateway to the FDLP EC. When GPO catalogs electronic publications they are added to the EC.

14. To minimize undue complexity, maintenance, and expense, proprietary client software and other products with copyright-like barriers will be avoided, but, owing to agency decisions beyond GPO's authority, may be included where appropriate.
15. GPO's costs associated with developing and maintaining parts of the EC under GPO's control are generally borne by the Superintendent of Documents Salaries & Expenses appropriation.
16. GPO supports the use of open-system standards, media and formats.

V. DEFINING THE FDLP ELECTRONIC COLLECTION

The FDLP Electronic Collection is a comprehensive collection of remotely accessible and tangible electronic Government publications. The EC includes electronic Federal Government publications that have been created at taxpayer expense and demonstrate public interest or educational value. Publications determined by their issuing agencies to be required for strictly administrative or operational purposes or for official use only, or those classified for reasons of national security, are excluded.

Electronic resources in the FDLP must meet the same basic criteria as traditional publications in the program. According to Title 44 of the U.S. Code, publications must be produced at public expense, have public interest or educational value, not be intended strictly for internal use in the issuing agency, and not be classified for reasons of national security. Information and data stored in and retrieved by means of document or content management systems, dynamic databases, or otherwise not fixed such that a consistent rendering can be returned time and again, will be considered on a case-by-case basis, working as closely as possible with the publishing agency.

GPO does not, however, distribute, catalog, archive, assign persistent identifiers to, or otherwise make accessible information which is out of scope for the FDLP. Publications that have not been declassified or released by authorities for public access are not in scope for the FDLP. Occasionally there are situations in which the persistent identifiers in GPO cataloging records link to content at non-Governmental sites, such as educational institutions. These are cases in which GPO or the publishing agency has an official agreement with the institution that manages the site.

Every attempt will be made to make the FDLP Electronic Collection as inclusive and official as possible, within the limitations of available technology to preserve ongoing access.

The FDLP Electronic Collection has five major components:

- Core legislative and regulatory documents on *GPO* Access, such as the Congressional Record, Federal Register, and others.

- Digital publications published or made available by GPO, within specific agreements for services between GPO and the publishing agency
- Electronic publications published and made available by their publishing agencies, which GPO identifies, describes, and links to at the agency site or from an EC access site.
- Tangible electronic Government publications, such as CD-ROM or DVD-ROM, which GPO distributes to libraries.
- 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.

VI. OFFICIAL STATUS OF EC CONTENT

Information included in the FDLP EC is U.S. Government public information published by official sources. While all FDLP EC content is official information, the level of confidence in individual digital publications can vary. GPO provides EC digital content with varying levels of authentication dependant upon provenance, chain of custody, and level of quality assurance in or type of output from a legacy digitization process.

In order to be certified as authentic EC digital content must be obtained from or its data origin verified by the publishing agency. Typically this will be born digital content for which GPO has been directly involved in the publication process.

The next level of certification will be for content obtained from trusted sources, such as digital publications harvested from publishing agency Web sites or created from source data files used to create print publications. Partner institutions creating digital preservation masters in accordance with accepted program specifications are also trusted sources.

Other EC digital content, for example content derived from print publications distributed through the FDLP, may be accepted from unofficial sources such as institutions creating digital access copies that do not conform to the accepted specifications for preservation masters. Acceptable unofficial sources also include non-Governmental Internet archives from which GPO may obtain a digital access copy. Low confidence access copies thus acquired may be replaced with preservation quality files when an opportunity to do so occurs.

VII. DISCOVERY, ACQUISITION, AND BIBLIOGRAPHIC CONTROL

When GPO catalogs or applies other metadata services to digital publications they become part of the FDLP EC. There are numerous electronic Government publications that are not included in the FDLP EC because GPO has not yet brought them under bibliographic control.

This first-level collection management activity depends upon knowledge that the products exist. Even though GPO is engaged in information discovery and Web harvesting to acquire products for the Collection, this activity is by itself insufficient. In order to include the broadest range of products into the FDLP, and thereby ensure current and permanent access, GPO will employ a range of strategies. These include reliance on notification from and outreach to other agencies and notification from the depository library community. Online electronic products are identified and recommended by GPO or other program stakeholders.

After evaluation of the product, necessary contact is made with the publishing agency, a selection decision made, bibliographic control established, and a copy of the digital object captured for the EC. The harvested digital objects may be stored in GPO's electronic archive or at an FDLP partner site.

GPO catalogs publications in all formats using a variety of national and international standards for bibliographic data, ensuring that the resulting records will provide broad, consistent access. GPO-created bibliographic records form the basis of the *Catalog of U.S. Government Publications* <http://www.access.gpo.gov/catalog> and are used in the local catalogs of libraries nationwide to describe and locate Government publications.

GPO cataloging records for online resources include persistent identifiers, currently a PURL, that enables the user to link directly to the described resource. GPO's persistent identifiers currently link to the resource at the publishing agency site until that resource is no longer available. At that time the PURL resolver table is modified to redirect the user to the archived object on GPO's archive server or partner site. In addition to placing files on archival servers, GPO is creating, storing, and maintaining a limited set of preservation level metadata for all archived files.

VIII. USER COMMUNITY

A key user community for the EC gains access through the facilities and resources of the FDLP, including its geographically dispersed network of depository libraries. However, in the networked environment the public routinely uses the Collection directly, without the depository library as intermediary. GPO will strive to accommodate the needs of as broad a range of users as possible within the constraints of time and resources. Collection planning and the effective use of GPO's appropriated funds will focus on depository libraries and depository users as definable, known groups representing the public's need for access to Government information.

Even though the emphasis of Collection development is toward depository libraries and their users, GPO will strive to accommodate the needs of the broadest possible range of users who possess a wide range of technical capabilities, within the constraints of time and resources. In the context of the FDLP, accessibility includes the degree to which Government information is accurately identified and described bibliographically, the information's availability is made known to the public through the National Bibliography, and technological, social, economic, political, and physical barriers to access are minimized. Publications are made available using World Wide Web or successor technology, in formats that enable use by those who require assistive technologies.

IX. ASSURING ONGOING ACCESS

GPO's strategy for assuring access and integrity for electronic government publications is to direct users of GPO metadata services to content at publishing agency servers for as long as possible. GPO captures and maintains copies of digital publications in a GPO-maintained archive, to be invoked only at the point at which the publication is no longer available at the publishing agency site.

A. SELECTION FOR THE ELECTRONIC COLLECTION

Electronic publications selected for the EC meet the scope criteria for the FDLP, and fall into one of the following categories:

- Core legislative or regulatory publications for which GPO is responsible by statute.
- Publications for which agencies have contracted with GPO for access and storage.
- Publications managed on publishing agency servers (or by their official designees or partners) for which GPO has established links and bibliographic control.
- Publications for which access is managed by GPO partners under specific agreements.
- Publications digitized by GPO or one of its partner institutions.

B. GPO'S ROLE IN DIGITAL ARCHIVING

The EC archive employs a distributed architecture with storage shared among multiple locations. Archival servers are operated by GPO, by GPO partners, and by third parties operating under contractual arrangements. In all cases, archival copies are considered GPO property. Under certain circumstances, such as a GPO partner's inability to

continue to maintain a digital archive, the content will return to direct GPO control.

GPO employs a multi-tiered approach to preserving and providing access to digital content:

- Under the terms of the agreement with NARA designating GPO as an archival affiliate, *GPO Access* content is considered the preservation copy.
- Born digital publications on agency servers may be declared by the publishing agency to be permanent. This must be documented with a written agreement that includes failsafe measures.
- Content managed by a GPO partner other than the publishing agency (e.g., a depository library) will be documented in written agreements which include failsafe measures.
- Publications represented in the FDLP only in digital format will be archived for permanent public access in the Collection of Last Resort.
- The appropriate means for backing up born digital content by creating and preserving one or more tangible versions is under consideration.

C. SECURITY OF THE COLLECTION

Security of digital publications has many aspects. To gain a level of trust among users, digital files should be:

- Secure from active, malicious alteration
- Secure from inadvertent alteration due to error, mistake, or degradation of media
- Produced, shared, and offered for access in an environment that balances, in its policies and practices, concern with security of systems with concern for access.

The public must be able to be assured that EC content is consistently available, official, and reliable. The Collection of Last Resort, containing preservation copies of digital objects, will meet the highest practicable assurance level described in the *Decision Framework for Federal Documents Repositories* (See Appendix II).

To fully realize the potential of digital media in a networked setting, a fully redundant collection infrastructure, including mechanisms for access, organization, and preservation, must be created and maintained in more than one location.

To achieve a secure environment for access and preservation, security measures

should include:

- Fully functional data stored in a secure offline environment meeting national standards for geographic separation.
- Fully functional redundant systems of servers and other infrastructure, as well as storage, for continuity and disaster recovery.
- Fully documented policies and plans for addressing security concerns.

D. OTHER AGENCIES' ROLE IN DIGITAL ARCHIVING

Another Government agency, typically the publishing agency, may enter into a content partnership to preserve a portion of the FDLP EC. In addition, GPO may enter into an electronic product content partnership to expand the content available to the Federal depository libraries. In either case, as well as in "tripartite" agreements involving GPO, one or more agencies, and a library institution, the basic parameters outlined above must be represented.

E. NARA'S ROLE IN DIGITAL ARCHIVING

GPO's EC complements the strategic goal of the National Archives and Records Administration (NARA) to provide the public with access to the essential evidence of our Government. In general there are important distinctions in what is collected and maintained by NARA and GPO.

With the exception of the content on *GPO Access*, the FDLP Electronic Collection is not comprised of the record copies of electronic products. GPO's principal concern is with the information content of the product, not with the product's value as evidence of the activities of the Government. Inclusion of an agency electronic publication in the Collection is in no way intended to be a substitute for the issuing agency's disposition of that publication to NARA in accordance with a records schedule.

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. NARA intends to maintain electronic records in a format that is independent of specific hardware or software requirements, and requires agencies to transfer such records to NARA in accordance with regulatory specifications that support that independence. It is critical for NARA to maintain the provenance of the records and other contextual information in order to document how the records were used to carry out the functions and activities of the creating entity. This contextual information enables the records to provide evidence and accountability, and must be preserved along with the content of the Government publications that are archival records.

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 FDLP content.

APPENDIX II: PLANNING DOCUMENTS REFERENCED IN THIS PAPER

Collection of Last Resort, Revised Draft June 18, 2004

www.gpoaccess.gov/about/reports/clar0604draft.pdf

Decision Framework for Federal Document Repositories, Discussion Draft, April 12, 2004

www.access.gpo.gov/su_docs/fdlp/pubs/decisionmatrix.pdf

The National Bibliography of U.S. Government Information: Initial Planning Statement,

June 18, 2004 www.gpoaccess.gov/about/reports/natbib0604.pdf

DRAFT