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

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

June 18, 2004

This document is located on *GPO Access* at
http://www.gpoaccess.gov/about/reports/natbib0604.pdf

Comments on this document may be sent to Gil Baldwin, Director, Office of National Bibliography Program Planning, at ebaldwin@gpo.gov.

Comment period ends July 30, 2004

CONTENTS

I. PREFACE

By law[1] the U.S. Government Printing Office is charged with preparing and publishing a "comprehensive index of public documents," including "every document issued or published … not confidential in character."

Over time GPO's Cataloging and Indexing Program has become a catalog of the publications distributed in the Federal Depository Library Program (FDLP). This erosion of comprehensiveness has resulted from several factors, including fugitive documents not available to GPO, the rise in digital publishing, and ongoing constraints on program fiscal, information technology, and human resources.

GPO intends to develop a comprehensive and authoritative National Bibliography of U.S. Government Publications that meets the full extent of the statutory requirements of 44 U.S.C. 1710-1711.

This document, GPO's initial planning statement on the National Bibliography, will examine ways to improve the comprehensiveness of our cataloging and metadata programs, to enhance the usability of GPO's bibliographic products and services, to operate a standards-based national library quality program, to demonstrate GPO leadership in the bibliographic services arena, and to enhance the public's ability to identify and locate U.S. Government publications.

II. BACKGROUND

Typically a national bibliography covers works published in or about a specific country and generally is compiled and published by a country's national library. GPO's National Bibliography of U.S. Government Publications is narrower in scope than a full national bibliography, in that its focus is entirely upon U.S. Government unclassified publications in all formats.

Many national bibliographies attempt to document a nation's published heritage, and in doing so makes that heritage known and accessible to present and future generations. This type of national bibliography lists and describes a wide variety of publications produced in a given country, or published elsewhere but of special interest or significance to that country. The country's deposit law may form the most important acquisitions channel for the national bibliography. A full national bibliography may include books, periodicals, sound recordings, microforms, music scores, pamphlets, government documents, theses, educational kits, video recordings and electronic documents. It normally provides standard cataloging information for each item listed, and may include information on forthcoming publications to permit advance ordering of new products.

Historically, in the United States, the Library of Congress (LC) was responsible for the traditional national bibliography. LC compiled cataloging records for all types of works

---

[1] See Appendix 1.

from contributing libraries, ultimately resulting in the production of the *National Union Catalog* (NUC), which ceased publication in the 1980's. Today the nearest approximation of a national bibliography for the United States is contained in OCLC's WorldCat online catalog. WorldCat is a worldwide union catalog created and maintained collectively by more than 45,000 member libraries in 84 countries. With millions of online records built from the bibliographic and ownership information of contributing libraries, it is the largest and most comprehensive database of its kind. WorldCat contains a significant number of non-U.S. imprints, and is therefore broader in coverage than a national bibliography.

GPO has contributed in excess of 337,000 records for U.S. Federal publications to WorldCat since 1976, and other OCLC member libraries have contributed many additional thousands, resulting in a de facto national bibliography for U.S. Government publications being embedded in WorldCat. However extensive WorldCat is as a resource, there are several concerns with it. Among these are lack of visibility of Government information to users, inability to focus only on U.S. Government publications, multiple records representing the same resource, a general lack of holding or location information from depository libraries for Government publications, variations in cataloging level and application that have occurred over time, and the lack of unrestricted, no-fee public access to WorldCat.

III. WHAT IS THE NATIONAL BIBLIOGRAPHY OF U.S. GOVERNMENT PUBLICATIONS?

GPO's National Bibliography of U.S. Government Publications will be a comprehensive catalog containing descriptions and locations of U.S. Government unclassified publications in all formats.

It is GPO's goal to expand its cataloging initiatives to comply fully with statutory requirements, to increase the visibility and use of Government information products, and to develop a premier destination for information searchers.

This goal is consistent with, and enhances the ability to deliver on, GPO's strategic direction of "Keeping America Informed." It is also consistent with the more general Government goal to improve public-facing services to facilitate public interaction with the Government.

Functionally, the National Bibliography will be the principal application of the Integrated Library System (ILS), once the ILS is procured and implemented. GPO will develop one or more relational databases of library cataloging and other metadata records that describe and link to U.S. Government publications in a variety of formats and locations. The National Bibliography will become a premier discovery tool for users searching for U.S. Government publications, and will utilize metadata resources created by GPO, by other agencies, and by partner institutions. The ILS' metasearch function will provide the capability to search across a distributed set of non-homogeneous metadata resources

from other institutions and display a combined result set. The public face of the ILS will be GPO's Online Public Access Catalog.

Initially GPO's core National Bibliography database will consist of GPO-produced bibliographic records representing titles cataloged from 1976 to the present. Once the ILS initial implementation is completed, GPO plans to expand the National Bibliography database by adding pre-1976 records, records created by cooperative cataloging partners and other institutions, records converted from other bibliographic or metadata formats, etc. Building the National Bibliography legacy database is expected to be a multiyear project; adding records for new publications will be ongoing.

IV. SCOPE OF THE NATIONAL BIBLIOGRAPHY PROGRAM

In order to facilitate public discovery of the body of U.S. Government public documents the scope of the National Bibliography must comply with the full extent of the statutory requirements. The legal definition of the cataloging and indexing program, codified at 44 U.S.C. 1710-1711, does not specially address digital publications. However, the changes in the Government information environment require that the National Bibliography address digital publications. GPO's cataloging authority in the electronic arena is derived from 44 U.S.C. 4101(a), which directs the Superintendent of Documents to maintain an electronic directory of Federal electronic information.

GPO considers that information in scope for the National Bibliography is any information product, regardless of form or format, that any U.S. Government agency discloses, publishes, disseminates, or makes available to the public, as well as information produced for administrative or operational purposes that is of public interest or educational value. Publications represented in the National Bibliography will be those acquired from official sources or sites, and not subject to official use or security classification restrictions.

The National Bibliography will cover an expanded range of current and future Federal publications. In addition, the National Bibliography will also expand in historic depth, as GPO acts to include bibliographic data for legacy U.S. Government publications.

V. KEY ASSUMPTIONS

1. The Integrated Library System enables GPO to build the National Bibliography. A 21st century National Bibliography can not be operated on GPO's legacy information technology (IT) platforms.

2. GPO will continue its OCLC participation.

3. GPO's core National Bibliography database format will be AACR2 cataloging records in MARC21 format.

4. Other bibliographic formats may be ILS/National Bibliography system inputs or outputs. Possible system input or output formats, requiring crosswalks include:

   a) MARC
   b) ONIX
   c) GILS
   d) Preservation metadata
   e) Dublin Core
   f) COSATI

5. GPO will apply different levels of cataloging to different types of resources, following applicable standards. However, the default level of cataloging applied to U.S. Government publications, regardless of format, will be equivalent to OCLC "Full" level, when none of the alternatives is applicable.

6. National Bibliography cataloging records will come from multiple sources, including:

   a) Original cataloging
   b) Record imports
   c) Conversion of pre-1976 GPO bibliographic records
   d) COSATI-MARC conversions
   e) Harvested metadata
   f) Metadata from printing contractors
   g) Records contributed by partners
   h) Preservation metadata from archiving projects

7. The National Bibliography will provide public access and location information for all items in the Collection of Last Resort (GPO's distributed archival collections of tangible and digital publications).

8. GPO cataloging policies and procedures will be documented and communicated proactively.

9. GPO will continue its participation in the LC Program for Cooperative Cataloging (PCC) Programs. These include:

   a) NACO (name authorities)
   b) SACO (subject authorities)
   c) CONSER (serial cataloging)

VI. PLANNING ISSUES FOR INVESTIGATION

This section enumerates issues that GPO will investigate, document, and develop policies and procedures to cover.

1. The ILS is should produce outputs programmatically, with minimal reworking or modification.

2. GPO is considering changing its cataloging practice for archived documents to set the persistent link to the archived version immediately in order to reduce the necessity for constant verification, research, and re-establishing links when the agency version is no longer available.

3. GPO will investigate the benefits and costs of producing additional or more detailed analytics.

4. GPO's application of the CONSER "single record option" varies with circumstances[2]. GPO will investigate the benefits and costs of applying this option more consistently.

5. GPO will investigate the benefits and costs of GPO participating in standards-setting bodies such as NISO or MARBI.

6. GPO is investigating replacing its system of persistent identifiers. It is expected that Handles will replace PURLs. Some issues that will be examined include to what extent are these solutions backwardly compatible with GPO's installed base of nearly 50,000 PURLs, and whether GPO should continue to assign persistent identifiers to digital resources beyond its direct control.

7. GPO is investigating the application of industry-standard unique identifiers, such as ISBNs or ISSNs for all publications with National Bibliography records.

8. GPO is monitoring developments in the cataloging and metadata arenas in anticipation of potential impacts on the National Bibliography program. Some of the developments being monitored include:

    a) FRBR.
    b) METS.
    c) MODS.
    d) PREMIS.

9. GPO is seeking the most efficient and cost-effective strategy for performing the cataloging work for the National Bibliography, including:

    a) GPO staff.
    b) Cooperative partners.
    c) Agencies.

---

[2] GPO practices are described in *Format determination in GPO Cataloging Practices*, April, 2004, http://www.access.gpo.gov/su_docs/fdlp/cip/cataloging-formats.pdf

d) Contractors.

10. GPO is investigating various means of providing bibliographic records to various users, including:

    a) Libraries.
    b) Library of Congress.
    c) Bibliographic record vendors.
    d) OCLC and other bibliographic utlities.
    e) Book industry partners.

VII. RELATED ISSUES

1. Determine how to best provide bibliographic access to documents, and full text in the case of electronic documents, including documents created "on the fly" from dynamic data bases.

2. Need to plan enhanced OPAC content design, including the possibility of providing specialized services to International Exchange Program libraries.

3. GPO will develop a detailed Retrospective Cataloging Action Plan.

4. Classification systems:

    a) Appropriate use of the Superintendent of Documents (SuDocs) Classification system, such as application to non-depository publications.
    b) Determine the role of SuDocs Classification numbers not assigned by GPO.
    c) Determine if SuDocs class numbers should be assigned to intangible resources and by what method?
    d) Determine appropriate application of other classification systems, including LC Classification or Dewey.

5. GPO will investigate Item number issues, including:

    a) Whether item numbers ideally have a one-to-one relationship with class stems or individual titles.
    b) Whether the item number system has utility in the digital publications environment.
    c) What alternative methods or data elements GPO and/or bibliographic record vendors might use to create library selection profiles.

APPENDIX I: LEGAL BASIS FOR GPO'S CATALOGING PROGRAM

44 USC 1710. Index of documents: number and distribution

The Superintendent of Documents, at the close of each regular session of Congress, shall prepare and publish a comprehensive index of public documents, upon a plan approved by the Joint Committee on Printing. The Public Printer shall, immediately upon its publication, deliver to him a copy of every document printed by the Government Printing Office. The head of each executive department, independent agency and establishment of the Government shall deliver to him a copy of every document issued or published by the department, bureau, or office not confidential in character. He shall also prepare and print in one volume a consolidated index of Congressional documents, and shall index single volumes of documents as the Joint Committee on Printing directs. Two thousand copies each of the comprehensive index and of the consolidated index shall be printed and bound in addition to the usual number, two hundred for the Senate, eight hundred for the House of Representatives and one thousand for distribution by the Superintendent of Documents.

44 USC 1711. Catalog of Government publications

On the first day of each month the Superintendent of Documents shall prepare a catalog of Government publications which shall show the documents printed during the preceding month, where obtainable, and the price. Two thousand copies of the catalog shall be printed in pamphlet form for distribution.

There is additional authority for a cataloging operation under Chapter 41--Access to Federal Electronic Information:

44 USC 4101. Electronic directory; online access to publications; electronic storage facility

  (a)   In General.--The Superintendent of Documents, under the direction of the Public Printer, shall--
          (1) maintain an electronic directory of Federal electronic information;

Appendix II: References

British National Bibliography (BNB) description.
http://www.bl.uk/services/bibliographic/natbib.html

Canadiana: The National Bibliography of Canada.
http://www.collectionscanada.ca/canadiana/index-e.html

Defining What Government Information is to be Categorized: Statement of Requirements. (GPO, 2004)
(http://www.gpoaccess.gov/cgiwg/pdf/cgiwgroup/revMay2004.pdf

Functional Requirements for Bibliographic Records: Final Report. (International Federation of Library Associations, Cataloging Section, FRBR Review Group, 2003)
http://www.ifla.org/VII/s13/wgfrbr/finalreport.htm

[MARC Formats and Standards Description]. (MARC Standards Office, Library of Congress, 2003) http://www.loc.gov/marc/

Metadata Object Description Schema Official Web Site. (Library of Congress, 2004)
http://www.loc.gov/standards/mods/

METS: An Overview & Tutorial (Library of Congress, 2003)
http://www.loc.gov/standards/mets/METSOverview.html

Template for Standard NAL Metadata. (NAL Metadata Task Force, 2001)
http://www.nal.usda.gov/cataloging/TEMPLATE2.pdf

United States Code, 2000 edition. Title 44, Public Printing and Documents, Chapter 17, Section 1710, Index of documents: number and distribution,
http://www.gpoaccess.gov/uscode/index.html

United States Code, 2000 edition. Title 44, Public Printing and Documents, Chapter 19, Section 1711, Catalog of Government publications,
http://www.gpoaccess.gov/uscode/index.html

United States Code, 2000 edition. Title 44, Public Printing and Documents, Chapter 19, Section 1901, Definition of Government publication, available at
http://www.gpoaccess.gov/uscode/index.html

United States Code, 2000 edition. Title 44, Public Printing and Documents, Chapter 19, Section 1902, Availability of Government publications …, available at
http://www.gpoaccess.gov/uscode/index.html

APPENDIX III: CATALOGING ACRONYMS USED IN THIS PAPER

- AACR2 – Anglo-American Cataloging Rules, 2nd edition.
- CDS – Cataloging Distribution Service, Library of Congress
- CONSER – Cooperative program for serial cataloging.
- COSATI – Cataloging format used by scientific and technical agencies.
- DC, or Dublin Core – Metadata format developed by OCLC.
- FDLP - Federal Depository Library Program.
- FLICC – Federal Library and Information Center Committee, LC.
- FRBR - Functional Requirements for Bibliographic Records.
- GILS – Government Information Locator Systems.
- IT - Information technology.
- ILS - Integrated Library System.
- ISBNs – International Standard Book Numbers.
- LC - Library of Congress.
- MARBI - Machine-Readable Bibliographic Information Committee of ALA.
- MARC – Machine-Readable Cataloging, a data encoding structure.
- METS - Metadata Encoding and Transmission Standard.
- MODS - Metadata Object Description Schema.
- NACO – Cooperative program to establish name authorities.
- NISO - National Information Standards Organization.
- NUC - *National Union Catalog*.
- OCLC – Online Computer Library Center, Dublin, OH.
- ONIX – Online Information Exchange; book industry cataloging format.
- OPAC - Online Public Access Catalog.
- PCC - Program for Cooperative Cataloging, LC.
- PREMIS - PREservation Metadata: Implementation Strategies, an OCLC working group.
- PURL – Persistent Uniform Resource Locator.
- SACO – Cooperative program to establish subject authorities.
- SuDocs - Superintendent of Documents, GPO.
- U.S.C. – *United States Code*.

APPENDIX IV: PLANNING DOCUMENTS RELATED TO THIS PAPER

Collection of Last Resort, Revised Draft June 18, 2004
www.gpoaccess.gov/about/reports/clr0604draft.pdf

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

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