



VERSION CONTROL

U.S. Government Printing Office
Office of Information Dissemination
Program Development Service

Washington, D.C.

June 21, 2005

Contents

I.	Preface.....	3
II.	Overview.....	3
	A. Key Definitions.....	3
III.	Scope.....	3
IV.	Key Assumptions.....	4
V.	Current State.....	5
VI.	Key Issues.....	5
	A. Version Definition.....	5
	B. Version Control.....	5
	C. Version Triggers.....	6
	D. Version Detection.....	7
	E. Version Identifiers.....	7
	F. Version Crosswalks.....	8
	G. Depository Library Responsibilities for Superseded Versions.....	8
VII.	Conclusion.....	8
VIII.	Resources.....	9
	A. List of Acronyms.....	10
	B. Cataloging Rules and Standards.....	11

I. PREFACE

The mission of the Government Printing Office (GPO) is to provide comprehensive, timely, permanent public access to official U.S. Government publications. This includes providing access to and information about tangible publications (e.g. printed documents, maps, CDs, or DVDs) as well as intangible (electronic) publications produced by or for the United States Government. In keeping with this mission, GPO has identified the need to develop concrete policies on the management of different versions of electronic publications. Therefore, this white paper has been developed, focusing on the issues and assumptions surrounding version control of electronic publications disseminated via the Federal Depository Library Program (FDLP).

II. OVERVIEW

GPO has a rich historical tradition of ensuring that citizens have timely access to current and historical Government publications. In keeping with this tradition, version control of electronic Government publications is a key component of GPO's future Digital Content System. Version control refers to the process of evaluating electronic publications against a set of standards that define GPO's version control policy. The purpose of version control is to analyze a publication and assign an appropriate version designation. It will be necessary to evaluate electronic publications throughout their entire lifecycles in order to provide accurate and relevant information to end-users via metadata describing that publication.

A. Key Definitions

The following definitions will be applied to the terms below throughout this paper.

- Version – Unique manifestation of a publication.
- Version Control – Process of evaluating electronic publications against a set of standards that define GPO's version control policy.
- Version Trigger – Changes beyond an agreed upon threshold or tolerance that constitute a new version.
- Version Detection – Activity of inspecting a publication for changes and responding to a version trigger.
- Version Identifier – Information stored in the metadata about a publication that identifies the version of the publication.
- Version Crosswalk – Information stored in the metadata about a publication that provides permanent links to all other versions and metadata about the other versions.

III. SCOPE

Policies, procedures, and guidelines put forth by GPO on version control will apply to all publications that are deemed to be within the scope of the FDLP, with a particular emphasis placed on publications that are disseminated electronically. This document will

not address version control issues related to tangible publications or documents that have not been approved by Federal publishing agencies for dissemination to the general public.

IV. KEY ASSUMPTIONS

1. GPO or other publishing agencies are able to change electronic publications without notification, detection, or identification to end-users.
2. GPO will acquire, authenticate, catalog, preserve, and provide access to all published versions of official and authentic electronic Government publications that have been approved for dissemination to the public and are within scope of the National Bibliography and the FDLP.
3. GPO needs to address the issue of version control in conjunction with authentication, harvesting digital content, and permanent public access.
4. GPO should work in conjunction with agency publishers on the issue of version control. For electronic publications that are marked "Draft" or that are marked "For Official Use Only" or "For Internal Use Only," GPO will coordinate with agency publishers on the issues of cataloging and access.
5. A set of rules should be developed based upon criteria that define what constitutes a new version of an electronic publication. The rules should be flexible and provide guidelines on when a digital version becomes a distinctly different entity.
6. Policies will be developed to address version control for different types of electronic resources (e.g. monographs, corrections or errata or updates, serials, series, databases, Web sites, video, audio, data streams, and multimedia).
7. While GPO's version control system will be automated, subjective evaluation and/or interpretation will be necessary at various points during the version control process, including but not limited to version detection and trigger activation.
8. Version information will be stored in the metadata associated with a given publication. The metadata will include information about the version of the publication being described and alternate versions of the publication.
9. End-users want to be able to track the history and/or changes that have occurred in Government publications. This information will be available in the digital signature or other future means of authentication for a publication.
10. GPO may explore value added products that augment the attributes of a publication (e.g. structure, content, format) for the purpose of distinguishing one version from another.
11. GPO will provide notification of version changes to its users (including depository libraries) through a number of channels that may include push and pull technologies (e.g. notifications lists, RSS feeds).

V. CURRENT STATE

Currently at GPO, version control is combined with GPO's bibliographic control processes. GPO provides bibliographic control by observing a set of rules, standards, and local practices to create and maintain bibliographic records. The standards, rules, terminology, and definitions were originally formulated to address "book format" terminology and the traditional relationships within the book industry with publishers, distributors, and libraries. GPO's current functional requirements for bibliographic control are derived from the Machine-Readable Cataloging (MARC) 21 standard.

The electronic document standards and rules have only recently begun to deal with the nuances of the electronic publications. Even though electronic publications have been around for more than a decade, they still represent new territory for GPO, depository libraries, and the information industry. Currently, GPO applies the Cooperative Program for Serial Cataloging (CONSER) "single record option" as needed for cataloging purposes. GPO has investigated the implications of creating a unique bibliographic record for each manifestation or version of a publication and will begin implementing separate bibliographic records for titles available in multiple formats in conjunction with the implementation of the ILS.

Appendix B contains additional information about GPO's current cataloging rules and standards that apply to version control.

VI. KEY ISSUES

A. Version Definition

In the print publishing industry, a new version of a publication can be equated to a new edition of that publication. In the realm of periodicals and serials, substantive changes to the content are defined as a new issue. According to GPO's Concept of Operations Version 2.0 for the Future Digital System, a version is defined as a unique manifestation of a publication.* In order to move forward with a comprehensive version control policy, GPO will need to fully define what constitutes a "unique manifestation of a publication" across all publication formats (e.g. monograph, serial, etc.). GPO should follow established best practices and standards when developing this policy.

B. Version Control

GPO envisions that the future process of version control will include acquiring, cataloging, storing, preserving, and retrieving different versions of publications. This process may be accomplished by assessing various document attributes (e.g. structure, content, format, etc.), creating metadata about these attributes, monitoring changes to the attributes, updating the metadata to indicate changes to the attributes, and creating links to related documents. In order to improve internal efficiencies and streamline workflows,

* According to the ConOps for GPO's Future Digital System Version 2.0, The term manifestation refers to the form given to an expression of a work; a different version.

the future version control process will be automated whenever possible, but subjective evaluation and interpretation may be required at various points through the process.

In order to perform the version control function, GPO should implement a workflow technology solution that will have the capability to identify versions that can be versioned automatically, and subsequently route those versions that cannot be automatically detected to the proper GPO personnel to make the version determination.

The Concept of Operations Version 2.0 for GPO's proposed Future Digital System envisions a scenario under which GPO will work with content originators (e.g. publishing agencies) to obtain version information, whenever possible at the point of ingest. GPO should acquire, authenticate, catalog, preserve, and provide access to all published versions of official and authentic electronic Government publications that have been approved for dissemination to the public and are within scope of the FDLP. In conjunction with this work, it is recommended that GPO and other agencies take proactive steps to make customers better aware of the version of a publication they are using, through the use of both technology and the implementation of guidelines or standards for identifying versions of government publications. The standards setting process should involve the wide array of stakeholders who use government information.

C. Version Triggers

Version triggers are changes to a publication that progress beyond an agreed upon threshold or tolerance that would constitute a new version of that publication. In other words, Version triggers should be the main indicators by which new versions of publications are detected. GPO will need to develop a comprehensive list of version triggers for all types of electronic publications that are made available through the FDLP.

Examples of possible version triggers include:

- Additions or deletions to the content of a publication.
Standards must be developed to determine what threshold will be used to determine what the type of content changes will activate a version trigger. For example, a simple spelling correction may not be a version trigger, but an addition or deletion of an entire sentence may be one.
- "Last updated" date changes on html pages.
- Cues from publishing agencies, whether submitted directly to GPO or discovered by GPO
- Change in publication title
- Expected dates of new issues of upcoming serials
- Changes in file formats
- Changes in metadata
- Change in effective date of publication
- Change in expiration dates of old data

D. Version Detection

The issue of identifying and detecting various versions of electronic publications represents a highly complex situation for version and bibliographic control. GPO's future version control process must deal with ambiguities routinely encountered for various kinds of documents and publications. These ambiguities make it difficult to detect and determine if a publication is a version of an existing publication or a completely new publication for which a new record must be created. For example, an agency publisher makes an electronic serial available each month on their Web site, and it is also issued monthly in paper. The electronic version of the serial is updated frequently throughout the month, and the monthly updates are made available without discrete identification of the change, except that some content has been revised, added or deleted. The monthly issue is then added to the agency publisher's archive as the "official monthly issue."

In this scenario, version control may differ from traditional bibliographic control. The original electronic version issued by the Federal agency on their Web site would not qualify as a serial under the standard definition used by libraries because some of the content within the publication remains the same while other parts are replaced. Instead it would be cataloged as an integrating resource. GPO's future version control system must be able to detect new editions, new issuances, reprints, for all kinds of electronic publications as described in cataloging standards: serials, monographs, multi-part monographs, maps (serials and monographs), and integrating resources. However, version control must differentiate in cases where the original document is overlaid on the Web as a correction versus being overlaid by a new edition or next issue.

GPO should continue to investigate technologies that can provide an automated version detection system that is capable of handling electronic government publications. This system should follow a set of rules based on pre-determined version triggers to activate those triggers and subsequently detect a new version. The version detection mechanism should then notify the entity or system responsible for creating the proper metadata required to manage the acquisition, cataloging, storage, preservation, and retrieval of the publication. While version detection is projected to be a largely automated process, it is reasonable to assume that certain situations may require subjective interpretation and evaluation.

E. Version Identifiers

Version identifiers should be stored in the metadata for a given publication. Functionally, they should identify the version of the publication. Version identifiers should be created at the time the version detection mechanism has activated a version trigger and detected a new version. GPO will provide public access points to current and historical versions so that end users may access any version for needed information.

Version identifiers must be easily read and understood by humans. End users to which publications are disseminated should be able to easily detect which version they are accessing. Information contained in version identifiers should conform to or at the very least should not be in conflict with existing standards being used for metadata created for cataloging, preservation, or any other functions.

Version identifiers will include, at a minimum, the date and time of the publication version. Other possible information contained in version identifiers may include but are not limited to volume or issue numbers, publishing agency, file formats, change in ISBN or ISSN, and various milestone dates (e.g., reaffirmation dates, copyright dates, GPO receipt date).

F. Version Crosswalks

Version crosswalks are essentially a subset of version identifiers. Version crosswalks will be stored in the metadata about a given publication. They will provide links to all other versions and corresponding metadata, if the information is not already contained in a single record. The version crosswalk links will be permanently available and persistent, to ensure that all versions will be accessible to all users permanently regardless of where they may reside. Version crosswalks will eliminate the need to create multiple records for a given publication.

Version crosswalks should also provide persistent and permanent links to other in-scope resources that may relate to the publication in question. For example, an entry for an issue of the *Federal Register* could cross-reference with the corresponding records for parts of the *Code of Federal Regulations*. In addition, the record for the print version of the U.S. Budget would cross-reference with records for the online electronic Access version and the CD-ROM version.

As is the case for all version identifiers, version crosswalks should be easily detectable and readable by both humans and machines.

G. Depository Library Responsibilities for Superseded Versions

In the online environment where new versions of online publications supersede older online versions, depositories should continue to consider “FDLP Guidelines for Determining Superseded Materials”. However, depositories may apply their local best practices regarding identification of new and older versions of publications. Depositories where online publications are cataloged may, for example, add public supersession notes in the OPAC display of appropriate bibliographic records. They may also choose to remove bibliographic records representing older versions from their catalogs if they determine that this practice best suits the Federal government information needs of their community. GPO will retain all bibliographic records as a part of the National Bibliography of U.S. Government Publications. For more information on superseded items, please consult the Revised Federal Depository Library Manual.

VII. CONCLUSION

This document describes the definitions, scope, assumptions, current state, and key issues relating to version control of U.S. Government publications that fall within the scope of the FDLP and the National Bibliography, and will serve as the starting point in the GPO-wide policy-making process for Version Control. Feedback has been obtained from the Federal Depository Library community at the Depository Library Council (DLC) Meeting held in Albuquerque, NM. The next phase of this process will be to solicit

DRAFT

further feedback and comments from various user communities on version control policy decisions.

VIII. RESOURCES

Draft Access and Preservation Plan for U.S. Government Publications, December 10, 2004.

FDLP Guidelines for Determining Superseded Materials
http://www.access.gpo.gov/su_docs/fdlp/coll-dev/supersede.html

GPO's Future Digital System Concept of Operations, May 2005,
http://www.gpo.gov/projects/pdfs/FDsys_ConOps_v2.0.pdf

Superseded List, 2002,
http://www.access.gpo.gov/su_docs/fdlp/pubs/suplist/index.html

Federal Depository Library Manual, Revised
Not yet published

Appendices

A. List of Acronyms

- AACR2 – Anglo-American Cataloging Rules, 2nd edition.
- BIBCO - Bibliographic Record Component Program of the Program for Corporative Cataloging.
- CONSER – Cooperative program for serial cataloging.
- COSATI – Council on Scientific and Technical Information.
- FDLP - Federal Depository Library Program.
- FLICC – Federal Library and Information Center Committee, LC.
- FRBR - Functional Requirements for Bibliographic Records.
- IFLA - The International Federation of Library Associations and Institutions.
- LC - Library of Congress.
- MARBI - Machine-Readable Bibliographic Information Committee of ALA.
- MARC – Machine-Readable Cataloging, a data encoding structure.
- NACO – Cooperative program to establish name authorities.
- NISO - National Information Standards Organization.
- OCLC – Online Computer Library Center, Dublin, OH.
- PCC - Program for Cooperative Cataloging, LC.
- PURL – Persistent Uniform Resource Locator.
- SACO – Cooperative program to establish subject authorities.

B. Cataloging Rules and Standards

In order to effectively provide version and bibliographic control over Government publications, GPO adheres to national and international standards for descriptive and subject cataloging. These standards include:

- Anglo-American Cataloguing Rules, Second Editions (2002 revisions and 2004 updates)
- BIBCO Participants' Manual
- Bibliographic Formats and Standards, Third Edition
- Cartographic Materials, Cartographic Materials: A Manual of Interpretation for AACR2 (AACCCM) 2nd ed. (2002 Revision)
- CONSER Cataloging Manual, 2002 Edition (Update 1, Spring 2004)
- CONSER Editing Guide, 1994 Edition (Update 15, 2003)
- Descriptive Cataloging Manual, Z1. Name and Series Authority Records
- Government Printing Office Cataloging Guidelines, 4th ed. 2002
- Library of Congress Rule Interpretations, Second Edition (1989)
- MARC 21 Format for Authority Data, 1999 Edition (Update No. 4, October 2003)
- MARC 21 Format for Bibliographic Data, 1999 Edition (Update No. 4, October 2003)
- NACO Participants' Manual, Second Edition
- SACO Participants' Manual Subject Cataloging Manual: Subject Headings, 5th Edition, with February 2004 Update