



# **Versioning: The Future of Managing Changes in Digital Content at the U.S. Government Printing Office**

## **SETTING THE STAGE**

Digital content may change on a very frequent basis, and it is easy to transform content without identification or notification to users. Users want to be certain that they are using the version of information that meets their needs and to be able to track the history of changes that may have occurred. In the case of Federal information, multiple versions of Government publications are often publicly available on Web sites, which can be confusing and potentially damaging to users who are not aware of the version status of the document. Version tracking is a necessary operation in the management and dissemination of Government information to ensure that citizens are accessing the appropriate documents. The following facts are presented to assist in the scheduled discussion.

Version Control is a critical function of GPO's planned Future Digital System (FDSys). As outlined in the FDSys Concept of Operations document, the Version Control function will establish the version of a piece of content and subsequently track it through its entire life cycle, based on best practices.

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.

Electronic document standards and rules have only recently begun to deal with the nuances of the electronic publications. 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.

It is easier to control versions in the print environment, because a new version is generated when an agency sends an updated publication to GPO to be printed. GPO has the responsibility to preserve information for permanent public access in the digital environment, but agencies are constantly changing information without officially issuing new versions. In many cases,

documents are changed and re-posted with the same file name on Web sites, and users do not know that they are looking at a different version. GPO's commitment to version control will ensure that users have permanent public access to the appropriate versions of official Government publications.

## **NEW INFORMATION**

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). GPO will follow established best practices and standards when developing this policy.

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

Version triggers are changes to a publication that progress beyond an agreed upon threshold or tolerance. 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.

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.

Version identifiers distinguish between versions and are stored in the metadata for a given publication. Version identifiers should be created when a new version is detected. GPO will provide public access points to current and historical versions so that users may access any version for needed information. Version crosswalks will be stored in the metadata about a given publication, providing links to other versions and corresponding metadata of that publication. These links will be permanently available and persistent, to ensure users have access regardless of where they reside. Version crosswalks should also provide persistent and permanent links to other in-scope resources that may relate to the publication in question.

In the online environment, where new versions of online publications supersede older online versions, Federal depository libraries 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. 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. GPO intends to maintain and periodically update the *Superseded List* to assist depositories with version identification. GPO will continue to provide public notes regarding supersession where appropriate in its online catalog.

## **ASSUMPTIONS**

- 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.
- Policies will be developed to address version control for different types of electronic resources (e.g. monographs, serials, series, databases, Web sites, video, audio, data streams, and multimedia).
- 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.

## **QUESTIONS**

1. Are the assumptions in this document correct?
2. As GPO works toward the implementation of its strategic vision, are we approaching the issue of Version Control appropriately?
3. What information should be included in policies that pertain to what constitutes a new version?
4. Version triggers are discussed in the Version Control White Paper. Are the outlined version triggers correct? Should other version triggers be included?
5. Should a format change be considered a version trigger?

## **CONTACT**

Selene Dalecky, Development Project Manager  
Program Development Service  
U.S. Government Printing Office (stop: IDPD)  
732 North Capitol St., NW  
Washington, D.C. 20401

Phone: 202-512-0108  
Fax: 202-512-1262  
E-mail: [sdalecky@gpo.gov](mailto:sdalecky@gpo.gov)