

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

Contents:

I. Briefing Topic: Versioning	2
I.1. Setting the Stage	2
I.2. New Information.....	2
II. Revised Assumptions	3
III. Questions to Council, with Council Discussion	4
III.1 Are the assumptions in this document correct?	4
III.2 As GPO works toward the implementation of its strategic vision, are we approaching the issue of version control appropriately?	4
III.3 What information should be included in policies that pertain to what constitutes a new version?	4
III.4 Version triggers are discussed in the Version Control White Paper. Are the outlined version triggers correct? Should other version triggers be included?.....	4
III.5 Should a format change be considered a version trigger?	4
IV. Questions from Council Addressed at the Meeting	5
V. Audience Questions Addressed after the Meeting	5

I. BRIEFING TOPIC: The Future of Managing Changes in Digital Content at the U.S. Government Printing Office

I.1 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.

I.2 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. Functionally, Version identifiers should identify the version of the publication and 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, 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.

II. REVISED ASSUMPTIONS

- II.1. 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.
- II.2. 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).
- II.3. 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.

III. QUESTIONS TO COUNCIL, WITH COUNCIL DISCUSSION

III.1 QUESTION: Are the assumptions in this document correct?

DISCUSSION BY COUNCIL

Council initially skipped this question, and due to time constraints did not revisit it.

III.2 QUESTION: As GPO works towards the implementation of its strategic vision, are we approaching the issue of version control appropriately?

DISCUSSION BY COUNCIL

Council stated that GPO should address the issue of corrections to data, such as changed numbers in tables, and how these are, in a way, another version or corrected version.

Assumption 2 was clarified by adding the language “corrections or errata” to the examples of electronic versions.

III.3 QUESTION: What information should be included in policies that pertain to what constitutes a new version?

DISCUSSION BY COUNCIL

Council agreed that a change in format is considered a version change, for example, a file converted from Word to PDF. Further discussion centered on what constitutes a unique manifestation of a unique publication. One example provided by Council was if periods are added it wouldn't constitute a new manifestation, but other things that would change the stylistic intent involving punctuation might constitute a new version. Council also stated that GPO would need to consider how much the presentation is related to the content.

Council stated that GPO should seek input from other groups such as the GODORT cataloging committee that deals with Metadata issues. Council also asked to review the Versioning white paper after the meeting and respond to the questions and assumptions at a later date. It was also suggested that Peter Hemphill could review the white paper and share his thoughts with the Council.

Mr. Hemphill was recognized as a speaker. He stated GPO should take a look at the rules-based system for determining what constitutes a change because what constitutes a change can change over time. GPO responded that it is the intention of the Future Digital System that it be rules-based and that GPO is in the process of gathering input in order to create rules that can be incorporated into the Future Digital System.

III.4 QUESTION: Version triggers are discussed in the Version Control White Paper. Are the outlined version triggers correct? Should other version triggers be included?

III.5 QUESTION: Should a format change be considered a version trigger?

Due to time constraints, questions four and five were not addressed during the session. Council suggested the topic of Versioning be included as a session at the fall Council meeting.

IV. QUESTIONS FROM COUNCIL ADDRESSED AT THE MEETING

IV.1 QUESTION: Is the issue of corrections to data, such as corrections in tables, addressed in the white paper? If not, it should be.

RESPONSE: The first draft of the white paper was an attempt to pull together the concepts that were discussed in the fall session and it is more of a 50,000-foot view and did not get into the specifics. We do know there are some things that we need to look at, but they weren't really presented in this particular white paper.

IV.2 QUESTION: If you have a change, you're going to have to integrate that version concept depending upon your first push versus whole model. If somebody is out there pulling this information down, they're going to have to go look for it to know that it's changed. The question is, is that going to be acceptable? Do you want as librarians to be notified when this change happens or is it acceptable that if somebody goes out and looks they're guaranteed of getting the latest version?

RESPONSE: The assumption is that GPO would have a unique record for each version, and since GPO would be pushing records, there would be an effected notification and the library could then decide whether they needed the multiple versions or whether they wanted only the most current. Our expectation was that our means of pushing, at least at this point, would be primarily by the bibliographic record.

V. AUDIENCE QUESTIONS ADDRESSED AFTER THE MEETING

V.1 QUESTION: If version 1.2 is authenticated after version 1.3, will the authentication date/time stamp only indicate when authenticated, but the sequential relationship of the versions will be clarified through the crosswalks in metadata?

RESPONSE: Yes, version crosswalks will show the sequential relationship of versions, but the authentication date and time stamp will only indicate when the file was authenticated by GPO.

V.2 QUESTION: How will version trigger thresholds be set to capture minor size yet major substance corrections? E.g., correcting 40 instances of incorrect spelling of "library" may be larger size, yet insignificant compared to a single insertion of the word "not" that reverses the meaning of an entire paragraph.

RESPONSE: It is GPO's assumption that the version control process will not be completely automated. The examples cited in the question are cases in which GPO will invoke a staff member to make a determination of version trigger thresholds. The policies

for this process will be developed in conjunction with the overall policies GPO is developing for Version Control.

V.3 QUESTION: Why is GPO re-defining what the library authorities have already defined to differentiate “editions”, “versions”, etc.?

RESPONSE: The idea behind the development of a Version Control policy within GPO is not to re-define editions or versions as libraries have already defined them, but to ensure that version control is handled using the same rules and procedures across the FDL community. Part of the reason GPO has sessions on such issues is to make sure we are adhering to rules and procedures already developed and tested in the library community.

V.4 QUESTION: I have a friend who, until recently, was an editor for Elsevier. She tells me that if a database user finds an errata sheet for a journal article, the user will be guided back to the original article. But, if the user finds the original first, they are not guided to the errata sheet. How does GPO plan to address this issue?

RESPONSE: These are issues that GPO is aware of internally, and GPO recognizes the need to have a policy in place to handle them. GPO will need to consider whether to imbed the errata pages or make a correction to the file with links. GPO will also need to come up with a policy on how to deal with the fact that documents with changes incorporated no longer represent the typeset page that was officially issued by the publisher. The McCarthy hearings were released with more than 30 pages of errata the day they were issued. GPO released one consolidated PDF file with the errata. If users search GPO Access and download the file that does not include the errata, they will not know those errata were issued. These are issues GPO will need to explore further as a Version Control policy is developed and implemented by GPO.