



Managing the FDLP Electronic Collection A Policy and Planning Document, Second Edition

Contents:

I. Briefing Topic: Managing the FDLP Electronic Collection: A Policy Planning Document	2
I.1 Setting the Stage	2
I.2 New Information.....	2
I.3 Micro Recap.....	2
II. Revised Assumptions	4
III. Questions to Council, with Council Discussion	4
III.1 Does Council agree with the key assumptions in <i>Managing the Electronic FDLP Electronic Collection: A Policy Planning Document, 2nd Edition</i> ?.....	4
III.2 Are there any unforeseen or unintended consequences of varying levels of authentication for EC content?.....	4
III.3 Can digital files created by sources other than the publishing agency be considered “official content”?.....	5
III.4 What does Council advise as the preferred procedure for handling electronic publications withdrawn from the FDLP at an agency’s request?	5
III.5 Should GPO continue its current practice of pointing the PURL to the publication on the agency’s Web site or link directly to the GPO archived version?.....	6
IV. Questions from Council Addressed at the Meeting	6
V. Audience Questions Addressed at the Meeting	6
VI. Audience Questions Addressed after the Meeting	10

I. BRIEFING TOPIC: Managing the FDLP Electronic Collection: A Policy Planning Document

I.1 SETTING THE STAGE

The Government has an obligation to provide permanent public access to its information, and the U.S. Government Printing Office (GPO) carries out this responsibility for all Federal Depository Library Program (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.

The FDLP Electronic Collection is a comprehensive collection of remotely accessible and tangible electronic Government publications. Electronic resources in the FDLP must meet the same basic criteria as traditional publications in the program. According to §§1901-1902, 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.

In 1998 the GPO published the first edition of *Managing the FDLP Electronic Collection: A Policy and Planning Document* and established the FDLP Electronic Collection (EC) 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.

The second edition of the FDLP EC plan 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 81% of titles are online, and that every title in FDLP will be available in digital format within five years. The EC is part of GPO's Collection of Last Resort.

This second-generation plan defines parameters and requirements for the 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.

I.2 NEW INFORMATION

Managing the Electronic FDLP Electronic Collection: A Policy Planning Document, 2nd Edition was developed and released for public comment on June 18, 2004. The original deadline for comments, July 30, was extended to September 7, 2004.

I.3 MICRO RECAP

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.

Information included in the EC is U.S. Government public information published by official sources. While all 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.

Given this overview, GPO seeks your advice and guidance based upon questions surrounding these issues that we have been asking ourselves that need to be considered and evaluated as the agency moves forward on the planning and enhancement of the FDLP EC.

II. REVISED 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:

- II.1 No-fee access to Government information is a right of the people.
- II.2 The Government has an obligation to provide permanent public access to its information.
- II.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.
- II.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.
- II.5 Information included in the FDLP EC is U.S. Government information published by an official source, i.e., the publishing agency or other trusted source.
- II.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.
- II.7 A distributed system for preserving and disseminating digital Government information will produce the greatest benefit in return for resources invested.
- II.8 Managing and maintaining the infrastructure, including hardware, software, and content, to provide permanent public access to FDLP electronic Government publications is an inherently governmental function, and the costs will be borne by the Government and its contractual partners and not the end user.
- II.9 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.

- II.10 When digital publications are described in the National Bibliography, they are considered to be added to the EC.
- II.11 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 unavoidable.
- II.12 GPO's costs associated with developing and maintaining parts of the EC under GPO's control are primarily borne by the Superintendent of Documents Salaries and Expenses appropriation.
- II.13 GPO supports the use of open source standards, media and formats. When necessary GPO may “normalize” digital content to facilitate access or preservation.

III. QUESTIONS TO COUNCIL, WITH COUNCIL DISCUSSION

III.1 QUESTION: Does Council agree with the key assumptions in Managing the Electronic FDL Electronic Collection: A Policy Planning Document, 2nd Edition?

DISCUSSION BY COUNCIL

Council discussion resulted in the Revised Assumptions shown above.

III.2 QUESTION: Are there any unforeseen or unintended consequences of varying levels of authentication for EC content?

DISCUSSION BY COUNCIL

Yes, this raises the question of whether an authenticated publication would be usable in a legal proceeding, or whether the authentication could itself be called into an administrative law review situation.

We can agree that there are other questions, and we don't yet know what they are.

III.3 QUESTION: Can digital files created by sources other than the publishing agency be considered “official content”?

DISCUSSION BY COUNCIL

It seems that depository libraries could not participate in the legacy collection digitization project if we said that only the publishing agency could create them.

Although some agencies do not exist anymore and cannot certify their publications, we still want to include these materials.

If a library acquired an item directly from the agency, do we have to look at a shelf list and try to determine if it is an official Federal publication? How can we certify that what we're digitizing was an official federal document in its tangible form and therefore remains an official document when digitized?

We know that some of the older files that have already been digitized were never official. They were transcripts, annals, or the register of debates. We now consider them official.

III.4 QUESTION: What does Council advise as the preferred procedure for handling electronic publications withdrawn from the FDLP at an agency's request?

DISCUSSION BY COUNCIL

[Tabled and not revisited]

III.5 QUESTION: Should GPO continue its current practice of pointing the PURL to the publication on the agency's Web site or link directly to the GPO archived version?

DISCUSSION BY COUNCIL

If an agency has something that traditionally gets updated over time, we need to be cognizant of versioning problems. Currently, the agency may replace a file with a newer version, while GPO is still pointing to it. GPO could point to the site but also ascertain that GPO has the most current version.

Often there are multiple editions of geospatial data, for instance the national map. As these things change, they will have to be lodged in separate files.

Generally agencies update a document because they want their public to have the most current information. There has to be a way of keeping the archive up to date.

IV. QUESTIONS FROM COUNCIL ADDRESSED AT THE MEETING

[Tabled and not revisited]

V. AUDIENCE QUESTIONS ADDRESSED AT THE MEETING

The facilitator of the Council sessions accepted questions from the audience written on GPO-supplied cards. Eleven of the thirteen questions were answered during the Council session. Those questions and their answers are summarized below. Questions held to answer at a later date, either because of time constraints or the need for a subject matter specialist to provide more detailed answers, follow the questions answered during the session.

V.1 QUESTION: Does pointing the PURL at GPO instead of the agency site affect the official nature of the information? GPO should archive versions and not ignore agency/official locations.

RESPONSE: It's the identical copy. When we find it on an agency web site we currently point to it on the agency web site and harvest that identical document. On the positive side of pointing to the agency web site, the user might browse and look at other things at the site. On the negative side, many of our institutions are spending so much time

verifying those PURLs over and over again and cleaning them up and making adjustments, whereas if they point to GPO that isn't needed. The documents remain in a permanent place.

- V.2 QUESTION:** Regarding the PURL pointing to the GPO server, there was a public comment on this that should be kept in mind, which is that linking to the agency gives the user continual information and allows the user to discover other related stuff at the agency site. It would be cool to include context in the metadata if the PURL goes only to GPO server copy. Something to work on some day?

RESPONSE: That's a very good comment and one that we certainly will take into account.

- V.3 QUESTION:** I wonder if we need to add an explicit assumption that EC is a convenient label during the transition period that eventually with digitization of the legacy collection and with the trend toward a hundred percent of new material in E form it will simply be the collection. This may seem obvious and therefore not necessary but sometimes I think it's helpful to remind ourselves conceptualizing the FDLP collection as tangible versus electronic is artificial.

RESPONSE: That is a good point. We currently have placeholder names for all of these things: the national bibliography, the collection of last resort, the electronic collection, etc. As we rewrite these we need to find nomenclature that we're all comfortable with and that will mean the same thing to everyone.

- V.4 QUESTION:** There are several questions that have specific assumption numbers on them. The first involves assumption number 9.

The assumption is: 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.

The question is: Please clearly articulate official partners so it is understood that Federal Depository Libraries will not be responsible for cost sharing or bearing.

RESPONSE: We were talking about official partners like the University of North Texas, which operates the cyber cemetery and has a contractual relationship with GPO. There is no depository right now that does not bear substantial financial obligations with respect to this program. Every one of these depositories pays for staff, pays for computers, pays for space. Even in the current program there's substantial financial commitment on the part of the depositories. But we have noted the comment and we need to be clear about who we mean by the official partners.

V.5 QUESTION: The next question refers to assumption number 11. The mix of institutions and users with interest 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. What exactly is an information intermediary?

RESPONSE: Librarians are information intermediaries. Information companies that build products are information intermediaries. It is somebody other than the end user facilitating access and interpretation of the information.

V.6 QUESTION: The next question comes from assumption 13. GPO's national bibliography services are a gateway to the FDLP EC. When GPO catalogs electronic publications they are added to the EC. Council needs to discuss the apparent change in number 13 from "the" to "a". This is far more significant change than it seems.

RESPONSE: We're not minimizing the role that the national bibliography will play, but there may be other metadata harvested and blended into other kind of systems. If we do OAI for metadata then many, many people would take that metadata and make it available in other ways that would bring people back to the collection. The national bibliography may be the primary gateway system that we operate, but it's not going to be the exclusive one.

There may be many people who get into this content through a Google search who never go through national bibliography, don't even know it exists. That's the reality of the way things work right now. There are lots of people finding things through Google without going through the OPACs in their institutions.

V.7 QUESTION: Assumption 16: Previously the statement said GPO supports the use of open systems standards, media and format. Do you mean open source or open access or both?

RESPONSE: We changed it to open source, but in fact, as they say in the example of the metadata, we're looking at OAI, which is open access, so both probably is the answer.

V.8 QUESTION: Is the depository library online catalog a major delivery system of government information or replaced by new centralized systems?

RESPONSE: The Monthly Catalog will be superseded by the integrated library system and by the national bibliography. The integrated library system and the ability to search the online public access portion of it will become the database that previously has been the Monthly Catalog database. In fact, it will go beyond records that were originally in the Monthly Catalog as we add retrospective records, and part of the input is digitizing the Monthly Catalog. Right now the Monthly Catalog database doesn't even go back as far as the early 1990s. So the integrated library system will bring in all of the records from the Monthly Catalog and more besides. But it also is independent. Many of the libraries download their own catalog and records for their own holdings, whether electronic or tangible, and develop their own access that way.

V.9 QUESTION: At the session on copyright the speaker emphasized the distinction between government inputs, government HPO and government workers. The first assumption referred to free access to all government information. Does that include work done by contractors and other copyrighted work?

RESPONSE: That may be one we're going to have to follow up with later and see if we can get clarification or answer it. There is legislation pending to make it clearer that government-funded health- and medical-related information remains in the public domain whether it's published in copyrighted journals or otherwise.

V.10 QUESTION: This next question comes from two key assumptions. The first key assumption being number 8: A system of shared responsibility for preserving and providing access to government information will produce the greatest benefit in return for resources invested.

And assumption 15 states: GPO's costs associated with developing and maintaining parts of the EC under GPO's control are generally borne by the Superintendent of Document's salaries and expenses appropriation. What happens if Congress refuses to fund? What is the contingency plan?

RESPONSE: There is absolutely no indication that Congress is not heavily committed to having this information available to the public. At last fall's Depository Library Council meeting, when the Public Printer made the statement about wanting to achieve 30 to 50 million dollars in revenue from the sales program, that was interpreted by some as meaning that there was an intent to charge for GPO Access and to charge depository libraries for content. And practically before we had returned to the office we were hearing from a number of Congressional offices about the fact that if we didn't have enough money we ought to be telling them that.

So while there may be an interest in having this hypothetical conversation, that isn't the sort of thing that's likely to happen precipitously, and it is highly unlikely to happen at all. And if Congress should make a decision at some point in the future that there was no need for a Government Printing Office, that function would be moved somewhere else and funded in that location. So it's not a risk that needs to be addressed for the short term.

V.11 QUESTION: I'm quite concerned that Council members would recommend that a GPO PURL go to the archived version instead of the live version of a publication in its natural environment. The user needs all the options available to him or her to see related publications and see how the report is presented in the agency structure. If the PURL goes only to the archived version this is limiting the user's options. And if that was the case I would not use the GPO Access to search for publications. I would use a search engine that brings me to the original site. The most helpful option would be to include both URLs. This would mean two PURLs. Much greater workload?

RESPONSE: In fact, one of the things we're looking at internally is digital object identifiers which allow you to have multiple locations for the same item. Users are able to find them in multiple locations even as the things move around. The digital object identifiers can track them far better than PURLs. So PURLs are the current technology we're using but maybe not the be all and end all.

VI. AUDIENCE QUESTIONS ADDRESSED AFTER THE MEETING

VI.1 QUESTION: What is the status of an inventory of digitization projects in libraries (and other institutions)? I seem to remember that GPO offered to take a lead in this.

RESPONSE: GPO is working with ALA-GODORT's Government Information Technology Committee (GITCO), to transition the information from the Clearinghouse of Government Documents Digital Projects to GPO to populate a registry of digital projects. This project has begun with discussions between GITCO and GPO on the best way to effect the transfer of information, and coordinate the structure and functions of the registry. The goal it to have the new registry available in late spring/early summer.

VI.1 QUESTION: Explain how copyright information will be handled.

RESPONSE: The general rule is that Government publications in the Electronic Collection are in the public domain, free of copyright restrictions on copying or reuse. In the rare cases where a publication includes an element that is copyrighted, for example, an image, the metadata should indicate that to users.