

Depository Library Council's Advice to the Public Printer January 22, 2004

At the Depository Library Council Meeting in October 2003, Public Printer Bruce R. James challenged the Depository Library Council (DLC) to provide advice on three key issues integral to the envisioning process Mr. James instituted upon his assumption of office in December 2002. Mr. James specifically highlighted these three topics for Council input:

- What constitutes a version of information and how will it be authenticated?
- How should GPO store information, especially in a digital form? Related to that, what should GPO do to ensure permanent public access to the Legacy Collections that have been developed over the course of government printing?
- What types of revenue generation might GPO engage in to further enhance, improve, and develop new technologies that will ensure GPO's future in the electronic milieu?

Over the course of the past several weeks, DLC has engaged in spirited debate over these issues. Additionally, Council sought feedback from the depository library community, library associations, and information specialists outside the purview of the depository library program. Much of this feedback is incorporated into our responses.

Version/Authenticity Control

Context and Problem:

A characteristic of the digital information environment is the ease with which electronic content may be changed. Two major issues that ensue from this malleability of digital information are version and authenticity control. With respect to version control, multiple versions of information are often publicly available on multiple Web sites, which can be confusing and sometimes damaging to information users who are not aware of the version status of the material they are using. With respect to authenticity, multiple copies of government information available on the Web (and on other digital media) allege to be the same version when, in fact, some are authentic "official" versions and others are unauthentic copies. Again, this situation leads to confusion and can be damaging to end-users.

The problems associated with verifying the version and authenticity of digital information products are particularly critical in the genre of government information. Authenticity implies that the internal integrity, origins, and completeness of digital products are verifiable. Users of digital government information must be able to verify the authenticity of digital products. This is especially true for information products having primary legal authority, such as statutes, regulations, opinions, decisions, and

guidelines, as well as other categories such as health advisory bulletins and Census reports, the authenticity of which users will expect to be able to verify. Similarly, users of public data files need to be aware of the date, scope, content and latest modifications of datasets. Historians need to be able to verify versions of documents that have evolved over time.

Council urges GPO to institute practical steps to improve user awareness of the version status and authenticity of government information they are accessing, such as through adherence to best practices and, where available, standards that can identify versions and the authenticity of government information resources. The standards-setting process should involve the wide array of stakeholders who utilize government information and should begin immediately. GPO, as the agency responsible for coordinating access to and dissemination of a substantial volume of digital federal government information, should be among the lead agencies in inter-agency efforts to develop the standards. As a Congressional agency, GPO should take the lead on these issues as they pertain to congressional information.

Further, Council recommends that a method of authentication should be developed and implemented as soon as possible that is widely recognized, flexible, open (e.g., open source and "General Public License," [<http://www.fsf.org/licenses/gpl-faq.html>]) and portable-over-time. Tools for verifying authenticity should be made freely available to federal depository libraries and the public. GPO should acknowledge that all technological means of authentication are subject to human and machine error, and intentional disruption or fabrication. GPO should, therefore, develop policies that reflect secure, long-term methods of ensuring authentic documents in a widespread dissemination and distribution environment such as the FDLP.

Council commends GPO for the vision it has shown regarding the version issue, and for its preliminary exploration of Public Key Infrastructure (PKI) as a mechanism for authentication. Council encourages GPO to continue its efforts to move ahead to adopt strategies for resolving version and authenticity issues in the near-term as standards are developed. GPO should explore multiple approaches to the long-term assurance that government information is authentic, including, but not limited to, dark archives on both government and non-government sites, and widespread distribution of authentic products to the FDLP.

Version Categories and Challenges:

The following list is a core set of categories of government publishing and associated versioning challenges. Guidelines for retention may vary by category. This is intended neither as a comprehensive nor prioritized list, but rather to be illustrative of some of the challenges that will be faced by those attempting to develop best practices and standards in this area:

Monographic Materials

Documents intended as a final product. Are copies of each revised or corrected version needed? Have users already referenced the earlier versions? Policy decisions or research

might have been based on earlier versions. Is there a difference for retention between revised and corrected?

Legislative and Regulatory Documents

Draft vs. final (interim)

Daily and weekly version (interim) – need to ensure that all data is included in annual cumulation rather than aggregated monthly or quarterly

Errata (e.g., hearings)

Process Documents

Each version of the document has value, e.g., versions of bills.

Drafts

How to distinguish which are replaced by final version and which remain valuable as record of the process.

Datasets

Datafiles where information is updated and corrected.

Example: Table with ten years of historic and actual figures and five years of forecast or estimated figures. If estimates are replaced with the actual as time passes, researchers need mechanism for obtaining the original estimates to evaluate accuracy of forecasting methodology.

Date of last update

Freeze on regular basis

Errata

Serials

Date of issue

Revised or original version

Errata

Databases

Freeze on regular basis

Errata

Authentication Challenges and Strategies:

The following list includes a core set of categories of government information authentication challenges with suggested mechanisms that might address these challenges. Many of the concepts in this section have been culled from two key resources:

- Council on Library and Information Resources, *Authenticity in a Digital Environment*, May 2000: <http://www.clir.org/pubs/abstract/pub92abst.html>

- David Bearman and Jennifer Trant, *Authenticity of Digital Resources Towards a Statement of Requirements in the Research Process*, June 1988:
<http://www.dlib.org/dlib/june98/06bearman.html>

Challenge:

Verify the provenance of government information.

Verify the originality (or faithfulness to the original).

Verify the publication is uncorrupted.

Mechanisms:

Mark (watermark) the original publication with document identifiers.

Register document identifiers.

Maintain “collections of record” which contain original or verified copies.

Register “key” data about documents which, “when hashed, or otherwise calculated in a publicly available way, should match that of the document in hand...”

Structure metadata to carry document authentication declarations or proofs.

Digital Information Storage and Legacy Collections

Description of a Legacy Collection

The goal of the FDLDP has been one of assurance of permanent public access with respect to historical and current information as published by federal government agencies. This authority has been codified in *44 U.S.C. Chapter 19*. The FLDP Legacy Collection now includes all government information products, regardless of form or format, which are of public interest, provide educational value (with the exception of material that is produced strictly for administrative or operations purposes), or are classified for reasons of national security, or the use of which is constrained by privacy considerations.

Further, government information in electronic form is now part of the scope of the FDLDP Electronic Collection. This Collection is composed of four major components:

- Core legislative and regulatory materials such as the *Congressional Record*, *Federal Register* and other materials that are found on *GPO Access*.

- Internet products that have been published or made generally available by GPO.
- Internet products made available by the originating federal agency and which GPO has identified and provided links to the information content.
- Products that are distributed by GPO in a tangible digital format.

Therefore the Legacy Collection shall include all tangible materials distributed to the Federal Depositories Libraries throughout the life of the Program and all publicly available federal government information distributed via the Internet. Tangible formats should include but are not limited to all types of print on paper, micro-formats, and digital products such as floppy disks, CD ROMs, and DVDs. The singular “collection” is used to imply the universe of FDLP government information available through the program both online and in hard copy, as opposed to the plural “collections,” which refers to holdings of FDLP items in individual FDLP institutions.

In its deliberations Council discussed strategies for assuring on-going access to both the tangible and Internet-based segments of the Legacy Collection. The following constitute scenarios Council believes merit further examination as plausible options for preserving the Legacy Collection. Council realizes that the models described will entail a reorganization of FDLP relationships, a reassignment of traditional FDLP responsibilities, and the adoption of a key new responsibility – the provision of permanent access to the Legacy Collection. Recent closer ties between GPO and NARA provide the opportunity to explore archival issues pertaining to FDLP materials.

Preservation of Tangible Copies:

A model for assuring ongoing access to the tangible format portion of the legacy collection would be the designation of a new tier of libraries that would acquire, preserve, and provide access to archival FDLP collections. An important assumption of this model is that the general-use copy would be digital. An electronic collection would make legacy government information widely available and save wear-and-tear on the original copies. In concert with GPO maintaining an archival collection of last resort, the new tier of archive collections may be organized along these lines:

1. Regionals and selectives would maintain as many tangible copies as they deem useful for their communities. These would also be “general use” copies, still subject to the access mandates of Title 44. Preservation of these copies would be up to the holding institutions.
2. GPO and the FDLP community would cooperate to establish several geographically dispersed archival FDLP collections that would be as comprehensive as possible and be preserved using archival standards. The archival collections would be accessed only if no copy of a document were available in the general use collections. The intention of establishing multiple archival collections is to avoid potential catastrophic loss – physical or institutional -- inherent in having only a single collection. Problems in one archival

collection would not undermine the whole system. Council notes that attention to potential natural or man-made disasters should be given when designating archival collection sites.

3. GPO would create an archival collection of “last resort.” This collection would be a non-circulating collection and would not be publicly accessible. The collection would be the most comprehensive archival collection, utilizing state-of-the-art standards. The collection would be accessed only when a copy of a document could not be found among the general-use collections or other archival collections. Materials requested from this collection would be digitized and made available through the Electronic FDLP Collection.

GPO and the depository community should investigate previous and current cooperative preservation programs and evaluate the outcomes in order to determine the optimum number of artifact copies needed. This optimum number may differ by content, format and potential use. The community also needs to investigate the number of tangible copies GPO should require from the agencies for all current and future acquisitions to assure permanent public access.

To build the comprehensive Legacy Collection, GPO should distribute its own needs list with regular updates and should closely monitor existing offers lists to acquire items being weeded from depository collections.

Digitization of Print Materials for Improved Access

Beyond preservation of the Legacy Collection, digitization provides the opportunity to improve access, usability and visibility of government information. GPO must coordinate the implementation of user-friendly searching of all digitized government information, a single gateway for access to the collection, and multi-format digitization where appropriate, such as maps and statistical information.

GPO and the community should begin prioritizing and digitizing the tangible materials in the legacy collections. These tangible materials would include print, microfiche, CDs, videos and any other tangible formats not mentioned here. Floppy discs are archived at <<http://www.indiana.edu/~libgpd/mforms/floppy/floppy.html>>.

There should be coordination among digitization projects in order to avoid duplication of effort and to ensure comprehensive coverage. The FDLP should lead this effort. All digitization projects should follow the same standards, perhaps those being developed as a result of the ARL prospectus. Standards for different types of materials should be considered. For example, statistical information could be done in spreadsheet or statistical package-friendly formats instead or in addition to simple images. All depository libraries should be included in the process, and the FDLP should plan for format migrations. Furthermore, GPO should establish an authentication program to

validate digitization projects produced by depository libraries in accordance with the established standards.

Digitized copies should also be preserved or archived, perhaps by using a distributed file system such as LOCKSS. These preservation procedures should be applied not only to materials digitized from the tangible materials in the legacy collections but also “born-digital” materials.

Preservation microfilm copies of legacy print publications should be made if they can feasibly be created as part of the digitization process.

Council advises that a minimum of five artifact copies of originals should be retained.

Cataloging

All materials, tangible and electronic, in the Legacy Collection must be cataloged in accordance with generally accepted cataloging and metadata standards. Complete cataloging of tangible legacy collections will help establish an accurate estimate of the number of existing copies and provide the primary means of access. Materials that have not been cataloged prior to selection for digitization must be cataloged as part of the digitization process.

The GPO should identify pre-1976 catalog records for the Legacy Collection. Such records would reflect the digitized item, and should be made available to all depository libraries at no fee.

Council reaffirms the historic role of GPO as cataloger and indexer of U.S. publications (whether published by GPO or other agencies). GPO should continue to serve as metadata provider for both retrospectively digitized materials as well as prospectively for both print and web-based content.

Council's Suggested Digitization Priorities - unranked

Titles

- Budget of the U.S. Government
- Climatological Data & Local Climatological Data
- Code of Federal Regulations (back to 1938)
- Congressional Record - bound
- Congressional Serial Set
- Current Population Reports
- Department of State Bulletin
- Economic Report of the President
- Ethnology Bulletins

Federal Register
Foreign Relations of the United States
Labor Bureau Bulletins
Public Papers of the Presidents
State Department Bulletins
Statutes At Large
Treaties and Other International Acts
State Department Bulletins
Supreme Court Briefs
U.S. Reports

Type

Agency Annual Reports
Commission Decisions
Children's and Women's Bureau
Census Reports
Congressional Hearings
Environmental Impact Statements - draft and final
Presidential Commission Reports

Revenue Generation

Given the declining sales revenue resulting from no-fee Web-based access to GPO databases, the Public Printer requested the DLC outline principles Council believes should not be compromised by future GPO initiatives intended to generate GPO revenue streams. The Public Printer believes that a revenue stream beyond Congressional appropriations will be necessary to assure that GPO will be fiscally able to adopt new information technologies and improve GPO products and services. Noting that in the print-on-paper environment the GPO sales program provided a viable source of cost-recovery revenue, Mr. James launched the discussion with speculation that a return to the old print-on paper model, which provided no-fee access in depository libraries, fee-based elsewhere, could provide the fiscal resources GPO requires.

Debate was particularly spirited on establishing parameters for revenue generation. Council spent a great amount of time and energy reviewing the information received during and shortly after the fall Council meeting. Council discussed various specific operational suggestions GPO might consider to generate revenue – many of which were raised during the open question-and-answer period at the Council session -- but concluded that the best advice at this time should concentrate on principles rather than on operations.

Council appreciates the need for GPO to upgrade facilities and equipment, to introduce new technologies into the information dissemination process, and the desire to operate in a fiscally responsible manner. Council also understands the significant historical role GPO has played in the development of government information dissemination practices including the cataloging, storage, and maintenance of that information, and encourages careful consideration of the balance between principle and operations as the organization moves to transform itself. Council appreciates the many challenges the Public Printer and GPO face in the coming months and years.

Government information is a public good which should always be freely and widely available to the citizens of the United States. As stated in *OMB Circular A-130* "Government information is a valuable national resource. It provides the public with knowledge of the government, society, and economy--past, present, and future...The free flow of information between the government and the public is essential to a democratic society." (1) The GPO and FDLP possess critical roles in maintaining this national resource.

As Council reaffirmed recently in the [*Envisioning the Future of Federal Government Information*](#), "Government information is a strategic national resource owned by the people and held in trust jointly, for the public good, by GPO and by federal depository libraries. Together these institutions provide stewardship for government information throughout its life cycle, ensuring timely access as new information is produced and permanent public access in the future." (2)

The fundamental purpose of the FDLP is to assure that citizens of the United States have no-fee access to information by and about their government. As information technologies have evolved over the past two decades, institutions involved in information production and dissemination have necessarily changed to adjust to the demands and requirements--and opportunities--of digital information. To date the FDLP has done well in transitioning to the digital era; Council encourages FDLP leadership to continue taking advantage of opportunities presented by digital information technologies to advance to the greatest extent possible the FDLP's underlying principle--informing the nation.

Council believes that in the current environment of web-based information - an environment in which libraries take every opportunity to make digital information available to their clientele at their desktops - the FDLP best serves the government information needs of its constituencies by extending no-fee access to web-based government information beyond library walls. In this way the FDLP should make every effort to optimize recent library trends providing for remote access to digital information.

Council strongly believes that it is the responsibility of our elected and appointed officials to ensure broad no-fee access to federal government information. According to GPO's 1996 report on the successful transition to a more electronic depository system: "The Government should not only allow public participation in the democratic process by providing access to its information, but should encourage public participation and use of

Government information through proactive dissemination efforts that ensure timely and equitable public access." (3) Council believes such proactive dissemination is best achieved by continuing the no-fee basis for access to all government information, regardless of form or format, to all citizens of the United States.

At a minimum and in order to ensure that the public has no-fee access to authentic government information via federal depository libraries as well as other traditional and electronic means, Council recommends that GPO seek full funding to support the FDLP from Congress. Council further recommends that the GPO seek additional Congressional funding that would help achieve other initiatives such as equipment modernization, information storage and retrieval systems and other technologies as articulated or envisioned by the Public Printer.

Given the recent fully-funded appropriations GPO received, Council believes that Congress has a knowledgeable grasp of GPO's needs and is also mindful of the myriad of initiatives GPO is pursuing in order to maintain its leadership role in the dissemination of government information. As active partners in the dissemination of government information depository, librarians can participate in this process by urging their members of Congress to approve full FDLP funding as well as full funding of other GPO initiatives.

Should congressional appropriations fall short of funding needed to operate the FDLP, Council encourages the GPO to seek revenue from sources other than the sale of digital FDLP or *GPO Access* information resources. In the current digital age characterized by ubiquitous access to all Web-based information, citizens expect to have no-fee access to information by and about their government. Charging fees for access to FDLP information would limit the public's ability to participate fully in our democratic society.

Council realizes that the evolving information technologies extending the reach of FDLP collections to all American computer workstations have, at the same time, posed a serious challenge for the GPO Sales Program. Council recognizes and applauds the role that the sales program played in the age of tangible information products -- that of assuring the availability of government publications on a cost-recovery basis beyond the confines of depository libraries. Given the program's importance in the past, Council urges the GPO to analyze the sales program both to determine the viability of continuing this operation and determine new parameters under which it will operate in the digital age.

Council recommends that in this analysis GPO, as the federal government's lead information agency, should strive to assure that access to FDLP information is as free and ubiquitous as possible to support America's democratic system and economy. The public should have free access to all content, as publicly generated, in whatever formats and via whatever search interfaces are developed at GPO. Even XML-marked-up text should be available to the public since in the future this will be critical content format for depositories and their constituents.

Council applauded the decision GPO made to provide no-fee access to *GPO Access* in the mid-1990s. Today Council cautions that efforts to support the sales program not undermine levels of no-fee public access established at that time. Council believes that proposals to provide no-fee access to one level of information products while selling more sophisticated products to generate revenue could promote an environment of information rich and poor. Council is concerned that establishing this dichotomy of products may prompt some future GPO administration to disproportionately direct resources toward for-sale products at the expense of no-fee products.

There is a category of digital publishing services that might be open to GPO that would not risk exacerbating or promoting the information rich/poor divide. These are variously termed as "customization" or "personalization" services. Personalized services that might be acceptable to the library community include such features as: individual e-mail notification (of relevant proposed regulations from the Federal Register, for instance); portals for specific audiences (the trucking industry, for example); pushing downloads through programmed extractions; and personal logins that allow users to remember their searches and retrievals, create footnotes on the fly, and store documents online in temporary folders.

All of these personalization services are geared toward individual users. None of these services challenge either the public's right to unfettered, no-fee access to all government content or government metadata.

Endnotes

- (1) Office of Management and Budget. *Circular A-130. Revised (Transmittal Memorandum No. 4). Memorandum for Heads of Executive Departments and Agencies. Section 7. B.*
- (2) *Envisioning the Future of Federal Government Information: Summary of the Spring 2003 Meeting of the Depository Library Council to the Public Printer.* http://www.access.gpo.gov/su_docs/fdlp/council/EnvisioningtheFuture.html
- (3) *Report to the Congress: Study to Identify Measures Necessary for a Successful Transition to a More Electronic Federal Depository Library Program as required by Legislative Branch Appropriations Act, 1996. Public Law 104-53.* Washington, DC: U.S. Government Printing Office, June 1996.
- (4) Richard Grauss. *Versions.* Photocopied summary of remarks provided to Council, October 21, 2003.

