

GREGG HARPER, MISSISSIPPI
CHAIRMAN

RODNEY DAVIS, ILLINOIS
VICE CHAIRMAN

BARBARA COMSTOCK, VIRGINIA
MARK WALKER, NORTH CAROLINA
ADRIAN SMITH, NEBRASKA
BARRY LOUDERMILK, GEORGIA

SEAN MORAN, STAFF DIRECTOR

Congress of the United States

House of Representatives

COMMITTEE ON HOUSE ADMINISTRATION

1309 Longworth House Office Building
Washington, D.C. 20515-6157
(202) 225-8281

<https://cha.house.gov>

ROBERT A. BRADY, PENNSYLVANIA
RANKING MINORITY MEMBER

ZOE LOFGREN, CALIFORNIA
JAMIE RASKIN, MARYLAND

ONE HUNDRED FIFTEENTH
CONGRESS

JAMIE FLEET, MINORITY STAFF DIRECTOR

October 17, 2017

Laurie Hall
Acting Superintendent of Documents
U.S. Government Publishing Office
732 North Capitol Street N.W.
Washington, D.C. 20401

Dear Ms. Hall,

Thank you for testifying during the Committee on House Administration's October 17, 2017 hearing entitled, "Transforming GPO for the 21st Century and Beyond: Part 3 – Federal Depository Library Program". The Committee requests you respond to the additional questions (below) that will be made part of the hearing record. Please provide your responses to the following questions to the Legislative Clerk, Dan Jarrell (Dan.Jarrell@mail.house.gov), with the Committee by **Thursday, October 26, 2017 at 5:00pm.**

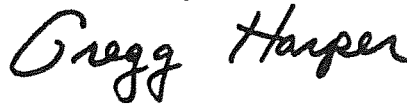
Questions for the Record Submitted by the Majority

1. In 2003, GPO and NARA signed a memorandum of understanding (MOU) to establish an affiliated relationship between GPO with preservation and access responsibilities codified in 44 U.S.C. Chapter 41, and NARA which operates under the authority of 44 U.S.C. Chapters 21, 29, 31 & 33. It would appear that through this MOU, GPO is providing preservation of all electronic information dissemination products in its possession. If this is accurate, can you help the Committee understand the concern by FDLP stakeholders about preservation of electronic documents?
2. The Committee is seeking clarification on the direct costs to provide tangible materials to regional depository libraries. In table format for each of the regional depository libraries, please provide how much was spent on purchasing and transporting tangible materials.
3. Please help the Committee understand where the SuDocs appropriations spent on tangible products go. In table format for FY2007 – present, for each of the four SuDocs appropriations, please identify the amount money spent for tangible products purchased from GPO's plant and the amount of money spent for tangible products purchased from private sector vendors.

Questions for the Record Submitted by the Minority

1. What is the status of the Trusted Digital Repository audit and the ISO 16363 certification process? What is the expected timeline for completion?
2. Does GPO Express provide a viable tool for capturing the printing of “smaller” jobs for inclusion in the Federal Depository Library Program (FDLP)?
3. What percentage of items purchased via GPO Express are eventually included as part of the FDLP program?
4. Do you recommend any legislative improvements to improve the operation of GPO Express as it relates to maintaining the FDLP?
5. In your prepared statement before the Committee, you reference sundry possible reforms to the FDLP but mention the list is not exhaustive. What are the possible reforms you did not mention?
6. We have learned that often, smaller selective depository libraries receive titles they did not select and are burdened with the attendant cataloging and storage responsibilities. How is your office working to improve the selection process for selective depository libraries?

Sincerely,



Gregg Harper
Chairman

cc: The Honorable Robert A. Brady
Ranking Member, Committee on House Administration

U.S. Government Publishing Office

Responses to Questions for the Record Submitted by the Committee on House Administration

October 26, 2017

Questions for the Record Submitted by the Majority

1. In 2003, GPO and NARA signed a memorandum of understanding (MOU) to establish an affiliated relationship between GPO with preservation and access responsibilities codified in 44 U.S.C. Chapter 41, and NARA which operates under the authority of 44 U.S.C. Chapters 21, 29, 31 & 33. It would appear that through this MOU, GPO is providing preservation of all electronic information dissemination products in its possession. If this is accurate, can you help the Committee understand the concern by FDLP stakeholders about preservation of electronic documents?

RESPONSE: GPO is preserving all electronic information dissemination products in its possession. The FDLP stakeholders' concern for preserving electronic documents is not with those maintained in FDsys/**govinfo**, but rather with those documents that are not held by GPO within FDsys/**govinfo**. Receiving certification as a Trustworthy Digital Repository will provide an additional level of confidence that the content in FDsys/**govinfo** is being preserved.

Of concern to FDLP Stakeholders is information not in GPO's possession. GPO is taking a proactive approach to acquire born digital content by harvesting agency websites. However, GPO cannot keep pace with the scale of agency publishing directly to the web. To mitigate this, GPO, the Library of Congress, NARA, national libraries, and other Federal agencies are collaborating through the Federal Web Archiving Group to harvest agency content and preserve this content on Library of Congress, NARA and other servers spread throughout the Government. This Group is working together to develop best practices for cataloging, harvesting, and preserving digital agency content and other guidance documentation for agencies in establishing principles for web-based information dissemination. Additionally, the library community as a whole are working together and exploring issues in many forums related to digital preservation, redundancy, data refuge, and related issues. Born digital fugitive documents are unsettling to FDLP stakeholders, who fear a "digital dark age" of Government information.

Another concern of FDLP stakeholders are the questions of redundancy and content being held solely in Government digital repositories. Regarding redundancy, GPO's FDsys/**govinfo** maintains at least two archival copies of information content in two separate geographic locations, has a secondary standby disaster recovery instance of the system for continuity of operations, and performs weekly backups over a redundant array of disk storage systems. All data is kept on enterprise-level storage with block-level version control and history, in addition to repository-level versioning and history. All archival copies are kept on separate physical and logical spindles and groups. Content is synchronized across geographic locations and regularly verified by both bit-level processes and content-level hash value verification checks.

Maintaining a redundant copy of FDsys/**govinfo** with a third party or specifically outside the jurisdiction of the United States is not an established best practice for digital preservation. Nor is it a requirement found in ISO 16363, *Audit and Certification of Trustworthy Digital Repositories*, the international standard and highest level of certification that is obtainable.

2. The Committee is seeking clarification on the direct costs to provide tangible materials to regional depository libraries. In table format for each of the regional depository libraries, please provide how much was spent on purchasing and transporting tangible materials.

LSCM does not routinely keep this kind of data. However, we have done some sampling of two regional libraries selecting nearly 100% of depository items. Purchasing and transporting costs for each library (even regionals) can vary based on the amount and type of material selected, the number of packages sent, and the physical location of the library in relation to the FDLP Distribution Facility in Laurel, MD.

For example, in FY 2017 the postage costs for the California State Library in Sacramento were \$1,551. For the Indiana State Library in Indianapolis, postage costs were \$885.00.

We used data from the month of May 2017 to determine the number of titles distributed and calculate purchasing copy costs. For titles distributed during that month we estimated costs by using the jacket costs and sampling the rider rates. We were not able to include costs for production associated with microfiche distribution since even the regional libraries that we sampled have unique selection profiles that require further data gathering and analysis for this format.

For FY 2017, 5,824 tangible titles were acquired for distribution to libraries in the FDLP. On average, based on available data, we estimated that approximately \$14,846 was spent during the year to acquire these titles for a regional library, or approximately \$2.55 per title.

3. Please help the Committee understand where the SuDocs appropriations spent on tangible publications go. In table format for FY 2007 – present, for each of the four SuDocs appropriations [sic], please identify the amount of money spent for tangible products purchased from GPO's plant and the amount of money spent for tangible products purchase from private sector vendors.

The single appropriation for the Public Information Programs of the Superintendent of Documents funds the Federal Depository Library, Cataloging and Indexing, By-Law Distribution, and International Exchange programs. Printing costs are incurred by two of those programs: the FDLP and International Exchange.

From FY 2007-2016, the majority of printed materials distributed to libraries in the Federal Depository Library Program were congressional hearings, reports, the *Congressional Record*, and related materials, all which are produced in GPO's plant. Regional depository libraries have the option to select a tangible format of congressional materials by committee, whether print or microfiche. Plant costs indicate that many regionals continue to select paper for many

committees and large selective libraries do the same. This data indicates the importance of these publications to the FDLP.

In that same timeframe, executive branch printing (represented in the “purchased” column) declined as more Federal agencies were creating born-digital publications and choosing not to print. Additionally, an increasing number of selective libraries in the FDLP were choosing to receive fewer tangible print publications and opting for access to born-digital publications by downloading bibliographic records from GPO with links to online versions on GPO servers or on agency websites, or other GPO partnership websites.

**S&E/PIP Appropriation
Printing Costs, FY 2007 - FY 2016**

FY	FDLP		Cataloging & Indexing		By-Law		International Exchange	
	Plant	Purchased	Plant	Purchased	Plant	Purchased	Plant	Purchased
2007	\$ 5,840,215	\$ 2,445,748	No Printing Costs for this Program	No Printing Costs for this Program			\$ 281,497	\$ 154,606
2008	7,611,936	2,021,000					425,715	84,639
2009	6,632,470	2,084,350					374,519	84,261
2010	6,245,347	2,046,530					545,099	79,021
2011	6,207,917	2,280,846					547,019	88,814
2012	7,352,223	1,688,731					626,278	54,221
2013	6,705,324	1,375,055					532,252	75,949
2014	6,848,068	1,288,180					534,709	77,697
2015	5,150,125	1,189,419					362,881	113,997
2016	5,961,868	1,506,012					373,189	188,910

Questions for the Record Submitted by the Minority

1. What is the status of the Trusted Digital Repository audit and the ISO 16363 certification process? What is the expected timeline for completion?

On September 26, 2017, GPO posted a solicitation to obtain the services of an auditing body to perform the external audit and assessment of GPO’s FDsys/**govinfo** repository through certification under ISO 16363:2012. GPO expects to begin the proposal evaluation phase for this procurement on October 30, and to award the contract in the winter of 2017/18. GPO anticipates the audit to be completed in 9-12 months (by the 1st or 2nd quarter of FY 2019).

This is the background to this process:

On December 18, 2014, GPO announced that we were preparing to become the first Federal agency named as a Trustworthy Digital Repository. To prepare for an external audit, GPO was named by the Library of Congress and the National Institute of Museum and Library Services as one of 5 institutions to receive a resident as part of the National Digital Stewardship Residency

program. This resident spent one year performing an internal assessment of GPO's digital repository against the ISO 16363:2012 standard. This internal assessment resulted in a set of recommendations and listed opportunities for GPO to expand current digital preservation practices. Between June 2016 and August 2017, GPO responded to the recommendations from the internal assessment. Of significance, the ISO 16919:2014 standard, used to provide accreditation for ISO 16363:2012 certifying bodies, was recently released to the public. At the time of the internal assessment, no certifying bodies or auditing bodies had announced accreditation under this new standard in order to perform ISO 16363 audits. In August 2016, GPO released a Request for Information in order to discover if any bodies had yet been accredited. GPO received responses to this RFI, and from the information gathered, as well as market research, GPO proceeded to seek external certification with the confidence that certifying bodies would be available to perform an audit in FY18.

In order for GPO to receive certification as a TDR, an accredited auditing body must perform an external audit of GPO's preservation repository, FDsys/**govinfo**, in the areas of Organization Infrastructure, Digital Object Management, and Infrastructure and Security Risk Management against 109 enumerated criteria within the ISO 16363:2012 certification process. A certification under ISO 16363:2012 will reinforce GPO's commitment to its mission of ensuring permanent public access to Federal Government information. Additionally, this certification will further validate that FDsys, its infrastructure, and its supporting organization are reliable and sustainable, in helping to ensure the long-term integrity of its digital holdings.

2. Does GPO Express provide a viable tool for capturing the printing of "smaller" jobs for inclusion in the Federal Depository Library Program (FDLP)?

GPO's Library Services and Content Management (LSCM) business unit participated in the GPO Express program from FY 2007 through FY 2013. Based on our analysis and review of materials coming through the program for the FDLP, we determined that it was not a cost effective source for capturing fugitive documents. On average, during these years, LSCM harvested only 13% of files submitted. What became clear was that agencies were using the GPO Express Program for a variety of purposes:

- To make copies of materials for presentations (i.e., PowerPoint slides);
- The files were a page or two, a section or chart of document and therefore incomplete;
- The files of documents were ephemeral, of short duration (i.e., a poster announcing a blood drive at an agency facility);
- The files were drafts, and a check of the agency website identified a final, or newer version was available; or
- The files were out-of-scope of the FDLP.

While the majority of documents that were harvested from GPO Express were often pamphlets, or one or two page titles, LSCM staff routinely searched agencies websites to determine if a more up-to-date or complete and final version was available directly from the agency site. Staff routinely chose to harvest documents from the agency website instead of using the GPO Express file to maintain digital file provenance.

3. What percentage of items purchased via GPO Express are eventually included as part of the FDLP program?

In our analysis from FY 2007-FY 2013, LSCM staff reviewed 19,382 files through the GPO Express Program. Of those files reviewed, only 2,599 files were included in the FDLP, or 13% were harvested of the total that were reviewed.

4. Do you recommend any legislative improvements to improve the operation of GPO Express as it relates to maintain the FDLP?

LSCM believes that while the GPO Express Program appears to work well for the agencies that use its quick and convenient printing services, it is not a viable or cost-effective method to identify fugitive documents for the FDLP. In FY 2014, LSCM began to accelerate our use of web harvesting technologies to acquire agency digital content for the FDLP. This current method and approach, harvesting directly from agency websites, ensures that we capture and preserve agency content that is no longer being issued in tangible format. The approach is being utilized by others in the Federal library community, including the Library of Congress, the National Archives and Records Administration, the Smithsonian Institution, and other national libraries.

5. In your prepared statement before the Committee, you reference sundry possible reforms to the FDLP but mention the list is not exhaustive. What are the possible reforms you did not mention?

We left the list open at the time because it was provided in advance of our annual Federal Depository Library Council Conference, where there were opportunities for further community discussion of legislative options. At this time we would add the following:

- Rename Chapter 19 to Public Information Programs of the Superintendent of Documents and include a new section 1900 that conveys that purpose;
- Incorporate Chapter 41 into Chapter 19;
- Revise sections 1710-1711 and incorporate them into Chapter 19; and
- Insert a provision for GPO to provide support services to depository libraries (e.g., education and training, preservation, cataloging records).

6: We have learned that often, smaller selective depository libraries receive titles they did not select and are burdened with the attendant cataloging and storage responsibilities. How is your office working to improve the selection process for selective depository libraries?

GPO allows selective depository libraries to select materials for their collections by using “depository item numbers.” This selection mechanism and approach dates back many years. The technology used to provide items for selection by libraries is up-to-date, but the selection approach has been modified only slightly over the years.

As background, selective libraries select depository item numbers representing an agency title or category of titles that meets the needs of their patrons and community. This selection constitutes what is called the depository library profile. New depository item numbers are created daily based on the agency publication activity. GPO notifies libraries of new depository item numbers that are created daily. Selective libraries are allowed to add or deselect new digital items to their selection profile at any time and drop any tangible item (microfiche or paper) at any time. Any material received by the library after the date of deselection, can be offered to other libraries or recycled as the library does not have to keep these and include them in their collections. Libraries are responsible for checking their item selection profile periodically to ensure that it still reflects their current needs.

GPO makes some selection decisions on behalf of the libraries based on prior preferences for similar items, but this practice is very minimal. In FY 2017, only 4 new titles were sent in tangible format under this practice. The majority of new material coming into the program is digital. Thus, selective libraries receive very few tangible items that they do not select.

To provide increased flexibility for selective libraries, GPO has recently taken two actions:

- GPO recently eliminated a depository item number that was sent to all libraries, regardless of selection profile. GPO felt that titles warranted broad distribution based on it being a “hot” topic or if something was deemed to be of interest to all libraries. At the request of many libraries that were moving to mostly digital, this practice was stopped in 2016.
- In 2015-2016, GPO enhanced the ability of selection by format. Thus, selective libraries can select a tangible format, microfiche format, or a digital format for many of the titles and categories of titles that are offered. This modification helped libraries with space issues and allowed them to choose formats that work for their community.

GPO catalogs all formats of titles distributed to libraries in the FDLP. GPO provides many avenues for libraries to obtain free cataloging/bibliographic records for the depository items that they select. Unless a library has unique or specific cataloging requirements, GPO’s cataloging follows all national and international standards which enables most libraries to use GPO cataloging with little or no modification.

For clarification, in the past several years, there has been a significant decline in the number of tangible titles selected by libraries. This is due to more agency titles being published only in digital format. In addition, the selective libraries in the FDLP select fewer items in tangible format.

GPO recognizes that changing this method of selection by agency and depository item number is needed. This would be a significant change to the FDLP operation, requiring not only changes in the LSCM organization and workflow at GPO, but also major changes in workflows at libraries.