Digital Deposit Working Group report to the Depository Library Council
Proposed Pilot

Revision History

Revision	Date	Revision Description	Author
1.0	July 19, 2019	Initial Draft	Jessica Tieman
2.0	February 12, 2020	Draft	james

1.0 Introduction

1.1 Background

In December 2018, GPO released a document titled "Developing a Multi-State Comprehensive Collection: FDLP Preservation Services Pilot Strategies,"[1] in which GPO proposed a pilot strategy to explore the topic of digital deposit as a service. To begin, GPO proposed its goal "to define a pilot project, including identification of needs and resources, in order to implement effective solutions for depositing digital content within the FDLP program with respect to needs for content integrity, preservability, collection development, usability, and user community priorities." In response to the proposed pilot, the Depository Library Council issued a recommendation calling for "the creation of a working group to explore current and future needs related to digital deposit - both dissemination of content and acceptance of content by GPO."[2] This working group, with representatives from Council, the FDLP community, and GPO, hosted two programs to explore community definitions and expectations of "digital deposit" as a service model. Based on feedback from the programs, the working group defined a vision for "Digital Deposit" as an approach to publication acquisition, preservation, and dissemination. The working group is proposing a pilot that will explore technological aspects of digital deposit and will inform changes to the collection and preservation of "lost/fugitive" documents that will grow the National Collection by contributing content to GPO and its digital repository, govinfo, as a service hub.

1.2 Digital Deposit and the FDLP

Though GPO envisions Digital Deposit as a potential service or mechanism that could support both the lifecycle of digital content as well as a comprehensive collection; GPO is uncertain what specific needs from the FDLP a service like Digital Deposit could fulfill. Since 2005, GPO distributed biennial surveys posed varying questions around the topic of digital deposit.[3] The variance in the questions asked make collecting qualitative information about the topic very difficult.

In April 2007, GPO issued a white paper on what was then called "Digital Distribution"[4] which stated that 93% of all new titles made available through the FDLP were available in electronic form; thus GPO would consider "an affirmative distribution of authenticated and official published digital content to Federal depository libraries." To this extent, the focus of digital deposit was centered on simply disseminating publications as digital files.

---

[1] Developing a Multi-State Comprehensive Collection | FDLP Preservation Services Pilot Strategies (December 20, 2018)
[2] Recommendations & Commendations from the DLC to GPO
[3] https://freegovinfo.info/node/12457
[4] Digital Distribution Issue Brief (April 2007)

At this time, GPO had some early assumptions about what "digital distribution" was. For instance, it was assumed that FDLs would want to download publications or receive them from GPO through a "push" mechanism. In 2007, the questions and concerns relating to digital distribution reflected concerns on how to authenticate files, synchronize files across libraries and sites, requirements for selective and regional libraries to retain derivative files, versioning of electronic publications, and the role of libraries and redundancy of digital information for preservation.

Between 2007 and 2009, GPO also witnessed a growing interest from FDLs now identifying digital deposit as something which could be a collective effort, focused not just on access, but also preservation. The increased interest in preservation led GPO to participate in the Lots of Copies Keep Stuff Safe (LOCKSS) USDOCS program along with 36 FDLs.[5] With the existing ambiguity around whether or not FDLs were interested in receiving content from GPO's digital repository or simply having the option of distributed copies of content across the FDLP, GPO established an internal policy that reflected the capabilities of FDsys at the time. This internal policy, Superintendent of Document Policy Statement (SOD) 321 Digital Dissemination of Access Content Packages for FDLP Digital Depository (2010) stated that rather than "pushing" digital content through an automated means, FDLs would instead proactively "pull" content from GPO's online access interface whether by means of saving and downloading individual content directly, or harvesting content through an application programming interface or similar technology. Currently, libraries still have the full capability to harvest PDF content directly from govinfo. GPO also released a govinfo API which allows users to retrieve bulk summaries of metadata, individual PDF files, and html files from over 30 different collections in govinfo through server requests.

How FDLs define "digital deposit" and expectations of digital deposit as a service continue to be complicated. In the 2014 *FDLP Forecast Study*[6], GPO queried FDLs about their "most pressing issues, goals and viewpoints." Responses specific to "Digital Deposit" were mentioned in both parts of the survey relating to access and preservation; this suggests that, across the FDLP, library expectations and definitions of "Digital Deposit" as a service may vary from library to library. More frequently, through survey data, it seemed to be that interest in becoming an "all-digital depository" was becoming synonymous with the language of "digital deposit," thus it became unclear if digital deposit should be considered a service model or a method of item selection. In 2016, GPO's *National Plan for Access to U.S. Government Information*[7] highlighted a dedicated goal to increase the number of digital depositories in the FDLP, but did not describe if digital deposit would be an explicit service or activity going forward.

Going forward with the information collected so far, GPO recognizes that the demand for digital deposit has been prevalent across the FDLP for nearly 15 years; however, there are differing definitions and expectations of what "digital deposit" as a service provides to the FDLP and what priorities it is best suited to solve for.

1.3 Working Group Activities to Date

The Digital Deposit working group convened in early 2019. At that time, the group consisted of three members from the Depository Library Council, five community members, and three representatives from GPO. Early group

---

[5] "Lots of Copies Keep Stuff Safe" https://lockss-usdocs.stanford.edu

[6] FDLP Forecast Study

[7] National Plan for Access to US Government Information

discussions aimed to define digital deposit for the FDL community and how it supports the National Collection. The group explored options for digital deposit pilots defining for each the pilot objectives and the roles of pilot participants.

As a means to share with the larger community internal discussions and ideas, the working group offered a session during the 2019 Spring FDLP Virtual Meeting, Digital Deposit: A Value Proposition. The session queried conference attendees to gauge their understanding of digital deposit, reception of participating in this work, and offered a value proposition for participating in this work. Based on answers to questions posed and discussion via the conference chat, it was clear that confusion about what digital deposit *is* persists.

In June 2019, a few members of the digital deposit working group were able to meet during the ALA Annual Conference where a consensus was reached on proposing three pilot projects – deposit from GPO to Libraries; deposit from Libraries to GPO; deposit from agencies to GPO. To share the pilots and continue to solicit feedback from the community, the working group hosted an in-person session at the Fall 2019 FDLP Conference. During this session the working group queried participants on the value of digital deposit, the role of the FDLP in this work, need for this work, and needed support among other things; see appendix for questions and feedback.

Since the FDLP Conference, the group gathered additional information about the possible expectations of libraries that might receive digital content and GPO's ability to share content. The working group has also learned more of GPO's expectations for potential pilot outcomes and how efforts could align with proposed legislation that would impact GPO. Based on this feedback, the working group proposes a single pilot project[8]: Lost Docs reporting and ingest.

2.0 Project Overview

2.1 Goal

The goal of this pilot is to implement "Digital Deposit" as a full ecosystem approach to reporting lost documents to GPO which are ingested and made available through govinfo.gov with an option to be disseminated to FDLs through API technology. The reporting of documents to GPO could require a human-centered approach, such as forming a collaborative relationship with the Rare and Endangered Publications Committee of GODORT or some other entity. In addition, a technological solution to replace the existing LostDocs reporting process is intended to be a piloted component of this project.

2.2 Objectives

---

[8] Since receiving its charge, the working group considered three pilot projects - the proposed Lost Docs project; a project in which libraries ingest content from govinfo; and, a project where agencies deposit content with GPO. Because of limitations on GPO's ability to share masterfiles, the group decided not to propose the GPO to FDL project at this time. The third project would require significant changes to GPO's current workflows and mechanisms; it is out of scope for DLC to recommend such changes and work demands of GPO.

- Scope the pilot to [House Document 116-4 – Reports to be made to Congress](#)

  It was decided to scope the project to the list of reports to be made to Congress in order to 1) keep the project scope to a narrowly focused list of agency documents for the pilot; and 2) to better understand the impact of passage of [H.R. 736: Access to Congressionally Mandated Reports Act](#) -- which has passed the House in the 116th Congress and is waiting on Senate consideration.
- Identify opportunities for FDLs to form a collaborative system for reporting publications to GPO
- Identify aspects of document reporting which can improve GPO staff workflows for cataloging and ingest
- Assess the usefulness of the current LostDocs reporting mechanisms
- Create from a user perspective requirements for a long-term solution to creating an ecosystem approach and eliciting minimum requirements for "Digital Deposit" and its various human-centered and technology-centered components.

2.3 Pilot Stakeholders

- Volunteers from FDLs and/or members of GODORT Rare and Endangered Government Publications Committee.
- GPO staff from LSCM and Program Strategy and Technology
- Depository Library Council
- DLC Digital Deposit working group

2.4 In Scope

The following items are the scope for this project:

- [House Document 116-4 – Reports to be made to Congress](#)

3.0 Project Approach

The following avenues of work are intended to be conducted concurrently.

3.1 Lost Documents Stakeholders

- The working group will recruit volunteers via various means (listservs, spring virtual meeting, LSCM news alerts etc).
- Capture initial stakeholder needs and user stories
  - Conduct group discussion meetings about the current workflows for lost documents
- Explore major aspects of lost documents as they relate to:
  - Collecting documents (FDL community)
  - Cataloging and indexing (GPO)
  - Capture current state-workflows, capabilities, and limitations

MARC XML and Metadata and PURL and PURL Server Technologies were considered as part of this phase, but decided that these would be better for future exploration.

Deliverables:

● Documented current-state workflows, capabilities, limitations, stakeholder needs, and user stories.

3.1.2 Govinfo as an Ingest and Service Hub

● Identify current mechanisms for reported documents to be ingested into govinfo
● Explore future mechanisms
● This will be the internal responsibility of GPO

Deliverables:

● To be determined by GPO and reported to pilot committee

3.1.3 Stakeholders work on fugitives

● Stakeholders will do their work, keeping track of metrics as work advances
● Workflows to be determined by the stakeholders

Deliverables:

● Fugitive documents from House Document 116-4 will be collected, described and ready for incorporation into govinfo and CGP. Metrics and tools for doing this work will be collected and reported to Council by the working group.
● Metrics and feedback will include, but are not be limited to:
  ○ Tools utilized
  ○ Time spent
  ○ Number of documents searched
  ○ Number of documents discovered
  ○ Barriers / challenges

3.2 Proposed Timeline

● April 2020: recommendation from DLC to GPO to implement pilot project

Timeline contingent on approval and recommendation from Council

● Spring 2020: Recruit volunteers; there is no cap on number of volunteers
● Summer 2020: Hold discussions and gather feedback around current workflows, user stories, project processes, reporting metrics, and projected outcomes (as described in 3.1).
● Fall 2020: e, pilot project participants will be under way with collection of fugitives (as described in 3.1.3).
● Spring 2021: Working group reports out pilot metrics and feedback at Spring Virtual Meeting

3.3 Roles and Responsibilities

Role		Responsibilities
Working Group Members	Cindy Etkin, Designated Federal Official Lisa LaPlant, GPO Liaison Jessica Tieman, GPO Liaison	Pilot facilitation Communication with DLC

	Robbie Sittel, DLC Member Will Stringfellow, DLC Member Heather Christenson, Community Member James R. Jacobs, Community Member	Communication with FDL community
Project volunteers	TBD	Locating and reporting documents from HD116-4
Project manager	TBD	Directing and tracking volunteer contributions Ensure pilot work and deliverables are completed accordance with pilot timeline Collating data Reports pilot activities to working group

3.4 Communication

The Digital Deposit Working Group has conference calls at least monthly, or more as needed on an ongoing basis. Members also communicate via email and collaborate on documents.

The Digital Deposit Working Group offered *Digital Deposit: A Value Proposition*, an information session and virtual discussion, during the Spring 2019 Depository Library Council Virtual Meeting.[9]

At the 2019 Federal Depository Library Conference, the working group held a focus group style session. See the appendix for questions and responses from this session.

Working group provides periodic updates to DLC.

Pilot Project Vision Document presented to DLC in March 2020.

4.0 Project Costs

4.1 Budget and/or Resources

TBD if pilot approved.

---

[9] https://www.fdlp.gov/events-and-conferences/3940-2019-depository-library-council-virtual-meeting

5.0 Project Conditions

5.1 Constraints

5.1.2 GPO Constraints
1. Resources to implement any recommended changes to existing reporting mechanism

5.1.3 Community/Volunteer Constraints
1. Time and workload constraints
2. Institutional expectations or limitations
3. Technological know-how

5.2 Assumptions

With the majority of information being published directly to the web by the issuing agency, it is assumed that GPO is not aware of all that is being produced. Because GPO is no longer part of the publication process, it is assumed a large amount of federal information is being excluded from the CGP and govinfo.

5.3 Risks

The 2013 NAPA report, *Rebooting the Government Printing Office*, described GPO's current landscape as having "exponential growth of 'fugitive' digital government documents, which has complicated GPO's ability to authenticate and preserve valuable history." With the majority of government information published on agency websites and circumventing traditional print publication workflows, the potential loss of information to Keep America Informed is greater than ever.

The unknown frequency of Congressionally mandated publications as outlined in HD 116-4 may result in a small number of publications for the proposed pilot.

5.4 Opportunities

GPO has the opportunity to explore new collaborative workflows to decrease the number of fugitive documents and increase content accessible in the CGP and govinfo.

The FDL community has the opportunity to comment on existing LostDocs workflow and be part of a potential new service model offered by GPO.

The FDL community has the opportunity to contribute to the collaborative building of the National Collection.

Community will get more information on the extent of fugitives, how the situation comes to be, and strategies for finding them. In general this will serve us well in bringing them into discovery and access paths.

6.0 Project Approvals

Appendix

A. Discussion Questions presented at 2019 Fall Depository Library Conference

1. What is the value of this work… to the community, to users, to others?
2. Who are the active participants in digital deposit?
3. How do you see the role of an FDL in digital deposit?
4. What does acquisition of digital-born government information look like?
5. How will these projects fill a need in the FDLP?
6. What risks do you see in these projects?
7. How will librarians engage with the projects? What type of engagement do you foresee with these or similar projects?
8. What kind of support do you envision librarians needing to participate?
9. How will local patrons use the content?
10. What are your thoughts about authenticity of publications, chain of custody within the framework of digital deposit, and authorization of various actors?
11. Do you have ideas/suggestions for other projects?

B. Discussion Question Responses

Q1. What is the value of this work... to the community, to users, to others?	Q2. Who are the active participants in digital deposit	Q3. How do you see the role of an FDL in digital deposit?	Q4. What does acquisition of digital-born government information look like?	Q5. How will these projects fill a need in the FDLP?	Q6. What risks do you see in these projects?	Q7. How will librarians engage with the projects? What type of engagement do you foresee with these or similar projects?	Q8. What kind of support do you envision librarians needing to participate?	Q9. how will local patrons use the content?	Q10. What are your thoughts about authenticity of publications, chain of custody withint the framework of digital deposit, and authorization of various actors?	Q11. Do you have ideas/suggestions for other projects?
fewer fugitives	selector	overseer	**right click save as, drops into workflow**	bnefit of participation (capacity building)	authenticity	time constraints	**staff time**	local repository?	**GPO is the trusted party**	have a data repository so data sets can be accessible (back up of data.gov) with better metadata/indexin/cataloging

preservation of materials that agencies may remove from their websites	technologists	provide policies if FDLs plan an active role in deposit and catalog and providing access	(or, what about an approval plan)	secures collection locally in case of need	value of content	academic public univ student volunteers	server space	local library federated search	**approve FDLs to accept deposits**	**serials archived by GPO**
more accessible to the public	users	all can provide access to deposited content	select by agency, format, topic	**skill development fo digital preservation**	authenticity of items	**library MLS students - an actual hands on course**	**admin support**	local library catalog	users don't usually care who hosts this, they care about finding	**figure out how to package together things that belong together (ata documentation or serials/monos)**
preservation	metadata people	not all will want to collect, not all can	NO ITEM NUMBERS please	**increase diversity of collections**	**authenticity/authority**	planning, coordinating, promoting	liaisons to agencies	open hub search	NO BLOCKCHAIN please	curation and quality control workflows
96% of new publications are online	systems people	whether or not they can accept deposit, they could	capturing and archiving documents and	help to make more accessible a comprehensive	data loss	local advocacy	$$$$	PURLS or PDFs	very important and must be maintained	finding ways to engage

		contribute	providing indexing and metadata to mak it searchable and accessible	national collection and record						
comprehensive	GPO	leading with standards for partners	central person everthing would go through		destruction of tangible materials	learn digital preservation and tech skills	tech skills - enable digital preservation and information architecture	CGP	**for authenticity both at ingest and being kept long term to avoid both loss and manipulation**	**automation and taking web archives extracting items to catalog**
shed light on government activities	administrators	organize the players into different projects / agency exploration	alerts with new content		lots of time planning and then nothing happens	**take ownership in the FDLP!!!!**	**community of practice**	local catalog records, local discovery layer?	provenance is critical	individual documents - (unreadable) and make discoverable
preservation	open ended	state higher priority than federal	centralization		**GPO changes priority**	fall to libraries with staffing and time	tool for content identification / submission	local repository storage?	provenance - a part of the record (ex. article donated by	record searches - what's missing from

							other than lost docs		USMC command, #28 from his person collection)	current catalog record
more accessible to the public	**open gov't / crowdsourcing**	**lost docs project is a great suggestion, contribute to this with pubs**	**"archival sliver" - what to capture (like a records schedule)**		new leginstation or "interpretation" of title 44	promoting GPO - early adopters	guidelines and standards for archival purposes	**research**	we do get this question about how we obtain items	regional collecting focus on materials of interest in region (e.g. calif related docs in calif)
public access	citizens	state docs program experience can share best practices, issues	**automatic alerts when new publications on agency website - with metadata harvesting and comparing**		hoarding is fun but runs up local tech costs	new coordinator don't have much time, a lot of things to learn.	**a platform or program for digitizing and injest**	research	CRITICAL	digitization data website
capacity building (local)	**agencies**	federal efforts could lead into developme	would SuDoc set a priority list		**climate change**	depends on staffing	internal administration support	**civic engagement**	minimize # of steps something travels through	**each region could focus on local area federal pubs and**

		nt of state efforts							**capture a lot of lost docs, in each region's interest, would help all libraries participate in a shift in thinking about depository collections but still in line with priorities, documents could feed into regional collection from any libraries (more people helping w/o smaller libraries absorbing workload) even if**

									materials stored elsewhere - Team: 1 large selective, 3 regionals, 2 GPO staff
fits library missions	open - please contribute	(FDLs can often see state docs as priority)	**a bot to detect publications posted on agency websites**		GPO	money - staff time, equipment, training and hiring, software and hardware	knowledge about agencies		central place for agencies to report their digitization plans
builds relationships	PQ paid (are they making money on the govt items thy digitize)	do other document pulls as was done in the 2006 EPA document pulls	priorities		ensuring authenticity	focus on agency outreach	**specific local stake holders**		digitization plans database - agencies and libraries can list what then plan to digitize
we used to act like islands but no longer a good idea in	Citizen Archivist	GPO could pull docs by agency and FDLs could catalog	publication equivalent to record schedule		criteria for inclusion	staff time	provenance is critical		

our current economic model		them or make recommendations							
national comprehensive collection	HathiTrust	**interested FDLs can sign up as GPO partners for wahtever level of involvement they are interested in/can manage**			ensuring the authenticity of the documents		training	View, download	
history of country	Internet Archive				inclusion		hiring	research/read	
revealing the existence of an unknown doc	agencies				valuable		software	analyze	
prevent loss and provide access long term	IT				If libraries host deposited data could their servers be overwhelm		librarians serving as liaisons to certain agencies	text mine	

					ed					
coordinated approach, more consistent	issuing agency						training	access during govt shutdowns		
leverage resources and share costs	**depository libraries - target an agency look for fugitive documents - check against CGP - harvest**							**having a single search (CGP) rather than digging through agency websites**		