Visualizing Deposit of Digital Content from the Government Publishing Office

Report of the
Visualizing Digital Deposit Subgroup
of the Digital Deposit Working Group,
a Working Group of the Depository Library Council

February 9, 2022

Members of the Visualizing Digital Deposit Subgroup

Julia Ezzo
Michigan State University
Member, Depository Library Council

Heather Christenson
HathiTrust
Depository Library Community

James R. Jacobs
Stanford University
Depository Library Community

Cynthia Etkin
Government Publishing Office

Jessica Tieman
Government Publishing Office

Table of Contents

Introduction

The Digital Deposit Working Group (DDWG) was established to "explore current and future needs related to digital deposit— both dissemination of content and acceptance of content by GPO". A subgroup, the Visualizing Digital Deposit Subgroup (VDDS) was tasked with exploring or "visualizing" what the deposit of digital content by GPO to Federal depository libraries might look like.

VDDS approached this charge by identifying use cases, developing high-level questions, and conducting interviews to garner in-depth feedback. Questions were asked of librarians representing each of the identified use cases to bring to fruition the concept of digital deposit to Federal Depository Library Program (FDLP) libraries.

Methodology

VDDS members created a chart for the different types of library and non-library organizations representing different use cases: FDLP regional library, various FDLP selective libraries (academic library, law library, public library, Tribal college library), individual library practicing digital preservation, an ASERL "Center of Excellence" library, digital preservation steward, and library-adjacent non-profit organizations working in the digital preservation and access space HathiTrust, and the Legal Information Institute (LII) at Cornell University. One representative organization per type was selected, primarily based on a positive response to past *Biennial Survey of Federal Depository Libraries* questions asking if their organization would be interested in digital deposit. For organization types that did not have a positive response to this question, subgroup members selected representative organizations using the FDLP directory, while taking in to account regional distribution to try to allow for cross-coastal representation by participants. Two of the organization types are not FDLP participants (HathiTrust and LII), though their institutional members and parent organization are FDLP libraries. Subgroup members selected interview candidates based on their knowledge of appropriate organizations.

The open-ended interview questions were designed to elicit detailed responses regarding interviewees' thoughts surrounding the feasibility of digital deposit. Questions were left intentionally nonspecific, with minimal definitions, clarification, or contextualization provided, in an effort to avoid influencing the responses. In the case of interviewees that

were not members of the FDLP (HathiTrust and LII), some of the questions were modified slightly to make them relevant to that particular organization.

Library Services and Content Management's eLearning platform was used to conduct the interviews. There were fourteen interviewees representing the ten business cases. And there were at least three, often more, subgroup members in attendance at the interviews. Each representative organization was interviewed separately by one member of the subgroup, though other subgroup members were in attendance. The interviewer read aloud each question, and allowed the interviewee(s) to ask clarifying questions prior to providing their response. Subgroup members were also able to ask follow-up questions to the responses given in the event that they wanted additional clarification or elaboration. Interviews were recorded in nine out of ten instances—one interview was not recorded due to technical difficulties. Responses were noted by one of the subgroup members during the interviews and the recordings were later reviewed by subgroup members to collate the responses for analysis.

Findings & Analysis

The interviews shed light on the libraries' point of view and provided details on how they envision different aspects of a digital deposit service. The following is a review of responses to the seven specific questions.

Question 1: What content would your library be most interested in receiving from GPO? For example, the latest hearings or maps for your geographic region.

While two of the libraries interviewed indicated that they wished to receive all available materials from GPO via digital deposit, the other eight interviewees preferred to receive a **selection of materials** through digital deposit. Interviewees were interested in selecting materials on a particular **subject**, such as legal or environmental materials. They also expressed a desire to receive all materials in a particular **format**, for example maps or XML files. Receiving all materials for specific **geographic** or regional areas was also suggested.

Question 2: How would you envision the selection process?

Interviewees want an easy, flexible method for selection that is clear as to what exactly a library is selecting. One librarian working with "a very diverse community ... would like to

be able to have easy access to a variety of topics and be able to modify selection of materials easily and quickly to meet patron interests."

A desire to move beyond the rigidity of the current item selection profile model to allow libraries the ability to **select materials by a variety of methods** was expressed. Examples noted in the interviews include: select all items in a particular **format**, select all materials that deal with specific **geographic** regions, select by **collections or general categories** (such as all legal-related publications, all material dealing with environmental **topics,** etc.), pre-made collections by library type (law library, public library, etc.).

Question 3: How would you make digital content accessible to the public?

The interviewees mentioned a **variety of modes for making digital documents accessible to the public**. The most frequently envisioned situation was to host the digital documents on a local platform, catalog them, and **provide links via library catalog**. Additional respondents described linking to the digital documents via a catalog, but did not specify that digital documents would be hosted locally. These responses included "on a server and point to it," and "online through a shared catalog." Related to the catalog model, we heard that "ideally cataloging and metadata come with the digital volume." Inclusion of bibliographic data as part of the process seemed to be an underlying assumption for many interviewees although they did not specifically point out that as a requirement.

Libraries were also **thinking beyond the traditional catalog access model**. Answers in this vein included "in catalog but also explanatory/promotional text so people know to look in the catalog," "links from **library website/web pages,"** "content visible on the web site," and "full text index." There was also an emphasis on **collections as a mode of access** by some interviewees, with mentions of "Browse by collection or SuDoc" and "Curated collections created by the institution."

Three libraries mentioned **provision of documents for computational and experimental use**, saying "present data and documents in a way that facilitates understanding by the general public and makes use of some computer science techniques to extract features of the documents themselves to make these more explicit," and "Allow experimenting with context for digital volumes" and interest in creating a "research 'sandbox.'" One interviewee emphasized the **preservation** aspect and envisioned they would "Preserve but point to GPO for access." One interviewee specifically mentioned "**downloadable for users"** as key.

Question 4: What would be the optimal mechanics of digital deposit (push/pull) for your library?

The responses to this question were mixed and somewhat inconclusive, with slightly more interest in a "push" model than a "pull" model, and with discussion of a number of ideas that don't fit neatly into either category but that are beneficial to designing a service. Of note is that this question was asked without specific context or definition of the models, other than the mention of push or pull, in order to give a starting point for the discussion and allow the libraries to put it in context themselves.

Four libraries emphasized the **push model** in their answers, envisioning a push from GPO to the library to be both **familiar and easier**, as it is "similar to what we have with tangibles", and that the library could "select what we want then it just shows up" with "one less thing to have staff do." Multiple libraries mentioned another advantage of this model, that it solves some of the challenges of "gathering materials in the digital age" and would help libraries "know that they are receiving all that they should be receiving." One library summed it up "this method keeps it **authenticated, authoritative, and complete.**"

Three libraries indicated an overall preference for a **pull model**, mentioning familiarity with this model and a **record of success in using it**, and also emphasizing that **this kind of model gives libraries more control and flexibility** with such comments as "pulling allows for flexibility to get content when you're ready" and "it would be nice to have control as to when to acquire." Some pointed out that their experience of push methods has been burdensome, noting that they "tend to have **issues and errors with push**. When pulling content either through URLs, Dropbox content, rsync, etc. These tend to work better and have less issues." and "with push it'd always have to be on to receive content when pushed." The conversations with libraries also brought up the point that a **pull model would necessarily require a method to share information on what items are available to pull** periodically, such as a "push notification from GPO to indicate material is available and then the library would be able to go in and retrieve those files." or "tools like bulk data and feeds that announce the changes in a standardized format." Some libraries cited pulling MARC records from GitHub and MARCIVE as examples of the kind of processes that serve their needs.

Several libraries noted that **having "cataloging records attached" would be key to their adoption of either a push or pull model**, and that "It'd be difficult to have several thousand files coming in at once without metadata attached." Some respondents expressed an interest in receiving metadata beyond bibliographic data that could serve a

wider set of use cases, wondering "what sort of metadata (description, technical, and preservation) would be paired with these records?" There was also mention of a need for "**a process to review the material** that is available to decide if we want it or not before downloading" with one library hoping for a way to craft a **selection process** that would allow them to "assume that when they're notified content is available then it is something they want to add without additional review."

Another concept that was brought up was a process that would also allow the library to contribute digital documents to GPO, "**would like for communication to be two-way** and all institutions to be able to deposit information into the system themselves rather than it being a one-sided process of just receiving materials."

Question 5: Would your library be interested in receiving physical formats (e.g., DVDs) of digital objects for offline digital access?

Libraries were **not interested** in receiving physical formats for offline digital access, with the exception of one library that allowed for the possibility, stating, "that falls into a conversation about what the library will look like in 5 years." Reasons for rejecting the idea of physical formats for digital access included lack of software and hardware needed to read the format, difficulty of "physical thing management" stewardship and format migration, and that users no longer understand why they'd need to use a DVD, or how to use it. Although not an access scenario, one library mentioned hard drives as a mechanism to transfer a large volume of files for ingest, "but this wouldn't be ideal or preferred."

Question 6: What kind of support/resources (technical, human etc.) do you think your library might need? What support/resources/infrastructure/staff are already in place at your library that would facilitate digital deposit?

In terms of support, libraries expressed a need for documentation, **training**, and clear **communications** from GPO. Having a **predictable process** would enable local workflows and automation. "Materials to promote the project, value, and ease" would help to get administrators and staff on board.

Most libraries responded that they **likely had some staff resources and infrastructure** to draw upon. Some of these resources were local, cross-campus, or at a state/consortial level. Examples of the kind of staffing available included IT staff, catalogers, and digital preservation staff. One (Tribal college) library had no local resources and potential

challenges connecting digital deposit with their shared catalog. This library said they wouldn't be able to do digital deposit locally, but would be interested at a consortial level.

Question 7: What do you see as potential FDLP requirements (conceptually similar to tangible materials or different in some way in terms of collection management, GPO support etc.)? What would be prohibitive to your library receiving files on digital deposit?

A majority of the libraries voiced a need for FDLP requirements to be **consistent, flexible, and relatively few**. Comments included, "keep things simple and consistent," "the fewer the requirements the better," "Keeping requirements consistent will be important to make it easier for staff," and "I don't know if there should be many requirements."

A number of libraries also wanted FDLP requirements to be **similar to requirements for tangible materials**, saying "would expect some of the same requirements as print" and "keeping it like the tangible program makes it easier to understand and doesn't have all new different requirements to track." Several libraries mentioned the dynamic between tangible and digital materials and that requirements for both should be in harmony. "Requirements would also need to depend on if there is a parallel process with print distribution" and "GPO would need to provide more guidance and assistance when it comes to collection management."

Several libraries mentioned the idea of a **tiered system that would offer flexibility** to libraries in terms of commitment. "Some would be fine with just a hard drive and basic hosting, others would do more" or there could be "access partners, preservation partners, etc." Libraries also noted that retention in a digital environment would "**need to be rethought because server space is different than physical space.**" Another notable comment was related to authenticity in a digital environment: "Would need to have some sort of connection between GPO's copy and distributed copies. Allow for provenance."

Some of the other specific comments regarding FDLP requirements were oriented around **digital preservation and access**, and tended toward one or the other. For example, a large library organization reported their "processes are geared around long-term preservation" and they are "willing to preserve long term." Some libraries "would like to not sign on to take material and preserve it all forever, potentially. Would want to be able to just keep the latest, for example." and one said "It would be nice to be able to remove items that aren't being used before 5 years."

In regards to the question of what would make digital deposit prohibitive to libraries, some libraries mentioned **technical challenges** such as server space, dealing with format obsolescence and migration, and access barriers such as "having adequate equipment to read and access the material." One library mentioned that they would need "more **support to get the program started**. To help coordinators get oriented, build this type of collection, do outreach in their community about this collection, etc."

Overall Observations

Digital Deposit is not just for one kind of library

The interviews were conducted during the covid-19 pandemic lockdown phase, and the real-life experience of remote library service provision most likely had some bearing on perspectives related to digital content. However, in our interviews with ten different organizations, each with their own institutional considerations and use cases for digital deposit, they all expressed interest and had ideas about how they would interact with a service. Additionally, most of the libraries indicated that they had access to some form of resources that would enable them to participate in a digital deposit service, whether those resources were local, or at a regional or consortial level.

Flexible selection is important to libraries

Based on this sample of libraries, a one size fits all digital deposit approach would not serve libraries' needs. Libraries expressed a desire to select digital materials to fit their own collection priorities, in a variety of ways such as topic, provenance, geographic range, format or other levels of granularity or time variables.

Bibliographic data is a requirement for digital deposit, and physical formats are not

The interviews provided an especially strong indication of specific requirements for two particular aspects of digital deposit. Responses showed that libraries view bibliographic data as inherent to digital deposit. We also heard strong opinions from our interviewees that they did not want to receive digital files on physical media.

A coordinated digital deposit service should be in harmony with existing FDLP model(s)

Based on the interviews, libraries see benefit in digital deposit being offered as a full program or service that includes not just deposit mechanisms but also coordination, documentation, training, and communications support from GPO. Libraries also

expressed a desire for an understanding of the dynamic between digital deposit requirements and existing FDLP requirements to be incorporated as requirements are developed.

Digital deposit could enable new use cases for Federal documents

Against the backdrop of the pandemic, the ability of a library to provide digital access is in the forefront of people's minds. In our interviews, discussions about expanding digital access stood out as an area where most libraries were looking ahead. The fact that a number of libraries were looking beyond traditional library catalog discovery, and brought up concepts of access such as computational access, seem to indicate a need to not "silo" Federal documents in bringing them to users.

Recommendations

1. GPO should undertake a pilot project with one or a small number of libraries to work out the details of a digital deposit service, in particular a service where GPO facilitates the deposit of digital files into the digital repositories of FDLP libraries. The pilot would:
    a. Use the findings from this group's interviews toward the project.
    b. Explore delivery issues (push vs pull, file types, etc.) that were brought to light by this group's interviews.
    c. Draft a high-level workflow for notification and push/pull delivery mechanisms.
    d. Document the necessary training requirements to support delivery mechanisms.
    e. Identify opportunities for technical solutions or enhancements to support delivery mechanisms.
    f. Report back to the Depository Library Council on the findings and outcomes of the processes explored in the pilot project.

2. GPO should take steps to outline a vision of how policy, coordination, and support for digital deposit could fit in with current and future models of the FDLP.

3. The Final Report of the Digital Deposit Working Group should be presented to the recently announced Task Force on a Digital FDLP.