

## 15 Amazing Facts about Digital Deposit You've Got to See to Believe and How They Could Change Gov Info Collections Forever – Audio transcript

---

[ Please standby for realtime captions ]

Hello and welcome to 15 amazing facts about digital deposits you have got to see to believe. I have some housekeeping reminders. Please use the chat box for questions, comments, and technical issues. Will keep track of everything and make sure they are addressed at the end of the presentation. Everyone will receive links to the recordings after the event. Please join me in welcoming Julia and James. Julia will start the session.

---

Thank you all for joining this working group presentation on 15 amazing facts about digital deposits you have to see to believe and how they could change government information collections forever. In addition to being a member of this group, I am also the librarian for government information, at Michigan State University. I am joined by the chair of the [ Indiscernible ] working group and a librarian at the University of North Texas and working group member James Jacobs, the government information librarian at Stanford University. I wanted to give you all a road back for our top. We know that this can be a bit complicated. It means different things to different people. We hope that you all will leave this presentation with a better understanding of what we mean when we say digital deposit and get you all thinking, to start thinking about the future of the FDLP and how it might become a service offer, as well as what role your library might be willing to play. Digital deposit has been the topic of regular discussion within the community for quite a while. The first appearance was in a survey in 2005. In 2009, the conference offered an early definition and workflows for the transfer of digital files. The modernization act of 2018 included as a service available for depository libraries optional digital deposit and I quote, the option -- they may receive directly from the superintendent of documents deposits in electronic format of information to disseminate X. In accordance with regulations under this subchapter -- in addition to all of this, informal discussion groups would convene at conferences. Because of these formal and informal thoughts to digital deposit, in 2018, the library counsel provided a recommendation to create a working group to explore current and future needs related to digital deposit, both dissemination of content and acceptance of content by GPO. Since early 2019, a group of DLC members, GPO representatives, and interested members of the FDL community have engaged in projects and discussions. They address the charge receipt by counsel. Over the past two years, the working group has identified a few amazing facts we would like to present to you today. We believe, if implemented, digital deposit could change government information collections. From discussion with the FDL community, it quickly became apparent a consensus did not exist on what digital deposit is and how it might work. Therefore, before the working group could develop an understanding, let alone communicated to others, we felt like we had to define it. To better illustrate the definition, there is a fairly simple graphic looking at the continuum of the work that it encompasses. For the FDL, for their world, should be a three-pronged approach. Flows from GPO to depository libraries, from agencies to GPO, and from libraries to the community to GPO. At the deposit library counsel meeting, the digital deposit working group offered this definition in their presentation. Digital deposit is practices, services, and workflows for the collaborative acquisition of born digital federal government information for our national collection of U.S. government information. It was not formally adopted or recognized by GPO but the working group feels this definition speaks to the broad what, how, and why of digital deposit as it applies to our work. I wanted to hand this over to James to discuss the messy problem and how to fix it.

---

Thank you, Julia. Yes, I have the ball now. Thank you, Julia. Yeah, hello, I am James Jacobs. Like ice cream, the Federal Depository Library community has long clamored for digital deposit. As Julia mentioned, the concept has been around since at least the early 2000's. I know I have been talking and writing about it for at least since then. They made their debut appearance as a question on the 2005 biannual survey. At that time, 33% of respondents indicated that they were interested in receiving digital files from GPO. Through the survey questions -- they changed over time and hundreds of libraries have consistently said in biannual surveys they were interested in this idea. 394 of them in 2005. 453 in 2007. 416 in 2009. 300 in 2015. You will get another chance in the next biannual survey coming up about digital deposits. Keep an eye on that. So it is clear that many libraries see a need for digital deposit.. I guess I had to -- sorry about that. Here we go. It should not be too much of a stretch to understand why digital deposit is so important to help grow the national collection of U.S. government information. In 2007, the GPO had a white paper on digital distribution that stated that during fiscal year 2006, 93% of all new titles were available through the FDLP in electronic format. The Internet has facilitated federal agencies publishing exponentially more public information every year. Therein lies the problem. More and more information is published online but almost none of it makes its way into the national collection. Aligned with that data point, the archive and other organizations have been working on the end of term project. This includes the GPO, Stanford, and the University of North Texas, represented on this panel. This archives the government web domain space. That is at the end of each presidential term, every four years. The outcomes for the 2016 crawl, as you can see on the slide, demonstrates the amount of information available online from the federal government. This is a huge amount of information. The 2020 crawl, which we just finished, they netted even more at around 300 terabytes of data. That is a lot of data. Publication used to be the work of one agency, GPO. Now it is going to take the work of many librarians and organizations to manage the entire lifecycle of public information from collection, description, and access and preservation. Preservation has always been a team sport, as we well know, and distribution has always been a key component of preservation to the FDLP . The FDLP has worked well as an access and long-term preservation mechanism for over 200 years. Because it consists of relatively simple collaborative process. They receive documents from government entities and catalogs indexes and deposits them in libraries and 1100+ libraries receive documents and provide storage access and public services for them to the public. The act of collecting, describing, giving access to and preserving documents remains critical to building the national collection. We now have to deal with digital and online formats rather than print and microfiche. In fact, it is probably disappearing in the next year or so. Digital deposit can be a key collaborative workflow and continuing to build and maintain the national digital collection. Lots of CAD -- I mean, copies, they keep stuff safe. Many libraries keep stuff safe. No one library or federal agency has the ability to collect, describe, and preserve everything. The agencies generally do not put much, if any, effort into preserving their public information. In fact, only around 1% to 3% of agency internal records are sent to NARA to permanent preservation. They do not see it is in their purview. As the network is necessary to ensure access to the historical documents, to everyone regardless of where they live, it is also necessary to assure that one of its members can recover from a devastating flood or other natural disasters. Redundancy is also critical in the born digital environment. In fact, it might be even more critical. Like -- they are more fragile than their paper cousins, who can be placed on a shelf and be safe for 100 years or more. How many of you have had a laptop hard drive crash and have been saved by your external backup drive? Imagine this happening on a countrywide or a global scale. Imagine if that pearl server crashed or if there were a government shutdown, where relying on a single point of access greatly increases the possibility of catastrophic failure or the entire network. We need many cats -- I mean, many libraries. So digital deposit can be a key tool and workflow for the care and feeding of the national collection. As Julia noted earlier, the concept of digital deposit has expanded to me deposit to libraries

but also to GPO from both libraries and federal agencies. That means the digital deposit can help the FDL DB with the issue of unreported documents. Unreported documents have been around since the beginning of the FDL . Every year, staff find thousands of unrecorded documents and received thousands more from the FDL community. In the last six months, 2072 unreported publications were catalogued from April to September of this year. Of those, 35% were results of depository libraries know to fighting them. Digital deposit from libraries to GPO and from agencies to GPO could greatly expand and enhance the national collection. After all, that is what we all want. Now I am going to pass the ball to Robbie. If I can. take it away, Robbie.

---

Can you hear me?

---

Yes.

---

All right. Okay. The digital deposit working group new from its inception we wanted to conduct a pilot. Initially, ambitions have led us to hopes of crafting them for the three workflows related to digital deposit to libraries, from libraries, and from agencies. They wanted to recognize three pilots would be unattainable and could be an overreach of our charge. We focused on a pilot that looked at the library to GPO model of digital deposit. Many are already engaged in this work, whether they realize it or not. Most of you might know this better as reporting lost documents and unreported publications. Once the working group decided on the workflow and the general framework, we had to scope it. The University of government information is pretty big and and pecking for unreported documents without a well-defined scope could prove time-consuming, tedious, and largely unmeasurable. As a way to provide structure to potential pilot volunteers, we decided to scope the project to this mandated report. This provided us with a list of materials required by Congress that, in theory, could be searched for and discovered and would allow the working group to apply metrics that would offer a number of data points. This brings us to our next amazing fact. In September 2020, the digital deposit working group sent out a call for fire tears . 18 individuals volunteered to join the project and embark on a search for congressionally mandated reports. Between October and December 2020, these 18 volunteers collectively spent 12,842 hours of time on the project. That was a very exhaustive amount of time searching for and discovering congressional mandated reports. The outcomes of the thousands of hours and volunteers spent searching is the focus of amazing fact number eight. Volunteers executed searches for 1049 congressionally mandated reports. From them we learned over half are still lost. This is likely the result of the agency not yet fulfilling their statutory requirements. The best news of the statistic is that 430 reports are now found and have been reported to GPO. 88 of those searched were never lost at all. The working group also hosted focus groups to learn how and if the community regularly searched for on catalogued materials and how, if they did, report those on catalogued materials to GPO. An exciting and surprising fact came from these focus groups and that is that the community is not overly familiar with searching for unreported and on catalogued materials. I thought they would all know that. I thought they would like to learn more and to make this part of the regular routine work as a depository coordinator. The working group also wanted to learn more about how ask GPO functions as a reporting tool for unreported publications. He hoped -- they allowed the working group to ask some comparison questions of participants who were familiar with reporting materials in the pre-and post askGPO on the new platform, which I had in my nose and I cannot remember what it is called. Previous discussions, we learned that community supports the use of reporting it. Those familiar with both iterations of askGPO complemented many of the new features within the products. That includes the new categories , automated responses , and the ability to attach files this is an especially useful aspect for those that added items to a spreadsheet or word document to allow for reporting in batches. When asked about potential improvements to askGPO, the main suggestion was a fielded form to ensure the key parts of a

catalogued item was reported. Now I am going to hand it back to Julia. Can I make my mouse work?  
There we go.

---

Thank you, Robbie. To explore the more philosophical aspects, the working group formed a sub group called digital lysing digital deposit. To accomplish that, they developed eight open-ended questions and interviewed representatives from 10 cases. They were law, public, tribal, center of excellence, preservation stewards, and other portals. All of our interviewees sported additional deposit and they were all receptive of the idea. Much to the surprise of the subgroup, no one said they could not participate due to lack of resources. All interviewed organizations stated that they had technological and personal resources that they could utilize for the purpose of digital deposit, either at their institution or through consortia. Involvement in the process vary depending on organization type. For example, some were willing to serve as hosting and digital preservation partners while others send participation more limited to organizational limitations. When it comes to accessing the delete deposited content, they offer creative ideas for access beyond the traditional catalog model. While the idea of locally hosted content -- they provide action on [ Indiscernible ]. They also suggested more creative non-catalog ways to provide access and increase discoverability. These include providing explanatory or promotional text to the digital content on the library website as well as collections as motive access model. Allowing users to browse by [ Indiscernible ] or curated collections. Three of the libraries we have mentioned use the digital content of computational or experimental use with an emphasis on presenting the documents in ways that the late understanding by the general public while making use of computer science techniques to extract features of the document. For example, text files for government publications. The creation of a research -- it was also mentioned by one library to give users a space to middling content in a low barrier access way. Through the interviews, we found flexibility was key and strongly desired in terms of content types, formats, and ability to select materials. The content types -- they were all file types. A variety of them will allow for nontraditional catalog access and the use of these materials with a number of libraries specifically requesting X amount of net files. Not only did they want a variety of file types but they also wanted the ability to request all with that format. Check a box and get all the XML files regardless of agency or other factors. There was also strong insists with subject areas like legal materials or materials related to environmental issues. Geographically focused selections also were desired. Some could receive all materials that covered specific regions or geographies. The interviewed librarians were clear that what they wanted most was an easy, flexible selection process that also made it clear what the library will be receiving through the digital deposit selection without having to do additional research or digging to find out what is included. Selection methods suggested selecting all of the format type topical subject [ Indiscernible ] by geography, agency, and selecting from collections by using a group interest area or library type, such as collections for law or public libraries. Just because there was a strong interest in digital deposit did not mean that everyone would abandon tangible items. These libraries indicated they would continue to select tangible formats as well as receive digital content, as they intended to use the digital to offer new services or in a different way compared to tangible items. As part of this extraordinary investigation into digital deposit, we asked these libraries if they had a preference as to whether they would receive push models or getting it themselves. They were split regarding their preference. Push would be similar to how they receive materials from GPO in tangible format. GPO will gather and prepare the material and give it to libraries. The library receives the items and does some processing activities to make them available to their patrons. Libraries that expect interest in the push model stated that they wanted these items to come with metadata to accompany the digital files, so that the objects will be easier to work with. Interest was expressed in making this a two way street, not only with GPO pushing content to libraries but libraries pushing digital content to GPO for dissemination. For those who are interested in a pull model, when they would go to a site and actively download the content found there, they requested that GPO would send notification that it is

available for retrieval and then the library staff would go and pull those items from that site. If it is a pull model, being able to control when the material is brought into the library site and with that, working to notice if there are errors or issues with the content files quickly. More investigation into the feasibility of these models will need to occur to determine what works best for GPO and the community. The work done by the visualizing subgroup was dependent upon discussing this topic with volunteers from the community and we are so thankful and appreciative of the community as we have no trouble getting 14 volunteers at an organizations to share their opinions with us during our interviews. To find volunteers by organization type, we targeted different types of libraries that indicated interest in digital deposit in the biannual survey. We know that we cannot interview everyone and we acknowledge each one is different, we try to to select libraries of different sizes and from a variety of regions around the country to provide a fuller perspective. One thing to note, this is hour bonus fact, is that our interviews were conducted during the lockdown phase of COVID-19. That may have influenced our responses. Libraries might have a traditional tangible focus but were suddenly met with unprecedented demand for digital only content to reach patrons that are no longer able to physically come to the library. This real-life. May have influenced them to focus on digital content acquisition and being enthusiastic to appear prior to the pandemic, they might not have thought this was an area of interest to their patrons. With the pandemic, the book I used to digital content and remote access to materials. We think it will continue, even as things return to pre-pandemic levels of access and service. I will hand this over to James for a recap.

---

Thanks, Julia. I have got the ball. This time, I'm going to move the slide forward before I start talking about the slide. So just a quick wrap up of what we discussed on the panel today. Through the digital deposit working groups work over the last two years, cannot believe it has been two years, we have affirmed for both research and discussions, as well as a pilot project, that, one, digital deposit is a really popular idea and, two, digital deposit could be a key workflow to the care and feeding of the born digital national collection and, three, there is a place for digital deposit in the ongoing work of libraries, no matter their size or their type of communities that are served. Digital deposit working group coalesced around three subgroups in our work. One group explored the philosophy and the concept and the devilish details of digital deposit with various library types. That was the visual deposit work group. The focus groups pull together volunteers from across the program to take the pulse of the community on GPO's current mechanism to see if it is working and how it could be improved upon. And then the third group did -- they ran a pilot to work through the unreported documents processed. This was seen as somewhat urgent for the community, to get a collective head around it, due to the reintroduction of the access to congressionally mandated reports act in April of 2021. I am at least hoping that it will pass this Congress , although it has been introduced in, I think, three past congresses but has not been passed yet. I would like to give a huge, huge shout out to all of the working group members. These people have put a lot of time and energy and brainpower, as well as quite a bit of good cheer and camaraderie, to the working group. It has been a pleasure working with everybody and I would like to give a tip of the hat to Robbie and Cindy, who have led the group and kept them on track and focused all this time. Now I will pass it back to Robbie. Ready, Robbie? There it goes.

---

Thanks, James. I am blushing a bit.

---

You cannot turn your camera on.

---

Let me -- there you go. We also want to definitely give a huge thank you to all of our volunteers without them, we could not have done our work. We want to thank the 14 individuals from varying types of libraries and organizations, who met with our deposit subgroup. These folks helps us learn more about

what this work might look like within individual libraries and what the value would be to their communities and what some of the obstacles are to fully engage in this work. We also think the 18 volunteers who spent thousands of hours developing searches and reporting back on their processes and workflows. We thank them for discovering so many new unreported congressionally mandated reports. Finally, we want to thank the folks who participated in our focus groups, not noted on this slide are folks that participated in our 2019 conference session. All of these individuals, regardless of the role, have provided information and insight that guided the digital deposit working group and have led to our project outcomes and recommendations. We could not have fulfilled our charge without your time and input. Thank you. Through the course of our work, they have evaluated and reflected on the feedback we received and the trends we observed in the work of our volunteers. Our evaluations and reflections are embodied in these four recommendations, which we plan to submit to counsel with our final report. We are leaving ourselves open to additional recommendations that may surface as we finalize our reports. We hope that the work of the group and the subsequent recommendations impact future activities related to digital deposits, which will assist with the access and preservation of born digital government information. I will not read the recommendations. To end things and before we open the floor for Q&A, we would like to request your participation in a short poll. Laura, if you are ready, will you bring it up?

---

Absolutely. I will -- you should see a poll come up.

---

Thank you.

---

[ Silence ] I will give it a couple more minutes for anyone else who wants to answer the poll.

---

[ Silence ] it looks like you have some good poll responses. I will close the poll. It will take a couple minutes to compile the responses and then you will see the recap.

---

Thank you. I was trying to go forward. There we go. Questions, if you all have them? We would love to engage in discussion or any questions you might have.

---

One question has come in from Andrea. Where can we read the four recommendations? Were they listed on the slide?

---

Oh, yes. Did I not put that slide up? I am sorry. I am working with two screens. I am a ding . It looks like from our poll results, we still have some work to do in bringing more people into the digital deposit fold but I know this group is determined to get that done. We will formally submit these recommendations to counsel and when we do, they will be in the public record and then hopefully counsel will take them to GPO as potential recommendations.

---

Another question has come in. What is your definition of unreported publications? This is from Canada.

---

I think that is a definition provided by GPO. Let me see if I can find it. I will put something in the chat that includes the official GPO definition. This was updated this month, actually, so that is kind of a yucky URL but hopefully it will get to you and then James also added it to the chat. Thank you, Megan. A good question.

---

To be clear, digital deposit does not deal with digital conversion of tangible materials. It is really just looking at born digital content. Yes, there are concerns in the community with regard to digitizing

content that has either personal identifying information or things that may be compromising of security issues. I do know that GPO embarked on a big reduction project with congressional records. We have done some of it here with our digital collections. Yes, a good question. It is not completely related to our topic. We would say that the poll method is different and I invite my other panelists to share this question as well. To me, they can be a very laborious and cumbersome project to go through by title and title. I would say that a poll method would be where perhaps where you get a selected box of material, just like you would get sort of a selected package of born digital content that you could pull into your digital repository.

---

Mary does make a good point, if I may jump in. The new electronic titles could be a way, if it was not just, you know, here is the list of new electronic titles for the last month, if there was away that they could be converted into more of a machine driven process, where here is the list of the new electronic titles and press a button if you want to get all of them or if you want to get only the ones from a certain agency, et cetera, et cetera. New electronic titles is a process that is already out there and it could be worked into a future digital deposit workflow.

---

Yes, and to -- there is a good follow-up question. Yes, the goal would be to [ Indiscernible ] into multiple repositories rather than just getting catalogued records that point out to other places. So kind of lots of copies but in a born digital world. Good questions. James, do you want to answer Amy's question?

---

Sure. Let me see what it is. I was pasting something into the chat. Yes, that includes metadata as well. Digital files are nothing without metadata, I would say. That is going to be key to doing this. Metadata is important for traditional library catalogs but also important for doing other kinds of digital projects, like corpus analysis or any of these newer forms of digital humanities and digital social sciences research.

---

And to answer Christine's question, the pilot that we have added as a recommendation, we would have to have that recommendation accepted by counsel and then taken to GPO, with their willingness to engage in a pilot for digital deposits. It is TBD.

---

Mary has another question there can be considered a hybrid system, where -- [ Indiscernible ] the whole document on an institution site?

---

I would think that is part of it and that goes back to the question of the major data. Yes, ideally, they live on a library's institutional repository site and contained in that would be all of the metadata that shows the origin and the [ Indiscernible ] of that individual item. I hope that answered the question. Do you want to chime in?

---

I think you answered the question.

---

I want to time in real fast and say if a library does not itself have the ability to have these materials hosted, they did indicate there are interviews where they would tap into consortium will repositories and maybe -- [ Indiscernible ] might have these copies on their local server and then they could tap into that and link to it.

---

A good point. I just lost my chat.

---

They have a new question that came in. I can give them to you. Are born digital hubs reported to GPO? Now I lost my chat.

---

Let's start again. Are born digital hubs reported to GPO and are the records available for discovery systems with pointers to fulltext? Is that included in the [ Indiscernible ] that has been referenced?

---

I would defer to GPO to answer that question. Cindy or Megan, can either of you answer that question since you are on the call? Cindy says get help. Is the practical guidance available as to how much storage space might be needed and time commitment from library staff, et cetera? I am interested philosophically but would have to be able to make a solid case for necessary library resources.

---

We have to address all of those larger issues with that proposed. Some of those questions could be answered through an actual digital deposit Pilate.

---

Yeah, just to clarify, get out is where one place where you can get the catalog records. That is what Cindy was meaning.

---

I do not know if that answered the question or not. Maybe what we could say is that digital publications are reported to GPO as unrecorded documents, for example and GPO does attempt to reach out to agencies to receive their digital documents . However, not many agencies do, at this time, send their digital documents, quote, unquote, to GPO. They do things like web archiving of agency sites and things like that. When a digital document is reported to GPO, they either point to it, if it is on a live site, something that is hard to collect digitally, or if it is one PDF and they save it to their digital repository, which is permanent access pick you might have seen that URL. There is lots of different ways that digital publications would have described and whenever GPO makes the catalog record, those are sent out either through GPO's -- if you subscribe to the archive, the documents for the service, you will get records to the digital documents. That is different from digital deposit. Sorry for the long-winded answer.

---

Yes, I agree. Space will always be a problem and I would say if we are able to launch a pilot with regards to digital deposit, it would be wonderful to have multiple types of libraries willing to volunteer to participate those that have limited space or limited availability.

---

I would say just as context, the project here, which I facilitate and run at Stanford, we require our locks partners to go -- [ Indiscernible ] they have three terabytes of storage required, which is not that much storage and we collect all of the documents on the info. We still have not but up against that three terabytes maximum storage. If you are talking about web archiving, they do collect a lot more data compared to just the digital files and the PDFs and that metadata.

---

[ Captioners transitioning ]

---

It is going to be somewhere in between depending on what your library wants to connect to and those types of issues.

---

I do just want to address the question quickly. The ability to manage and interact certified or official publications. This is something that would follow a similar model to digital content partnerships or preservation stewards and another version of a formal agreement between the library and whatever would be required of that in helping to assert that it is unofficial resource or copy but that can also be addressed in metadata. Hopefully pick

---

All right. We over? I want to be respectful of those coming behind us.

---

We are good. We have another 15 minutes or so.

---

Sorry.

---

Does anybody have any more comments or questions for Robbie and Juliet and James?

---

Thank you for attending and for all of the great questions that you did a lot to I think forward the conversation and I am very appreciative.

---

Thank you .

---

Okay. If there is not more questions are like to thank our presenters today for such a wonderful program. Up next in this room we will have global information landscape, a comprehensive overview of the professional report and or other virtual meeting room the next program is GPO introducing the serials at collection and if you want to join that program you have to close out of this room first and use the other URL to enter the meeting room which is found on the event page and for now it will take a short break and start again that 4:15 p.m.

---

[ The event is on a recess. The session will reconvene at 4:15 p.m. Eastern Time. Captioner on standby. ]

---