Weeding Digital Government Information Resources

Considerations and Strategies

Paul Riermaier, Will Bandoma, Sue Gagnon, Janet Marler, Sandra Standish, and Victoria Turner

Weeding is a systematic approach to the removal of resources from a library's collection. In the weeding process, materials are identified for withdrawal in order to maintain a collection that is accurate, updated, well-used, meets the needs of the users, and is in line with the library's mission. When weeding tangible resources that are part of the Federal Depository Library Program (FDLP), a depository library must ensure that its weeding policy follows the Legal Requirements & Program Regulations of the Federal Depository Library Program and any separate guidelines set by the Regional Depository. However, there are no specific rules or guidelines to follow when weeding digital FDLP resources. This means that individual libraries have more leeway to craft digital weeding procedures that best serve their institution, patrons, and the community at large. In this article, we will discuss initial considerations when developing a process for weeding digital depository materials, we will examine different methods for analyzing a digital collection's size and usage, and we will review methods for maintenance and weeding of digital resources.

Libraries in the Federal Depository Library Program (FDLP) act as stewards for the government documents in their collection. Physical space constraints and other resource limitations have long impacted decisions on holding tangible materials both in a library's general collection and its depository documents.[1] While the space required to house print and other physical materials provided functional limitations on collection size and required regular maintenance and curation, the same limitations do not exist for digital items. Absent such external pressures it is easier to neglect maintenance of the digital collection.

However, regular maintenance and curation are part of providing users with a useful and accessible digital collection.

The need for regular maintenance of a library's digital holdings extends across not only its general collection but also to its government documents collection. Initial considerations when developing a process for weeding digital depository materials at selective depository libraries include examining size, usage and maintenance.

Background

As library collections evolved to include digital materials alongside tangible materials, library processes have had to evolve to accommodate this shift. The shift to providing digital materials alongside, or in lieu of, physical materials was driven in part by the efficiencies inherent to digital materials and in part by user preferences. A 2006 user study conducted on the University of Montana campus showed that "although some respondents prefer printed materials for specific types of information, users generally want to find and access government information online."[2] The ease in which documents can be produced digitally has given way to more documents being produced than ever before, and "very much greater than the earlier production of printed government information."[3] In 2013, the National Academy of Public Administration reported that "97 percent of all federal documents are 'born digital.'"[4] Relying on that 2013 report, Representative Rodney Davis remarked in his opening statement at the Hearing for the Oversight of the Government Publishing Office (GPO) on March 3, 2020 that "most users expect to use those documents in their digital form."[5] In addition to born-digital documents, libraries have expanded their offering of digital versions of print-born documents.

While there are still work flows and procedures that must be followed, digital federal documents can be added to a depository collection more seamlessly than tangible resources.[6] This

also means that digital documents can more easily remain in a collection, unused by patrons and unreviewed by a depository librarian. Whereas space constraints, large shifts of materials, and other circumstances might prompt regular review of a tangible collection, those same pressures do not impact a library's digital collection. Yet, libraries still "need to continue to develop, analyze, and maintain [their] online collection."[7]

When weeding tangible resources that are part of the FDLP, a library must ensure that its weeding policy follows the *Legal Requirements & Program Regulations of the Federal Depository Library Program* (*LRPR*) and any separate guidelines set by the Regional Depository.[8] However, there are no specific rules or guidelines to follow when weeding digital depository items.[9] This means that individual libraries should consider how to best serve their patrons, institution, and community at large when crafting procedures for weeding digital depository items and ensure that those policies comport with the FDLP's mission of providing free and accessible government information to the public.

Approach

The lack of rules in the *LRPR* regarding digital weeding provides libraries with flexibility to craft policies responsive to their library's specific needs. However, without any basic guidelines to begin with (assuming there are no policies instituted by the Regional Depository), it can be daunting to craft a strategy from scratch. When developing policies and procedures prior to conducting an FDLP digital weeding project these principles may be considered:

- Digital depository collections should be developed to meet the needs and format preferences of primary library users and the general public.
- A digital depository collection should be curated so that results from catalog searches are not overwhelming.
- Methods of measuring digital depository item usage should be implemented.
- Statistics on resource usage should be maintained and reviewed in order to evaluate collection relevance and usefulness to patrons.
- Routine maintenance of the digital depository collection should include:

  ○ Identification of broken or dead URLs and replacement or removal.
  ○ Updates reflecting changes to an institutional author such as an agency's name or its role.

Development, Curation, and Weeding of Digital Depository Collections

Like materials in a library's general collection, materials in a digital depository collection should be developed in accordance with the needs and preferences of users. FDLP libraries, which include specialty and academic libraries, may have distinctions between their primary patrons and their depository patrons that need to be taken into account. A library's general collection may be developed solely to serve its primary patrons, but participation in the FDLP means that the depository collection must be freely available and serve the needs of the general public.[10] For many document types, users prefer to access a digital format.[11] As digital collections grow in response to user preferences for digital access, regular weeding can help ensure the collection is visible and usable for patrons.

A commitment to serving user needs and preferences does not mean unchecked expansion of digital materials. Athina Livanos-Propst stressed in the article, "Developing Weeding Protocols for Born Digital Collections," that weeding projects are necessary to maintain a collection that continues to serve the needs of library patrons.[12] To that end, Livanos-Propst identified several principles that should inform weeding decisions, two of which were especially pertinent to government document collections: (1) "Accuracy: Is the information in the resource technically accurate and factually correct?" and (2) "Usage: How often has the resource been viewed?"

Regarding accuracy, digital content produced or created by GPO, should, whenever possible, include the GPO seal of authenticity. "The widespread use of digital technology has changed the ways GPO's products are created, managed, and delivered to users. Because many of the official publications GPO provides online are in PDF format, GPO uses digital signature technology to provide evidence of authenticity and integrity and safeguard against unauthorized changes to these files."[13] One should note that the seal of authenticity may not apply to materials that originate outside of the GPO.[14] This does not mean that links to documents without the seal should be automatically weeded; rather, links to unauthenticated documents should be replaced with links to documents that carry the GPO seal of authenticity. Additionally, all future selections should opt for authenticated documents over unauthenticated sources.

Review of a digital collection should also focus on identifying links that are broken. The FDLP provides stable URLs for many digital items in the form of Persistent Uniform Resource Locators (PURLs). One advantage to PURLs over URLs is that GPO reviews and maintains PURL links for accuracy and functionality which should reduce the amount of broken links

in the collection.[15] GPO's commitment to maintaining PURLs makes it is less likely that a PURL link will become broken. However, broken PURL links may still happen and a digital collection will likely consist of both PURLs and URLs. Consequently, libraries should use a link validator tool to crawl the PURLs and URLs in their collection to identify and correct or remove broken links.[16] Broken PURLs should be reported to GPO for correction.

Digital depository materials can be made accessible to users through LibGuides and other discovery tools, including the library catalog. The increase of government documents, especially those that are born-digital, can mean that a simple catalog search recalls a tremendous amount of resources. The flow of government information and digital materials is unlikely to subside, but a careful review of digital items in a library's selection profile can help prevent clutter in a catalog. Weeding electronic resources can also reduce overwhelming results and focus a collection on those materials that are "vital, vibrant, and up-to-date."[17]

Reviewing Collection and Usage Statistics

How can libraries determine the digital depository documents that their patrons need and that keep their collections vibrant? Measuring usage of materials has been a key part of the weeding process for tangible materials and continues to play a part in the digital weeding era. Libraries can use circulation and in-house usage statistics for tangible documents but measuring the usage of digital items in a collection is not altogether straightforward. There are several different strategies that may be used, and the best strategy for a particular library will be dependent on past practices of how their collection has been processed and maintained. Staffing consideration should also be considered, as some strategies are more time intensive to set up.

For many libraries running a report using their integrated library system (ILS), with certain customizations, will be the most practical method for measuring usage. For example, one method of measuring usage of digital resources this way is to generate an ILS report, sorted by the MARC 074 GPO item number field, which could then be used to target titles in certain subject areas. GPO has also suggested running ILS reports of the MARC 856 (PURL), 040 (GPO) and 086 (SuDoc) fields as well.[18] Reviewing the results of these reports would help in identifying content that could be weeded. Care must be taken though in looking at the age of a record, as older records may still contain current information. Unlike fixed tangible resources, digital resources evolve as agencies continue to add more digital content to their websites. Deleting a record, therefore, may result in the discard of more information than expected.

A second approach suggested by GPO for weeding digital resources is to insert a standardized code, such as "REVIEW-2025-QTR.4", into the catalog record of a resource. This code could be inserted upon the record's import, into a local 5XX or 9XX MARC field (locally defined note fields), identifying that the resource be reviewed in a certain number of years. One advantage of this method would be that smaller sets of records would be able to be reviewed on a regular basis. However, this coding must be done when the record is entered into the catalog, not when the report is set to be run, meaning that planning and consistency is required for this method to provide meaningful results. The reviewer must be cautious to review content for collection needs, and not go by the age of the record. Some records may be old, but they are still current, such as with serial records where content is added periodically. Also, the reviewer must be aware that some items may have been superseded, but the newer document will not show in the results having been cataloged in a different date range not captured by the search.[19] More discussion on superseded documents can be found in the Personal Experiences with Weeding section below.

The GPO's PURL Usage Reporting Tool can provide insight into how much traffic a library is directing to a specific PURL.[20] PURLs are located in the catalog record of the resource in the MARC 856 40 or 856 41 fields (electronic location and access field). When patrons use a digital resource, the data from their selection of the PURL link is recorded with the FDLP. Depository libraries can receive reports on this PURL data to evaluate which digital resources patrons are utilizing from library-specified websites, databases, library guides, and other locations. You can access the PURL Usage Reporting Tool through the FDLP.gov website under the Collection Tools tab. Running this report and viewing the usage regularly allows depository libraries to understand which digital resources are being utilized and which can be weighed in their weeding decision making. Users of this tool should keep in mind that prior to 2008 GPO used a single record to catalog all formats of a title, so the PURL could be located on the print or microfiche record. After 2008 PURLs are only included in online resource catalog records. The PURL Usage Reporting Tool is easy to set up and begin receiving usage reports. The disadvantage is that the tool will not capture URL usage. It is possible to develop methods to monitor all click throughs of online resources in a catalog, but for initial steps or short-staffed institutions the PURL Usage Report may be much easier to manage.[21]

Personal Experiences with Weeding Digital Government Documents

Depository librarians on the GOVDOC-L listserv were asked for advice in regard to undertaking a digital weeding project. Their responses ranged from the practical to the theoretical and highlighted the different considerations individual institutions must undertake when contemplating or prioritizing a digital weeding project. One librarian stated that "any weeding project I'd do would be focused on tangible collections due to space considerations. I probably wouldn't contemplate any kind of organized digital weeding until much of the print had been reviewed."[22] In a similar vein, another librarian responded, "What is the reasoning for deleting these links?"[23] Diaz suggests that, "updating born digital documents is not as time consuming as most people think once you get the knack for it." He offered the following for consideration:

1. If you have an URL for a document, replace it with a PURL. That way GPO is in charge of updating the link.
2. If you find a PURL that is not working, look for the link on the document's website, copy the new link and report it to GPO and include the PURL with it so they can update it in the system.
3. Some libraries have link checkers. Make sure the IT department sends the list to you every month.
4. The reason most links go dead is because the website (agency) has redesigned their webpage and moved it someplace else. If they did remove it from the website and you think it's important, make contact with the agency through email. Also, report it to the GPO. They can persuade them to put it back on the webpage or GPO can archive it themselves.
5. Linking to the Internet Archive is better than not having it available at all. This is one of (his) favorite places to find deleted titles.[24]

The most detailed response came from a retired librarian, Michael McDonnell, as a private email to the author. He suggested a collection maintenance project to remove superseded documents. Using electronic resource catalog records received from MARCIVE, McDonnell and his staff created a report:

> With these fields: SuDoc number; Title; OCLC record number; and the Alma/Primo location code that tells us if a record is for a tangible, or electronic title. A student employee scanned through the Excel file and noted any titles that might supersede earlier editions.

After possible superseded records are identified, students searched the catalog for earlier editions. Search results and complete records werea printed out for all matching records including the newest. The librarian reviewed the printouts and decided which records to delete. Our documents holdings are recorded in the OCLC database so, when we deleted a record from our catalog, we also had to delete our holdings from OCLC.

Sometimes earlier editions were housed outside of "Documents." Sometimes superseded editions had different SuDoc numbers. This could be because of a classification error or because they were issued by another agency. These and other issues took up more staff time. Some SuDoc number inconsistencies required waiting for answers to AskGPO inquiries that held up the process.

My documents department was tasked with reducing the physical footprint of our collection. This is one reason we undertook this project. Most superseded documents are small. We did not gain much space doing this. However, we did add to our title count of discarded documents each month. Not having records and documents containing possibly outdated information was another benefit of the effort.

If you have an electronic only collection you have to ask if you have the staff time to take on this task. At first you might find a large number of superseded documents as you clean out multiple earlier editions. But the number will fall as you continue the process. My opinion is that the government does not issue as many of these superseding titles as they did in the tangible era. Some have been replaced with continually updated web pages. Your selection profile might also make this task unproductive. And, if you are working in another environment, you might not be able to use the process as described here."[25]

The inquiry to GOVDOC-L prompted a fair degree of skepticism about the value of digital weeding. Some respondents shared that it can be an onerous undertaking given the volume of digital resources a depository collection may contain. For instance, the Electronic Resources Specialist at Thrall Public Library reported that there were over 200,000 electronic government resources. More generally, articles have cited the effort required, or the decisions made in managing the volume of electronic records needing review. For example, Livian-Probst discussed the decisions made in reducing the reviewing

of 100,000 electronic records to just under 15,000 records. These decisions included focusing only on records from specific years, as well as certain types of resources.[26] Similarly, in the PowerPoint presentation *To Weed or Not To Weed: Managing An Electronic Collection*, Elizabeth Psyck humorously acknowledged that the process is "labor intensive and not particularly pretty." She was able to address the weeding of electronic documents only because her library had the "best staff and students ever."[27]

Conclusion

Depending on a library's current collection status and past practices, implementing a digital weeding process may require rethinking workflows and breaking the project into stages. However, it is still possible for depository libraries to put in place policies and procedures today that will allow for effective and worthwhile digital weeding in the future. As mentioned earlier, adding a date indicator to a record when it is cataloged would allow for more concise reports to be generated. Similarly, reports that target the MARC 074 GPO item number field would permit review of key subject areas.

Ultimately, weeding a digital collection and a physical collection share the same goal: provide users with resources that are helpful, easy to find, and relevant to their interests. The general principles that are used when weeding print materials will also inform decisions made when weeding digital items, but the specific processes are different. When first starting a digital weeding project those processes can be labor intensive or technically difficult to institute. The resources described in the annotated bibliography can help with the process of setting up a digital weeding system. Once the initial investment of time and effort is made in establishing a system for digital weeding the regular maintenance and review of the digital collection can be simplified. The result will be a digital collection that matches the print collection in usefulness and relevancy.

Paul Riermaier (prierma@law.upenn.edu), Reference Librarian, Biddle Law Library, University of Pennsylvania, Philadelphia, Pennsylvania. **Williams Bandoma** (bandoma@portlib.org), Business & Government Librarian, Portland Public Library, Portland, Maine. **Sue Gagnon** (sgagnon@anselm.edu), Head of Periodicals, Geisel Library, Saint Anselm College, Manchester, New Hampshire. **Janet Marler** (jmarler@ocean.edu), College Lecturer II—Library, Ocean County College, Toms River, New Jersey. **Sandra Standish** (nicke1sl@cmich.edu), Access Services Specialist, Central Michigan University Libraries, Mount Pleasant, Michigan. **Victoria Turner** (thrall2@warwick.net), Government Information Librarian, Middletown Thrall Library, Middletown, New York.

Notes

1. Celina McDonald, "Basic Project Management for Weeding Government Documents Collections," *DttP: Documents to the People* 44, no. 3 (2016), https://journals.ala.org/index.php/dttp/article/view/6120/7894.

2. Jennie M. Burroughs, "What Users Want: Accessing Government Information Preferences to Drive Information Services," *Government Information Quarterly* 26, no. 1 (January 2009): 203.

3. James A. Jacobs, "Born-Digital U.S. Federal Government Information: Preservation and Access," (background paper prepared for the Center for Research Libraries Global Resources Collection Forum: Leviathan Libraries and Government Information in the Age of Big Data, Chicago, April 24–25, 2014), https://www.crl.edu/node/10225.

4. National Academy of Public Administration, *Rebooting the Government Printing Office: Keeping America Informed in the Digital Age* (Washington, DC: National Academy of Public Administration, 2013), http://purl.fdlp.gov/GPO/gpo34038.

5. Rodney Davis, Committee on House Administration, "Opening Statement," *Hearing on the Oversight of the Government Publishing Office*, 116th Cong., March 3, 2020, https://cha.house.gov/committee-activity/hearings/oversight-government-publishing-office.

6. U.S. Government Publishing Office, "All or Mostly Online Federal Depository Libraries" (June 6, 2019), https://www.fdlp.gov/requirements-guidance/guidance/2124-all-or-mostly-online-federal-depository-libraries.

7. U.S. Government Publishing Office, "Transitioning Depository Collection to an All Online Depository" (February 4, 2015), https://www.fdlp.gov/guidance-article/transitioning-depository-collection-to-an-all-online-depository.

8. U.S. Government Publishing Office, *Legal Requirements & Program Regulations of the Federal Depository Library Program* (Washington, DC: Government Publishing Office, February 2018), https://purl.fdlp.gov/GPO/gpo89707.

9. Ashley Dahlen and Melissa Fairfield, "Weeding Print & Digital Resources," Session, *2019 Federal Depository Library Conference*, Arlington, VA, October 22, 2019, https://www.fdlp.gov/fall-2019-federal-depository-library-conference.

10. *Legal Requirements & Program Regulations*.

11. Peter Hernon and Laura Saunders, "The Federal Depository Library Program in 2023: One Perspective on the Transition to the Future," *College and Research Libraries* 70, no. 4 (July 2009): 351–70, https://doi.org/10.5860/0700351.

12. Athina Livanos-Propst, "Developing Weeding Protocols for Born Digital Collections," *Code4Lib Journal* 43 (2019), https://doaj.org/article/089c052c6fab48ae8b5003f45af4c1be.

13. U.S. Government Publishing Office, "Govinfo: Authentication," https://www.govinfo.gov/about/authentication.

14. U.S. Government Publishing Office, "Document Authentication Workshop," (July 18, 2010): https://www.govinfo.gov/media/authentication-workshop.pdf.

15. U.S. Government Publishing Office, "Persistent Uniform Resource Locator (PURL): Explanation, Purpose, and Tracking Usage at Your Library" (June 13, 2019), https://www.fdlp.gov/requirements-guidance-2/instructions/709-purls.

16. Donna Kraemer and Ashley Dahlen, "Intricacies of Digital Content and PURLs" (October 22, 2018): https://www.fdlp.gov/fall-2018-federal-depository-conference.

17. Juris Dilevko and Lisa Gottlieb, "Weed to Achieve: a Fundamental Part of the Public Library Mission?," *Library Collections, Acquisitions, and Technical Services* 27, no. 1 (2003): 73–96. https://doi.org/10.1016/S1464-9055(02)00308-1.

18. Dahlen and Fairfield, "Weeding Print & Digital Resources."

19. Dahlen and Fairfield, "Weeding Print & Digital Resources."

20. Users of the PURL Usage Reporting Tool must register a set of up to 10 hostname or IP address patterns (full or partial) for their institution to receive monthly reports of all PURL referrals from those hosts. Data is compiled through the previous day for the current month. Historical data is available for a maximum of 12 months.

21. Christopher C. Brown, "Knowing Where They Went: Six Years of Online Access Statistics via the Online Catalog for Federal Government Information," *College & Research Libraries* 72, no. 1 (January 2011): 43–61, https://doi.org/10.5860/crl-68r1.

22. Anonymous, Email message to author, November 4, 2019. Used with permission.

23. C.A. Diaz, Personal communication on listserv, GOV-DOC-L@LISTS.PSU.EDU, November 4, 2019. Used with permission.

24. C.A. Diaz, Email message to author, January 24, 2020. Used with permission.

25. Michael McDonnell, Email message to author, March 4, 2020. Used with permission.

26. Livanos-Propst, "Developing Weeding Protocols for Born Digital Collections."

27. Elizabeth Psyck, "To Weed or Not to Weed: Managing an Electronic Collection" (PowerPoint, DLC meeting and FDL Conference, 2016), https://www.fdlp.gov/file-repository/fdlp-academy-video-supplemental-materials/3609-to-weed-or-not-to-weed-managing-an-electronic-collection-captioning.

Annotated Bibliography

American Library Association. "Collection Maintenance and Weeding." American Library Association. Last updated January 2018, http://www.ala.org/tools/challengesupport/selectionpolicytoolkit/weeding

This section of the ALA's Selection & Reconsideration Policy Toolkit for Public, School, & Academic Libraries emphasizes criteria that should be used when making weeding or deselection decisions. It provides guidelines on creating policies to guide weeding and deselection decisions, focusing on both tangible and digital materials. Policy considerations are provided for Public, School, and Academic Libraries. Example policies for each are also provided.

Brown, Christopher C. "Knowing Where They Went: Six Years of Online Access Statistics via the Online Catalog for Federal Government Information." *College & Research Libraries* 72(*1*): 43-61, https://doi.org/10.5860/crl-68r1

The author provides general background on the GPO's transition to digital materials, with a focus on Colorado depository libraries and the author's own library at the University of Denver. The primary focus is an analysis and evaluation of the data collected from Denver's URL click-through tracking initiative that added a tracking prefix to every government document URL in their OPAC.

Dahlen, A. "FDLP PURL Usage Report—Analyzing Usage of Digital FDLP Resources." Video file, June 12, 2019, https://

www.fdlp.gov/fdlp-purl-usage-report-analyzing-usage-of-digital-fdlp-resources

This recording of a 2019 FDLP Academy webinar introduces the basics of PURLs and goes into detail on how to set up the PURL Usage Reporting Tool to track usage data. The webcast also discusses how to work with usage reports to understand what materials are being used and how users are locating those materials.

Dahlen, A., & Fairfield, M,. "Weeding Print & Digital Resources." Session, 2019 Federal Depository Library Conference, Arlington, VA, October 22, 2019. https://www.fdlp.gov/fall-2019-federal-depository-library-conference

This program from the Fall 2019 Federal Depository Library Conference, which includes the accompanying slide deck, focuses on the process of weeding print depository collections. Methods to weed digital collections are discussed including running reports on your ILS and adding MARC notes. Suggestions were also given regarding reviewing links and finding aids that direct to digital materials.

FDLP. "Weeding a Depository Collection." FDLP Requirements and Guidance. Last updated August 26, 2019, https://www.fdlp.gov/requirements-guidance/guidance/30-weeding-a-depository-collection

This GPO guidance explains the rules and regulations governing the weeding of tangible depository collections. The three ways that tangible depository materials qualify for weeding are discussed in detail.

Livanos-Propst, A. "Developing Weeding Protocols for Born Digital Collections." *Code4Lib Journal* 43 (2019), https://doaj.org/article/089c052c6fab48ae8b5003f45af4c1be

The author describes the process used to weed over 14,000 digital items in a free collection of classroom-ready educational resources. The weeding project is described in detail, providing assessment criteria, review of compiled assessment data, and policies implemented to maintain a vibrant, usable collection.

Psyck, E. "To Weed or Not to Weed: Managing an Electronic Collection." Session, DLC Meeting and FDL Conference, 2016, https://www.fdlp.gov/file-repository/outreach/events/depository-library-council-dlc-meetings/2016-meeting-proceedings/2016-dlc-meeting-and-fdl-conference/2778-to-weed-or-not-to-weed-managing-an-electronic-collection-slides

The author, a Government Documents Librarian, outlines some pros and cons of weeding electronic documents in this short PowerPoint presentation. She cites four steps in this process.

Sare, L. "Weeding in the Digital Age," Session, DLC Meeting and FDL Conference, 2015, https://www.fdlp.gov/file-repository/outreach/events/depository-library-council-dlc-meetings/2015-meeting-proceedings-1/2015-dlc-meeting-and-fdl-conference/2626-weeding-in-the-digital-age

This presentation focuses on how to identify print material that should be weeded and identifies criteria to consider when evaluating whether an item should be weeded.