



# The Preservation of Government Publications

October 2017

Roger C. Schonfeld | [roger.schonfeld@ithaka.org](mailto:roger.schonfeld@ithaka.org)

Chairman Harper, Ranking Member Brady, and Members of the Committee, thank you for inviting me to testify before the Committee on House Administration on “Transforming GPO for the 21st Century and Beyond.”

I am the director of the library and scholarly communication program for Ithaka S+R, a not for profit research and advisory service that focuses on higher education.<sup>1</sup> I am a librarian who has led a number of projects examining preservation issues broadly in the transition from print to digital content and collections. Nearly a decade ago, I led two major consulting projects focused on structural challenges facing the Federal Depository Library Program (FDLP) and how its vital work can be sustained.<sup>2</sup> Others have spoken with you about some of the benefits of the Program and the challenges it faces. I will focus my remarks on factors I hope you will consider related to the long-term preservation of government publications, in print and also electronic form.

## Print Preservation

The current enabling legislation does not assign preservation responsibilities to the FDLP. Rather, it assigns to regional depositories a responsibility for permanent retention of the publications distributed to them.<sup>3</sup> Preservation

---

<sup>1</sup> Ithaka S+R is a service of the not-for-profit organization, ITHAKA, which also operates the JSTOR, Artstor, and Portico digital library and preservation services.

<sup>2</sup> See Roger C Schonfeld and Ross Housewright, *Documents for a Digital Democracy: A Model for the Federal Depository Library Program in the 21st Century* (New York: Ithaka S+R, 2009), available at <https://doi.org/10.18665/sr.22358> and Ross Housewright and Roger C. Schonfeld, *Modeling a Sustainable Future for the United States Federal Depository Library Program’s Network of Libraries in the 21st Century: Final Report of Ithaka S+R to the Government Printing Office*, May 16, 2011, available at <http://www.uflib.ufl.edu/docs/ithaka-final-report-and-gpo-statement.pdf>.

<sup>3</sup> See 44 US Code 1911.

should nevertheless be understood as an imperative for ensuring permanent public access.

Understood comprehensively, preservation entails a variety of activities, processes, and conditions designed to extend the life of a publication and the information it contains. Preservation activities for an individual collection go far beyond simply providing the space necessary to store materials. It requires appropriate environmental conditions, which are not always present in legacy library facilities, to extend the life of paper-based collections. It involves physical security, facilities maintenance, and disaster planning. It involves processes to identify missing materials and fill gaps that arise. It involves item-level metadata to identify what is held and where as well as conservation. And of course it involves a commitment to retain materials, usually combined with provisions for access.

The FDLP provides for some but not all of the steps necessary to ensure the preservation of the print collection, mostly by having multiple copies of many items housed in 47 “Regional” depositories. About half of these are academic libraries.

## How Many Copies

Because of research that I commissioned from a University of California, Berkeley operations researcher, we have a framework for determining the number of print copies of a given item that are needed for preservation purposes.<sup>4</sup> Depending on a number of factors, including the condition of the collection and whether the materials are in circulation, we may need fewer than 10 copies to ensure their long-term preservation.<sup>5</sup>

Federal government publications with widely available trusted digital copies are well suited to this type of analysis, since digitization substantially reduces the use (and therefore risk of loss or damage) of physical volumes

---

<sup>4</sup> Candace Arai Yano, Zuo-Jun Max Shen, and Stephen Chan, “Optimising the number of copies and storage protocols for print preservation of research journals,” *International Journal of Production Research* 51:23-24 (2013), pages 7456-7469, available at <http://dx.doi.org/10.1080/00207543.2013.827810>.

<sup>5</sup> Among the other factors are the quality of the digitization, digital preservation, and terms of access, to ensure that access has migrated to the digital version; the degree of quality assurance provided for the print copies, and the degree of image intensiveness.

The Berkeley model imagines that there are at least some pristine copies of the items, which requires different kinds of preservation conditions than are currently in place through the FDLP.

The key takeaway is that there is a balance between the preservation approach and the number of copies that are needed in order to provide assurances that the collection will remain available.

## Trust Networks

The FDLP is under stress today for a number of reasons. It is my view that an essential reason is that it is organized on a state-based model, when academic libraries are increasingly striving to manage their print collections collaboratively in ways that do not align with state boundaries.

In the United States, academic libraries share much of the responsibility for preserving books, periodicals, and other print publications. But since the FDLP was created, how libraries structure their responsibilities has changed substantially. In a previous era, preservation was assured by many libraries collecting copies of the same work. Today, this is decreasingly the case.<sup>6</sup>

There are several changes that have faced academic libraries. Digital versions of many publications have become the main form of access for reading purposes, certainly for periodicals. And the logistics of managing and sharing print versions has improved substantially. As so much information of all types have been made freely available on the internet, libraries are pivoting to find a variety of ways to add value to their communities, and academic libraries are repurposing their central campus spaces away from collections storage. To allow them to do so while maintaining access to print publications, academic libraries are creating “trust networks” to take collective responsibility for storing and preserving print collections.

These networks can cover a variety of different content types, differ somewhat in their provisions around permanence, and involve a number of organizational

---

<sup>6</sup> The term for this strategy, “preservation through proliferation,” was defined in Stephen G. Nichols and Abby Smith, *The Evidence in Hand: Report of the Task Force on the Artifact in Library Collections* (Washington DC: Council on Library and Information Resources, 2001), available at <http://www.clir.org/pubs/reports/pub103/pub103.pdf>.

structures and governance models.<sup>7</sup> But they typically involve a more explicit commitment to retention and in some cases preservation than any single library was ever able to make for similar materials on its own.

The Association of Southeastern Research Libraries - whose 38 member libraries across 11 states form one of the strongest and most innovative trust networks we have - has established an innovative program for government documents.<sup>8</sup> Participating libraries create a center of excellence for a certain agency or set of agencies, committing to build, maintain, and catalog (and in some cases digitize for preservation and access) a collection more rigorously than the FDLP mandates. Today, there are 40 Centers of Excellence focusing their collection management work on agencies of different sizes and character -- from the US State Department to the historic House Un-American Activities Committee. This type of distributed but systematic approach is a fantastic model for what the future of the program could look like. The FDLP should evolve along with library preservation strategy to allow a group of libraries together to serve in the role of the Regional, combining the flexibility associated with such a collaboration with greater preservation assurances that this model provides.<sup>9</sup>

Of course academic libraries are not the only library members of the FDLP. But the program can and should evolve to accommodate changing strategies and organizational structures for preservation regardless of the type of library.

## Digital Preservation

As government publications are now typically issued in digital format, it is essential that they be preserved in digital form. Digital preservation has several discrete components to it. First, there are the technical issues around file preservation, format changes over time, preservation metadata, and similar

---

<sup>7</sup> A number of very interesting trust networks are profiled in an issue of *Against the Grain* that I edited. See volume 22, number 5 (November 2010), available at <https://doi.org/10.7771/2380-176X.5640>.

<sup>8</sup> Of ASERL's 38 members, 37 are Federal Depository Libraries.

<sup>9</sup> At the same time, I have expressed concerns about whether libraries participating in these trust networks have given enough attention to the long-term preservation questions. Roger C. Schonfeld, "Taking Stock: Sharing Responsibility for Print Preservation." Ithaka S+R. July 8, 2015, available at <https://doi.org/10.18665/sr.241080>.

issues. And second, there are governance issues, including how preservation is organized and funded. I want to focus on organizational issues.

Best practice is for preservation responsibilities to be transferred from the creator and publisher to one or more third parties representing the customers and users of publications.<sup>10</sup> For trade and scholarly publications, for example, individual libraries or groups of them acting collectively have typically served as the preservation agent.

Federal publications seem to represent a unique type of preservation problem. While the federal government publishes materials openly on the internet and they are freely available for all to read, there is a diffusion in responsibility for preservation that must be addressed.

GPO now operates FDSys and will operate GovInfo as a digital platform for free access and long-term stewardship of federal publications. Collecting all federal publications into a single platform is a worthy goal. However, many federal digital publications are not being gathered up digitally into this platform. It is therefore reasonable to worry if their preservation as a coherent collection is failing. GPO needs stronger focus or abilities to enable it to build this coherent collection of federal publications, or an alternative other than GPO must be found for doing so. In addition to solving this issue prospectively, it must also be addressed retrospectively, to ensure that gaps in holdings of digital and digitized publications are filled. But even if GPO can fully populate federal publications on this platform, we should question whether GPO should have sole responsibility for preservation.

It is my view that extending preservation best practices to federal publications requires a strong third party role for preservation in addition to the efforts of the government itself. This third party (or these third parties, since there can certainly be more than one) should follow a number of principles and practices:

- » It should have a formal agreement with the federal government outlining its responsibilities.
- » It should take custody of government publications when they are issued or on a timely basis soon thereafter.
- » It should maintain custody of these publications in a diversity of political jurisdictions, in this case to include at least one jurisdiction outside the United States.

---

<sup>10</sup> See for example the principles outlined in this statement issued by The Andrew W. Mellon Foundation based on a roundtable of library leaders. “Urgent Action Needed to Preserve Scholarly Electronic Journals,” available at <http://www.arl.org/publications-resources/1150>

- » At any point that publications become unavailable through FDsys, GovInfo, or their successors, the third-party would be obligated to provide permanent public access.

I want to emphasize that the current structure of the Federal Depository Library Program is not able to provide for this kind of third party. The FDLP could be restructured to allow for such a model, or GPO could contract with such a third party outside of the FDLP. Stable financial support must be provided.

In addition to this formal relationship, GPO should also allow complete and selected bulk downloads of all digital content and metadata so that any individual or organization can analyze, make accessible, and preserve federal publications.

## Conclusion

I thank this Committee for its interest in how the FDLP might be updated to ensure preservation and access to government publications. By modernizing the FDLP and aligning it with preservation best practices, you can help to ensure that these vital government publications will remain available for the American public for generations to come.