

Recommendations to the Library of Congress — Legislative Information Services and How They Could Be Improved

September 4, 2020

We are pleased to submit the following recommendations concerning the Library of Congress' legislative information services. As we discuss below, we recommend the Library of Congress increase its efforts to—

- Publish Information as Data
- Put the Legislative Process in Context
- Integrate Information from Multiple Sources
- Publish Archival Information
- Collaborate with the Public

Introduction

The Library of Congress recently announced it will hold a virtual forum on September 10, 2020, where it is expected to hear stakeholder feedback on the Library's legislative information services. The forum is being held at the direction of the House Committee on Appropriations pursuant to its report accompanying the FY 2020 Legislative Branch Appropriations Bill, which requested the convening of a wide range of stakeholders to explore how the Library's legislative information services could be improved.

We appreciate the Library hosting this conversation and are excited to provide these recommendations for its consideration. We thank the House Legislative Branch Appropriations Subcommittee for proposing the forum and recommend and request that there be many more like it.

The Library of Congress is unique because it is *Congress'* library. The broad original purpose of the Library of Congress was to provide a repository of knowledge from which legislators could learn so that the nation could benefit from a more informed national public policy. Today, it is clear that Congress' library must specifically help surface the information that is most important and most relevant to the nation's current challenges so that Congress, business, public sector groups, and the public itself can benefit from access to the raw sources of information and from the Library's cogent (nonpartisan), taxpayer-funded analyses.

Before we begin, we wish to recognize some recent activities of the Library of Congress.

First, the Library has made great strides in its efforts to modernize one of its digital information services, Congress.gov, and we thank that team for its concerted efforts to communicate with internal and public stakeholders. Congress.gov is a tremendous improvement over THOMAS, which we remember well, and the periodic blog posts that clearly and transparently explain improvements to Congress.gov are an appreciated effort to improve communication with the

public. The Library's collaboration with other Legislative branch agencies on the publication of bill summary and status information as structured data on GovInfo.gov is a particularly welcome and important change. So too are the continual efforts to enrich Congress.gov bill pages with additional information drawn from the House and Senate, such as documents submitted during the proceedings and (when available) video of those proceedings. In addition, we are pleased to see the publication of calendar information about House and Senate committee meetings, which was done pursuant to a directive from Appropriators, and the continual improvements the Library has made to that centralized committee calendar.

Second, we commend the Library for its efforts to make available online significant new resources on Congress' legislative activities. The Law Library of Congress' ongoing digitization efforts concerning its legal reports, the Statutes at Large, and the Serial Set are to be applauded and will provide new insights into legislative activities. (We also thank the GPO for collaborating with the Library on digitization efforts.)

Finally, we note the efforts by the Congressional Research Service to publish Congressional Research Service Reports online, pursuant to legislative directive, and the publication of the Constitution Annotated. Both sets of documents have long been sought by the public, and this public access is elevating policy discourse both outside of and within Congress.

1. Publish Information As Data

A number of the digitization efforts undertaken by Legislative branch support offices and agencies constitute transforming paper records into PDFs. There undoubtedly is value in making authoritative electronic copies of paper records available online. We believe, however, that even greater utility and value would arise from transforming paper documents into data, which, in turn, can be republished in a variety of formats, including PDFs and Excel/CSV for accompanying numeric and tabular data.

Take the effort to digitize the Statutes at Large, for example. Rekeying the statutes as data — whether in a structured, semi-structured, or unstructured format — would unlock a significant number of additional uses for the information. The data could be ingested into the Comparative Prints (Posey) Project of the U.S. House, which shows how a proposed bill would amend a law; parsed to infuse hyperlinks into other legislation; more easily found through search engines; used to analyze how legislation has changed over time; made more readily accessible to persons with disabilities; allow for permanent citations (see perma.cc); and more.

Given limited resources, when considering how to transform paper documents into digital form, the Library should evaluate what transformation provides the greatest utility to Congress, to the Library, to internal stakeholders, and to the public. In many respects, a paper record transformed from print into digital text (in a structured data format) is simply more versatile; additionally, while digital text can be easily turned into a PDF, the inverse is much more difficult to do.

The Library of Congress faces a number of challenges when contemplating digitization projects. We encourage the Library to strongly prioritize transforming paper records into digital text, especially in structured data formats (such as XML and JSON). This will not be true in all cases, such as when the vast majority of the utility comes from making the documents publicly available and the time involved in rekeying the data would be prohibitive. Similarly, in circumstances where the Library already has PDFs of the document but has yet to publish them online, it should not wait to publish that information. Nevertheless, we urge the Library to consider a digital data-first approach, and to create a priority list of paper records (or PDF files) that should be transformed into data.

When the Library already has records in a structured or semi-structured digital format, it should as a matter of course make that information available to the public in those formats.

Examples:

- A. **Publish Congressional Research Service reports as HTML.** CRS generates several thousand new or updated non-partisan reports each year; these reports are required by law to be publicly available. While the reports are published on CRS' internal website both as PDF and HTML files, they are only available as PDFs on the public-facing

website crsreports.congress.gov. This significantly impedes the utility of the public-facing documents and also decreases their visibility to search engines and visually impaired users. These reports should be published online both as HTML and as PDF.

- B. **Make an API for Congress.gov publicly available.** Congress.gov has an internal API that shares legislative information in a structured data format between internal stakeholders. It should make an API available to the public to facilitate access to legislative data.
- C. **Make the text of amendments promptly and readily available.** Floor amendments are often contained in the Congressional Record but their text is often very difficult to find via the congress.gov website. The Library should explore methods to make all amendments fully available and easily findable and readable on Congress.gov starting with amendments whose text appears in the Congressional Record.

2. Put the Legislative Process in Context

The Library has a role to play in supporting public and Congressional understanding of the legislative process. In an age of big data and fast-moving challenges, Congress looks to the Library of Congress to place information in context, and the Library must look to data sources across the legislative branch to make that happen.

In addition, the Library should be at the forefront of data collection and analysis that empowers the Legislative branch, and should lead the effort to connect datasets that provide important context to users. Furthermore, it should be cognizant of the changing responsibilities of information providers in a world in which misinformation and disinformation campaigns are on the rise.

Examples:

- A. **Congress.gov's excellent calendar should integrate upcoming floor proceedings** and, to the extent possible, proceedings before the House Rules Committee. This would supplement the information already available on the Committee Schedule. Congress.gov links to where some of this information can be found; it should go a step further and integrate the information onto the calendar page.
- B. **The Library should provide links to meeting transcripts** on Congress.gov as quickly as they become available. This would enhance hearing accessibility and help users identify proceedings on topics of interest. It should also consider publishing automatically generated unofficial transcripts (with proper notation that they are unofficial) to assist people with disabilities and others in accessing and using the information while the official transcripts are finalized, which can take months or years.
- C. **Information published on the Library of Congress website should be more timely.** We understand the Library of Congress may not be the original data source, but there can be a significant lag between when information becomes publicly available and when it becomes available on Congress.gov. There are even circumstances where legislation is passed by a chamber before the text is available on Congress.gov. A vast array of third-party paid services fill some of the gap between when information is released and when it becomes available on Congress.gov. Particularly for data originating with Congress, there should not be two-tiered access for those who can afford to pay for these services and everyone else. The Library should strive to reduce or eliminate this timeliness gap — perhaps by encouraging greater alacrity on behalf of its congressional partners or speeding up its own processes — so that public information is available to everyone as close as possible to when it is released. Congress is the first branch of government, and it cannot afford to be behind the Executive branch or the private sector when it comes to access to its own information.

- D. **Link directly to CRS reports in references to legislation.** CRS reports often refer to specific bills and provide useful context that enhances public understanding of those bills. Improving discoverability of CRS reports in the context of Congress.gov bill pages is important; thus, bill pages should link to CRS reports that, in the text of the reports, specifically identify and relate to that particular legislation.
- E. **Integrate historical information about members, committees, leadership, and offices into legislative data.** The National Archives conducts a “congressional web harvest” at the end of each Congress that archives all congressional websites. This archive goes back to the 109th Congress. These records contain a wealth of information that provide useful context to legislative activities. Congress.gov should consider how to integrate this historical information. For example, Congress.gov’s “committees profile” pages, which link to the current committee webpage, could also link to committee webpages for prior Congresses (with appropriate context to explain that it is historical material).
- F. **Expand implementation of human centered design principles.** The Library should consider expanding its work concerning human-centered design to further improve how it makes its most complex information products, such as the appropriations tables, available to the public. While great amounts of information are available on Congress.gov’s appropriations webpage, there is room for continued improvement to help users find the information they want and present it to them in an intuitive format.
- G. **Reimagine topical information products.** The Library produces a number of helpful reports and analyses, but perhaps more could be done to gather and synthesize information on topics of interest to Congress. There could be utility in a regularly-produced information product that draws upon the entire resources of the Library and presents that information to staff in a digestible format. For example, the topic of cybersecurity is both in the news and relevant to Congress. An information product — a newsletter, for example — that regularly provided updates to congressional staff and the general public on the latest topical developments is likely to be useful in many contexts. Such a product could, for example, include summaries or previews of hearings, highlighting the introduction or passage of legislation, providing new domestic law analyses from CRS or new foreign law compilations from the Law Library, indicating whether the Executive branch has recently taken or plans to undertake an action, identifying the holding in recent court opinions, and so on. The private sector has developed a number of such topical newsletters along these lines — for a fee — but many issues important to Congress remain unaddressed or underdeveloped.

3. Integrate Information from Multiple Sources

The Library of Congress, House Clerk, GPO, CBO, GAO, and other internal stakeholders steward a vast amount of information relevant to legislative activities. Congress.gov is the place to connect those disparate datasets, thereby enhancing user experience and helping them locate, interrogate, and retrieve a cross-section of valuable information.

Examples:

- A. **Ensure hearing videos are consistently available on Congress.gov.** As we discussed above, videos can be absent from hearing pages for a variety of reasons, some as prosaic as committee ID input and formatting errors. We understand that significant difficulties can arise in access to this information from the original information providers, and that this may not be fully within the control of the Library to address. Nevertheless, the Library should continue to work with House committees to ensure videos are identified correctly and explore alternate means to categorize the videos should the primary sources be unable to provide sufficient metadata. Similarly, Senate committees should be encouraged to publish their videos in formats that the Library can archive and integrate into meeting pages.
- B. **Integrate all federal laws into Congress.gov.** The Statutes at Large comprise all public laws enacted by Congress from the start of our country. They remain relevant to Congressional activities today. The Library already republishes all Statutes at Large on Congress.gov, but they should be able to be accessed from Congress.gov's main search engine by citation and via link from the text of legislation that cite them. In addition, as the Statutes at Large are not digitized as structured data, the Library should consider alternate means of obtaining those laws as structured data, such as requesting a gift or purchasing those laws from a third party.
- C. **Link to all previous iterations of a bill.** Frequently a bill is introduced during more than one Congress with identical or slightly modified text. The Library should identify iterations of a bill in a prior Congress and link to those past iterations on Congress.gov legislation pages similar to how it links to identical and other related bills within the same Congress.
- D. **Identify and incorporate datasets generated and held outside the Legislative branch.** The National Archives holds vast amounts of legislative information and is the final repository of enacted laws. The White House often issues Statements of Administration Policy or Presidential Signing Statements that relate to legislation; signing statements are in turn collected in the Compilation of Presidential Documents published by the Office of the Federal Register in the National Archives and Records Administration. Courts often provide interpretations of statutes. Regulations can further define the meaning of a statute. The Library of Congress should consider which of these datasets might be worth integrating into Congress.gov and how it might do so. This

would help provide more contextual information about legislation and how it is implemented.

- E. **Build a dashboard indicating what information is not yet available on Congress.gov**, as there is often a window of time between when a bill has been introduced and when that bill text is available on Congress.gov. In addition, collaboratively create a roadmap for information the Library wishes to transform into data.

- F. **Allow Congress.gov users to search by topic** across multiple Library of Congress silos, as the average person looking for information on Congress.gov is likely to think in terms of subject matter, not source. For example, a user may be interested in “encryption”; when they search for that term it would be useful if Congress.gov returned results from a variety of sources including CRS reports, Law Library reports, relevant hearings, relevant bills, and more. We understand that bringing together information from across the various Library of Congress silos would be difficult, but it makes sense to build towards this goal.

4. Publish Archival Information

The Library is the access point to great amounts of legislative information that is either generated in the Library or re-published from other entities. Oftentimes, datasets published by the Library are incomplete for various reasons. The Library should strongly consider publishing the rest of the dataset, or at least expanding what is available to a reasonable point based on current or anticipated resources, with the option of going further as more resources become available.

Examples:

- A. **The Library should review and publish CRS reports in the CRSX archive.** The Library already publishes recent CRS reports online, but the many already-digitized historic CRS reports in the CRSX archive are unavailable from CRS. (They are often available from third-party sources for a fee). The Library should evaluate and publish those non-confidential CRSX reports on its public-facing CRS reports website. Additionally, the Library should consider publishing non-digitized CRS reports going back to the modern founding of the agency in 1970. While those reports are older, they are often still relevant to political debates today.
- B. **The Library should make video archives of Congressional committee proceedings available.** Fortunately, the Library already has world-class capability in video curation. The Packard Campus of the National Audio-Visual Conservation Center maintains a high-resolution video archive of Congressional hearings. The Library should consider making these valuable historical materials much more widely available. In addition, we note that the Library of Congress is publishing links to some House committee videos. As noted above, the Library should work with its partners to make sure it is linking to all House videos; it should also endeavor to collaborate with the relevant Senate offices to obtain and publish videos of Senate committee proceedings. It should also work to preserve recordings of House and Senate committee proceedings. Video helps bring proceedings alive and provides valuable contextual information.
- C. **The Library should make historic bill text available online.** Currently, bills on Congress.gov only go back to approximately 1995, which is when THOMAS was created. Legislation considered prior to that date is of practical and historical interest and should be available as well. The Library should evaluate expanding the archive of bills online prior to 1995.
- D. **The Library should work with GPO to collect, publish, and preserve Congressionally mandated Executive branch agency reports** published in House Document at the beginning of each Congressional session. There is legislation wending its way through Congress to require this to happen, but the Library should proactively move to do this.

5. Collaborate with the Public

For almost a decade, the Legislative branch as a whole has engaged in an ongoing effort to collaborate with its internal and external stakeholders through quarterly meetings of the Bulk Data Task Force and the Annual Legislative Data & Transparency Conference. These thoughtful and informative meetings have allowed us to regularly engage with many of those doing the hard work of modernizing Congressional technology and making information about legislative activities available to the public. The Library should apply this approach to its own operations and across its various silos.

In addition, we have heard multiple disquieting reports that some staff in the Library have, at times, discouraged Library employee attendance and participation at these proceedings. The Library should make clear through official and unofficial channels that Library staff are encouraged to attend, participate, and collaborate. We have seen participation and collaboration by the heads of other Legislative branch offices and agencies at these events, and there would be great value in the Librarian of Congress providing her own stamp of approval.

We believe that the forthcoming forum focused on the Library's legislative information services presents an opportunity to further showcase the Library and more deeply involve public stakeholders. It also makes a space available for greater collaboration between the Library and the public. This is a tremendous opportunity that the Library should seize upon.

Examples:

- A. **The Library should formally embrace and announce a presumption of public access to legislative information with express limited exceptions for legitimately confidential material.** The Library of Congress should adopt an institute-wide posture that it will make legislative information available to the public — to “default to open” — except in those circumstances that warrant it be withheld, and it will proactively make requests to the original owners of information to publish that information instead of waiting for a directive to do so. The Library should prompt conversations concerning the release of information and development of new services — especially when that information is partially available or available in another format — and once given permission to publish information in one format, the Library should be able to undertake efforts to publish it in other formats.
- B. **The Library should meet quarterly with interested parties** to discuss its digitization goals, efforts, and identify areas to collaborate. It further should publicly announce points of contact — with relevant contact information — for its different divisions, so that people who wish to collaborate with the Library or have questions will have a person who is responsible for responding to them and building a positive relationship with the Library community. While there are some efforts to do this in some divisions of the Library, it should be expanded to all divisions (regardless of whether they would say they have a

public mission).

- C. **The Library should deepen and expand its relationship with the legislative data community** of political scientists, information publishers, application developers within Congress and congressional vendors, and enterprising journalists who share a similar mission as the Library does to create public access to legislative information and a better informed policymaking process. We are all on the same team. We encourage the Library's participation in conversations with the community on shared challenges and goals and participation in efforts like the United States Project (<https://github.com/unitedstates>), a collaborative GitHub repository focused largely on legislative information. The Library should allow the public to collaborate with the Library, as appropriate, on the development of its public legislative information tools. The LOC Github repositories (<https://github.com/LibraryOfCongress>) and LOC Labs (<https://labs.loc.gov/>) are excellent initiatives, worth building on. The Github repositories, in particular, could benefit from organization into a profile which guides users in how to contribute.
- D. **The Library should expand its crowdsourcing efforts** to include the transcription of records relating to legislative activities.
- E. **The Library should explain how it drafts bill summaries**, which are of tremendous importance to how most people understand what legislation is intended to accomplish.
- F. **The Library should include a summary of public input** and an explanation on how it impacted operations in its annual report to appropriators.
- G. **The Library should strengthen its collaboration with FDLP Libraries** and others in digitizing and drawing connections among legislative information records. With more than 1,100 FDLP libraries alone, the Library of Congress could expand its reach by strengthening its partnerships and collaborating on digitizing information. This also provides an opportunity to make Library information available in local contexts.
- H. **The Library should create a dashboard for data concerning Library of Congress web properties**. How many users are using Library of Congress properties? Which items are "hot?" How often are web properties "down?" How long does it take to process information for the website once it is received by the Library? All this, and more, would be interesting to know, and be useful in deciding how to allocate resources.
- I. **The Library should consider having a Wikipedian in Residence** and adopt other mechanisms to collaborate with non-profits and trade associations to make information held by the Library more widely available and provide additional context to information held by the Library.

Conclusion

Thank you for the opportunity to provide these comments in advance of the public forum on the Library's legislative information services. On behalf of the Congressional Data Coalition and many others, we are looking forward to the proceedings and would welcome the opportunity to discuss our recommendations further. Please contact Daniel Schuman, policy director, Demand Progress, at daniel@demandprogress.org.

Appendix

The following is the relevant section of the committee report accompanying the Legislative Branch Appropriations Bill, FY 2020, that directed this forum be held. (H. Rept. 116-64, p. 28).

Public Access to Legislative Information: The Library of Congress has become one of the pre-eminent sources of public information about federal legislation and congressional processes, particularly through its role in hosting the "Congress.gov" website but also through other digital activities. The Committee commends the Library for its achievements in this area.

The Committee also understands that users of the Library's legislative information services have suggestions regarding ways in which those services could be improved or expanded, and the Committee would like to encourage creation of a more structured process for receiving and considering such suggestions. Such a process should also recognize some of the limits on the Library's ability to implement improvements and expansions of these services, however. For example, much of the information presented on the Congress.gov website is not generated or controlled by the Library but rather comes from other sources--particularly various offices of the House of Representatives and the Senate and the Government Publishing Office--and expanding or changing the content of Congress.gov will often require those sources to do something new or different. Further, some improvements and expansions would require commitment of substantial time and resources by offices both within and outside the Library that already have a heavy IT workload.

To facilitate public input into the Library's legislative information services and how they could be improved, the Committee urges the Library to convene a public meeting at least once a year during the next two years to discuss these issues. The Library should invite a broad range of potentially interested parties to participate, such as journalists, academics, advocacy and public interest organizations, research institutions, libraries, and publishers of legislative information. The Library should also invite and encourage participation from representatives of Legislative branch stakeholders such as the Clerk of the House, the Secretary of the Senate, and the Government Publishing Office.

Following the meeting, the Library should prepare a report, to be submitted to the House and Senate Appropriations Committees and also to relevant authorizing committees and officers of the House and Senate. **The report should summarize the comments and suggestions made, indicate which could be implemented by the Library itself and which would require action by other entities, roughly estimate the cost and effort needed (where possible), and provide the Library's evaluation and recommendations for further action.**

(emphasis added)