

**Comments Submitted for the Record**

*Submitted by:*

**Joshua Tauberer, Ph.D.**  
Co-founder and Chief Technology Officer  
**POPVOX**  
1425 K Street NW, Suite 350  
Washington, DC 20005

*For:*

House Committee on Appropriations  
Subcommittee on the Legislative Branch

*Regarding:*

Budget Hearing - Library of Congress / U.S. Government Accountability Office / U.S. Government  
Printing Office / Congressional Budget Office (Feb. 7, 2012)

In particular, bulk data for legislative information.

*Submitted on:*

February 6, 2012

## Comments Submitted by Joshua Tauberer / POPVOX

February 6, 2012

Chairman Crenshaw, Ranking Member Honda, and members of the Committee, thank you for the opportunity to submit comments on the budget for the Library of Congress and the Government Printing Office.

I am the co-founder and chief technology officer of POPVOX, an Internet startup that helped constituents write 380,000 letters to Congress in 2011. I am also the founder of Civic Impulse, LLC, which operates the website GovTrack.us, a free and open source legislative tracking tool.

These remarks address the inability of the Library of Congress to meet the practical needs of the public it serves, the information demands of Members of Congress, as well as statutory requirements regarding the electronic publication of legislative records. In particular, the Library of Congress should be funded sufficiently and specifically to publish a bulk database of bill status.

**“Bulk data”** is today considered a core component of any government information dissemination program. In 2009, the Government Printing Office began offering bulk data for bill text, the Federal Register, and other publications. Its FDSys platform, crucial infrastructure for the dissemination of this data, has been regarded by technologists as a commendable success. The Office of Law Revision Counsel has long offered the United States Code in multiple bulk data formats, including XHTML. Earlier this year, the Clerk’s office, under the direction of the Majority Leader, began publishing bulk data for bills to be considered in the week ahead. Bulk data can be produced at a fraction of the cost of other information dissemination methods, such as websites.

Legislative data has a wide impact. Since 2004, GovTrack.us has been one of the most popular websites among the public to research and track the legislation that Congress is considering. Over the last six months, GovTrack and its data partners have been used by **5–10 million individuals**. These individuals often cannot understand the terminology or advanced capabilities of THOMAS and are frustrated by THOMAS’s limited mandate to publish only a small fraction of legislative records. GovTrack is a more accessible and more comprehensive reference for legislative status, Members of Congress, congressional districts, and bill text. The private sector, among it GovTrack and POPVOX, has stepped in to provide services the Library of Congress cannot or should not provide on its own. However, these private sector services rely on the ability to collect legislative information, and to do so affordably. The Committee should be proud of how GPO has supported these services to date and should expand on this priority — especially at the Library of Congress — going forward.

Legislative data can also make the **House’s internal work-flows more efficient** and therefore cost effective. GovTrack’s free data services have been used by dozens of House offices to help Members locate their own voting records, to add district maps to Member websites, and to support research tools used by staff. POPVOX helped constituents write 380,000 letters to Congress in 2011. Despite the overwhelming mail volume facing House offices, legislative correspondents consistently tell us they wished more constituents used POPVOX to write letters to their office. That is because we use data to ensure that letters sent through POPVOX are easier for staff to process when they are received. POPVOX also recently announced an iPad app for legislative professionals, including House staff, which can save offices time and money. The app, as with the other applications mentioned here, is based on legislative data. The private sector can use bulk legislative data to help Congress save money

and work better.

Because Congress does not publish bulk legislative data, **GovTrack is seen as the authoritative source for legislative data.** GovTrack serves as a free data provider to dozens of other websites which build tools to help the public stay engaged with government. While GovTrack has been honored to provide this service for the past seven years, it is unfortunate that Congress is not seen as the authoritative source for its own information. (I should note that GovTrack, POPVOX, and other private sector websites get much of their information from THOMAS, but creating a comprehensive database of the information can only be done at a high cost and with low accuracy.)

The Library of Congress's current priority, with regard to legislative information, is to upgrade the THOMAS website. This upgrade is long overdue: THOMAS was built in 1994. However, creating **bulk data is both far cheaper and will have a much wider impact** than a website upgrade. Far more people get their legislative information from free websites in the private sector than from THOMAS.<sup>1</sup> Bulk data from the Library of Congress will improve the services that are being offered in the private sector and will reach a wider audience than updates to THOMAS. In 2009, Congress adopted an explanatory statement accompanying the Omnibus Appropriations Act of 2009 in which it directed the Library of Congress to consider bulk data downloads. Hindsight today tells us the language in 2009 was both forward thinking and on the right track. Bulk data has proven effective at other government agencies, as I noted earlier. At a meeting on February 2, 2012 hosted by the Committee on House Administration, Robert Gee, Chief of Law Library Public Services at the Library of Congress, indicated that the statutory requirement had not been fulfilled. It is time to make bulk data a priority.

I urge the Committee to take whatever steps necessary to enable the Library of Congress to provide bulk access to legislative documents, bill status, summary information, and other legislative data. An advisory committee composed of members of the Library, GPO, and other legislative data agencies and members of the public that regularly meets to address the need for public access to this information and the means by which it is provided will be necessary to guide the technology choices that lay ahead. I also urge the committee to maintain funding for GPO's FDSys program, which is the infrastructure needed to ensure the private sector can obtain legislative records published by GPO.

I appreciate your attention to this matter.

---

<sup>1</sup> According to compete.com, comparing thomas.loc.gov and thomas.gov to the sum of govtrack.us and opencongress.org.