Joshua Tauberer, Ph.D., President, GovTrack.us
Testimony before the
House Select Committee on the Modernization of Congress
on the topic of
Opening up the Process: Recommendations for Making Legislative Information More Transparent
on
May 10, 2019

Chairman Kilmer, Vice Chairman Graves, and members of the committee, thank you for the opportunity to speak before you today and for your focus on legislative transparency. My name is Joshua Tauberer and I am the founder and president of GovTrack.us and a collaborator with my co-panelist Daniel Schuman at Demand Progress on EveryCRSReport.com.

Today, my organization helps almost 10 million individuals each year learn about the daily activities of the United States Congress. Our website is a research and legislative tracking tool used by journalists, lobbyists, legislative staff here on the Hill, advocates, students, educators, and members of the general public. We explain legislation, track provisions of bills when they are incorporated into omnibus bills, predict the likelihood that bills will be enacted, we tweet legislative developments and send emails to users looking to track developments about particular bills, subject areas, or Members of Congress. And we do this to make information about the United States Congress accessible, understandable, and actionable so that our users are effective advocates for their issues. We are nonpartisan and our service is free.

Seventeen years ago, when I got started in this field, transparency in Congress looked a lot different than it does today. Then, like now, Congress published a vast amount of information about bills and votes online on THOMAS.gov (now Congress.gov) and the House and Senate websites. As a student researching intellectual property legislation for a college class back in 2001, I was amazed, actually, by how much Congress put online. But it wasn't understandable. Available doesn't necessarily mean accessible. I wanted to know, what was *in all* of the legislation in Congress at the time? Who were the sponsors introducing the most intellectual property legislation? Could I get an alert in my inbox whenever a new IP bill was introduced?

There are ~10,000 bills introduced in each Congress. If you want to know what's in them – you ask your staff, they use a computer. It's too hard to answer basic questions like this about legislation before Congress without technology.

So I emailed the Library of Congress, the steward — then and now — of Congress's legislative data, and asked them if they could share with me their database of bills in Congress. A list of bills. A spreadsheet of sponsors. Something in a format that I could feed into a computer to have it search it, sort it, and transform it for me. They declined. Frustrated but ultimately inspired by this experience, I spent the next 15 years building the most detailed database of bills, roll call votes, and other congressional information and made it freely available for anyone to reuse. To do that, I wrote software that gathered legislative information automatically from THOMAS.gov and other congressional websites, using a method that technologists call "screen scraping." While I finished my graduate degree, my software was adding more and more information to the database.[1]

---

1   More background at https://medium.com/civic-tech-thoughts-from-joshdata/govtrack-now-actually-uses-open-government-data-5fc16f377e86.

I used that database to build my organization's website GovTrack.us in 2004. And because even you, as members of the House, would have gotten the same answer from the Library if you had asked back then, congressional staff used my database, as well, to keep their bosses informed about new legislation. Almost 100 Member websites have used our resources to share legislative information with constituents. We've broken national news with our data.[2] And we've helped your constituents stay informed and engaged on legislative developments that matter to them.

Data and technology is how Congress is accessible – to everyone. There is no way to deal with the 200,000 pages of legislation introduced in each Congress without technology. There is no transparency in Congress without technology.

Times have changed, and Congress has, in many ways, harnessed the power of transparency through technology, following the examples set by projects like GovTrack. In 2016, a collaboration led by my co-panelist Mr. Bob Reeves between the Clerk of the House, the House Administration Committee, the Government Publishing Office, the Library of Congress, and the Senate Sergeant at Arms created an *official* comprehensive legislative database, in the form of bulk and structured data, like the one I had created. So I deleted my database, and GovTrack now uses Congress's official legislative data instead. Since my goal was to ensure that this data was available to the public by *someone,* and because Congress met that public need for information, I was able to refocus on new transparency projects.

Later that year, I worked with Demand Progress to repeat a similar strategy for bringing transparency to Congressional Research Service research reports. Although CRS's reports were widely available to those who could pay for them through private services, these taxpayer-funded reports, including almost 400 reports that explain different parts of how a bill becomes a law and other aspects of Congressional procedure[3], were not generally available to the public. So we built the most comprehensive database of CRS reports and made them available for anyone to read for free — with redactions of author names to respect the privacy of CRS staff. I used the same "set it and forget it" strategy as I did for GovTrack. While I'm talking to you here, EveryCRSReport.com is adding more reports to the website.

Last year, House appropriators encouraged CRS to publish its reports, a requirement ultimately enacted into law.

In 2017, using the new official data published by Congress, I created an analysis of the text of legislation that revealed that twice as many bills had been enacted into law than anyone was really aware of. These bills were enacted by incorporation into other bills, often large omnibus bills. When you take this into account, the popular idea of a gridlocked Congress looks a lot different. And, let me tell you, the sponsors of these bills love that they're now getting credit on GovTrack for having sponsored more enacted legislation. We're showing a more complete picture of what's *really* been enacted than Congress's own website Congress.gov.[4]

When I build technology for government transparency, my strategy is always the same: Don't do too much, and make it last forever — or at least until Congress does it better. That's how I built GovTrack without any outside funding and how Demand Progress and I built EveryCRSReport on a meager budget of a few tens of thousands of dollars. Good technology doesn't go out of date as quickly as you might think.

---

2   e.g. https://www.washingtonpost.com/politics/members-of-congress-trade-in-companies-while-making-laws-that-affect-those-same-firms/2012/06/23/gJQAlXwVyV_story.html
3   https://www.everycrsreport.com/topics/legislative-process.html
4   https://www.govtrack.us/about/analysis%23ideology#textincorporation

The question for the Modernization Committee is whether it will make the progress of the last several years by the House Clerk's office and the legislative branch agencies last forever too.

I've talked about how technology is a crucial aspect of legislative branch transparency, but I will now be more precise about what the Modernization Committee could be looking at.

The collaboration between the House Clerk, GPO, the Library, and the Senate on creating public access to bulk legislative data, known as the House Bulk Data Task Force, has proven successful. This data is flowing out to the American public and is creating more equitable access to the legislative process — only Beltway insiders and the most resourced organizations had access to this information decades ago; today, everyone does. So I thank Mr. Reeves and the rest of the taskforce for their amazing work. But there is more work the taskforce needs to do, especially creating public access to the text of amendments to bills. As you know, on major legislation, amendments are where all the action is. The taskforce's ongoing work to convert legislative data files into a consistent format known as USLM is also important. Until that collaboration is institutionalized and funded, this collaboration is at risk of ending.

But there is a bigger picture here than just creating more data. We need to be making the data that Congress publishes more understandable. Let me give you an example. I have with me the official data file for House vote on passage of H.R. 2146 in the 114th Congress, the Defending Public Safety Employees' Retirement Act.[5] It says the vote is on a bill "To amend the Internal Revenue Code of 1986 to allow Federal law enforcement officers, firefighters, and air traffic controllers to make penalty-free withdrawals from governmental plans after age 50, and for other purposes." But it wasn't, at least not exactly. It was actually one of the very controversial bills enacting the Trans-Pacific Partnership, but you wouldn't be able to tell from the data. Similarly, last year's FIRST Step Act, the bipartisan criminal justice reform bill, was enacted using a bill to reduce marine debris as its vehicle for passage.[6] In both cases, the data *isn't wrong*. The TPP provisions were added onto the *end* of the retirement bill with an amendment. The FIRST Step Act was a substitute amendment. But these small legislative maneuvers make it *exasperating* for members of the public to follow the progress of Congress's most important legislation using services built on top of Congress's official data.

There is no quick fix to this problem. It's going to require big thinking to make the House's legislative data more understandable. And developing technology to support transparency in the House for decades to come isn't easy either.

That's why I hope the committee will consider ways in which transparency and data can be *elevated* as a part of the institution of the House of Representatives. The creation of a Chief Data Officer or Chief Transparency Officer for the House of Representatives would create a space for this big picture thinking, and a person in this role would be in a position to coordinate the ongoing efforts in the House to ensure today's successes continue.

I also ask the committee to make the House Bulk Data Task Force permanent as the Congressional Data Task Force. On its original mission, the taskforce has focused on just one way congressional data is published. An expanded mission for the taskforce or for a Chief Data Officer or Chief Transparency Officer would look at how data is handled throughout the legislative branch, including information

5   http://clerk.house.gov/evs/2015/roll374.xml (use View Source in your browser to see the data file)
6   https://www.govtrack.us/congress/bills/115/s756/text/is

used for oversight, and for constituent services, and it would be able to provide key insights about other aspects of the operations of Congress.

Lastly, I mentioned my strategy of creating simple technology that lasts forever — or at least until Congress does it better than I did. Technology strategy is an important part of any large organization, but the House's technology is spread across many organizations. I worry that the amazing work already happening is not being used to educate technologists throughout the branch. In addition to the work of the Bulk Data Task Force that I already mentioned, the Government Publishing Office has done stellar work creating the new GovInfo.gov website; the Office of the House Clerk is working on a wonderful new site for the public to watch live House proceedings; the Library of Congress's use of "agile methodology" to build Congress.gov is an example that the whole legislative branch should look to. So I also ask the committee to consider ways to fund and nurture the technology experience and talent that exists within the legislative branch, and to create ways for digital efforts throughout the legislative branch to draw on that experience and talent. A Congressional Digital Service, modeled off of the executive branch's Technology Transformation Service team at the General Services Administration, is something to consider.

This committee has the unique opportunity to build on the last decade's worth of advances and to put in place mechanisms to ensure that we continue making progress. I am excited for the opportunity and look forward to discussing it with you.