

**The Washington Post**

# **Congressional data may soon be easier to use online**

## **Congressional data may soon be easier to use online**

By [David A. Fahrenthold](#), Published: June 8

Online, searching for a bill in Congress feels a little like time travel: Go looking for legislation, and you wind up in the Internet of 1995.

At Congress's ['90s-vintage archive site](#), there's no way to compare bills side by side. No tool to measure the success rate of a bill's sponsor. And there's certainly no way to leave a comment. Congress makes it hard for outside sites to do any of this, either, by refusing to give out bulk data on its bills in a user-friendly form.

On Friday, there was a signal that might change, as the GOP-led House moved toward releasing an unprecedented trove of data on its doings.

On a 307 to 102 vote, the House created a task force to study speeding up the release of data in the user-friendly XML format. That release is still a good distance away: The task force must answer members' questions about whether releasing congressional data could lead to fake bills or fake speeches, doctored up online.

But for a legislature that has closely guarded its data, this was a big step. In a statement, House Speaker John A. Boehner (R-Ohio) and other Republicans called it “the moment lawmakers agree to free legislative information from the technical limits of years past.”

For the unique species of geek who wants to slice up Congress’s data, this is the sandbox that has long been awaited. What if somebody built an app to rate congressmen like restaurants? Others might follow the Angie’s List model and treat them like contractors: Which representative gets the most bills passed? Who talks a lot but gets the least done?

Or an app might treat lawmakers like potential blind dates, creating an eHarmony equivalent for people looking for just the right legislator to send a donation to. Just type in how you would have voted on a set of bills, and it could search the real votes to find your best match in Congress.

The libertarian Cato Institute is already working to digitally “tag” spending amounts in legislation so that a user can search for every bill that would spend more than \$10 million.

The data — which are currently stored on a clumsy, impenetrable Library of Congress site known as Thomas — could in the new format be the raw material for a Yelp for Congress, a way for modern users to evaluate lawmakers with the same kind of crowdsourced help that they use to evaluate lunch.

Which means it could be something that Congress eventually will regret.

“What could you do if you had an overview of every committee appearance by a particular witness over the last five years?”

said Thomas Bruce, a professor at Cornell Law School. He thought more: What if witnesses could be checked for their political leanings or their campaign contributions? What would that say about the reliability of the advice Congress is getting?

Bruce said the House was smart to realize that it couldn't hold on to this information forever, since others are already inventing ways to compile it themselves. "This really is a teen-sex problem," he said. "Your kid is either going to find out from you or from the other kids."

In fact, one person has been working for more than a decade to try to build a better Congress on the Internet: a 30-year-old in Columbia Heights named Josh Tauberer who has jury-rigged his own database of congressional stats.

"The world [of Congress] is more futile than you thought. It's actually sadder than you thought," Tauberer said.

The new data will make his job easier, but the information could also make him less relevant, since some users will be able to go directly to Congress to find the information he provides on his site, govtrack.us.

Currently, the Capitol's most noticeable online fossil is the Library of Congress site named for Thomas Jefferson. On Thomas, users can search for bills by name, sponsor and subject. But they can look at only one bill at a time — divorced from the patterns, history and context that make all the difference on Capitol Hill.

"If you're outside the Beltway, you can't understand Thomas," Tauberer said.

He began working to organize congressional data in 2001, while a freshman at Princeton University. “What I saw was an opportunity for something much better.”

The job wasn’t easy. Over 11 years, Tauberer has built a computer program that “scrapes” the balky Thomas site, searching twice a day for new bills, new votes and new co-sponsors. It takes that information and packages it in a more Web-friendly form on his site.

On Thomas, a no-chance measure like, say, H.R. 40 — which would establish a commission to study paying reparations for slavery — might look like any other bill. It has a veteran sponsor, Rep. John Conyers Jr. (D-Mich.), and it has been assigned to a House subcommittee.

On govtrack.us, it’s possible for an untrained user to find some of the context that a Hill staffer carries in her head. If you look up H.R. 40, Tauberer’s site notes that it’s a repeat bill that has failed again and again in Congress. Conyers’s party is out of power in the House. And, to begin with, only about 4 percent of House bills ever pass.

“This bill has a 0% chance of being enacted,” the site says.

A Conyers spokesman was only slightly more sanguine, saying: “Rep. Conyers hopes a greater understanding of the substance of H.R. 40 will increase the likelihood of its passage, but he understands it takes time to educate members and the public.”

In all, Tauberer has spent nearly a dozen years tracking the vast amounts of wasted time and dead-end bills that the new data should reveal to the world. “I don’t fault them for trying

anymore,” he said of America’s legislators. “I actually think it’s kind of sweet.”

Tauberer isn’t the only one doing this: CQ Roll Call, for instance, has its own way of assembling the same data, which it provides to paying clients. But Tauberer is one of the few people who offers the data for free. That has become fodder for sites like Maplight.org, which combines bill data with fundraising information to show how much has been donated by a bill’s supporters and opponents. At OpenCongress, users can find news reports about a bill and offer their comments on the text, line by line.

Tauberer says he wants Congress to release the data itself, because he worries about flaws in his homemade system. It lags behind the latest news, for one thing, and it can produce errors when the computer misreads something.

The nonprofit groups that depend on Tauberer’s data have a more morbid worry.

“What happens if he walks in front of a bus?” said Daniel Schuman of the nonprofit Sunlight Foundation. Where would the data come from then? “This is a question about basic information that the federal government should be giving to the American people.”

This is not the first time that Congress — led by middle-age legislators whose best skills are verbal — has come late to a shift in electronic communications. The Capitol had a telegraph office, for instance, until 2007. The Senate still sends its messages by teenager, with its 183-year-old page program.

The House bill is intended as a step in the right direction. But the issue will be handed over to an institution far slower and older than the Internet: a congressional task force. It will decide how the bulk data will be released and how to protect information from being falsified or corrupted.

“I think everybody in Congress is happy to say, you know, ‘Here’s what we’re doing,’ ” said Rep. Ander Crenshaw (R-Fla.), who helped craft the bill that was voted on Friday. He said he recognized that the current system wasn’t enough: “I’m not a tech guy, but there’s a better way to do that.”