

Joshua Tauberer
Civic Impulse, LLC
tauberer@govtrack.us
202-558-7227

August 14, 2013

Speaker John Boehner
H-232 The Capitol

Majority Leader Eric Cantor
H-329 The Capitol

Re: Bulk access to legislative data

Dear Speaker Boehner and Majority Leader Cantor,

I am the president of Civic Impulse, LLC, which runs www.GovTrack.us, a website that helps ordinary citizens find and track bills in Congress and one of the world's most visited government transparency websites. In 2012 we provided information on legislation before Congress to 5 million individuals, our free widgets were used on 70 official Member websites, and the House Democratic Caucus turned to us for legislative data because they could not get the information they needed from any legislative branch agency.

On April 29, 2011, you called for House stakeholders to work on publicly releasing the House's legislative data in machine-readable formats.¹ Last year, after House appropriators tried to stall legislative transparency, our users wrote 1,400 letters to Congress about getting better access to legislative information², and on June 6, 2012 you directed a Bulk Data Task Force to begin implementation on bulk access to legislative data.³ The Washington Post ran a front-page story on the issue several days later.⁴

The task force sent its recommendations to House Appropriations on December 31, 2012 supporting broader implementation of bulk and structured legislative data.⁵ While the task force began work on several projects — all of which I commend, especially docs.house.gov, the projects in the office of the House Clerk, and the LRC modernization project — core legislative data has still not been made available:

No legislative branch agency makes available a spreadsheet that lists every bill introduced in the 113th Congress. How can there be meaningful transparency if the public cannot get a simple list of bills?

As you know, the Library of Congress is currently responsible for public access to bill status information. When their website THOMAS.gov is retired in 2014, GovTrack and all of

¹ <http://www.speaker.gov/press-release/speaker-boehner-majority-leader-cantor-call-new-data-standards-make-congress-more-open>

² <https://www.popvox.com/bills/us/112/hr5882/report>

³ <http://www.speaker.gov/press-release/house-leaders-back-bulk-access-legislative-information>

⁴ http://www.washingtonpost.com/politics/congressional-data-may-soon-be-easier-to-use-online/2012/06/08/gJQAdikBNV_story.html

⁵ <http://www.gpo.gov/fdsys/pkg/CHRG-113hrg81322/pdf/CHRG-113hrg81322.pdf>, page 679. The task force's recommendations were only made available to the public last month, six months after the recommendations were submitted to House Appropriations.

our users and data partners will lose access to the rudimentary source of legislative information that we use now.⁶

I urge you to direct the Library of Congress to publish structured data for the House's legislative data as soon as possible. Specifically, this means publishing the House half of the Library's "bill summary and status" database as structured data, such as in XML.⁷ This database lists bills introduced in Congress, titles, sponsors, actions, and similar information.

Structured data is the fiscally responsible way to achieve legislative transparency. It is cheaper than website development (like Congress.gov) and reaches a wider audience by leveraging the private sector — including journalists and businesses — to keep the public informed.

Small steps with incremental progress would be more useful than waiting another year for a comprehensive system. An appropriate first step would be to direct the Library to:

1. **Publish a CSV spreadsheet listing every bill pending before the House**, with number, primary title, sponsor (with the sponsor's THOMAS ID), status, and introduced date (in ISO 8601 format), updated as often as THOMAS is.

It should take no more than 1 month for the Library to make this information available. After this, other data fields could be added in an XML file. A complete list of core legislative data fields is included as an appendix to this letter. Additionally we ask that you:

2. **Ensure sufficient resources** are made available to this and other open data projects.

Other recommendations can be found in my testimony submitted for the record for the FY2014 legislative branch appropriations bill.⁸

Thank you for your time.



Joshua Tauberer
President
Civic Impulse, LLC

CC. Rep. Alexander, House Appropriations, chair, legislative branch
Rep. Wasserman Schultz, House Appropriations, ranking member, legislative branch
Bob Reeves, deputy clerk

⁶ We "screen scrape" THOMAS.gov to get at the legislative data.

⁷ See 'Recommendations to the Bulk Data Task Force' at http://razor.occams.info/pubdocs/2012-08-24_bulk_data_recs.pdf

⁸ <http://razor.occams.info/blog/2013/03/18/public-comment-to-the-house-appropriations-legislative-branch-subcommittee-for-fy2014/>

Appendix: Legislative Data Field List

The following is a fairly comprehensive list of fields in databases maintained by the Library of Congress, Government Printing Office, and House Clerk related to core legislative information. This is the data advocates refer to when we speak of “bulk legislative data.” High-priority items are in bold.

Bill & Resolution Status (Library of Congress “bill summary & status”)

- **Congress number, bill type (properly normalized), bill number**
- **Sponsor** (using THOMAS’s legislator IDs, or for original bills the introducing committee using THOMAS’s committee IDs, or for bills without a sponsor an indication of that)
- Whether bill is introduced “by request”
- **Date of introduction** (ISO 8601 format, e.g. YYYY-MM-DD)
- **Primary title**
- Title list (whether title is for a portion of a bill, the title type (official/short/popular), the status of the bill when the title was applied (e.g. introduced, passed house), a sequence number to indicate which entries are newer than others, whether the title was 'identified by CRS')
- **Cosponsors** (using THOMAS legislator IDs) w/ join and withdrawn dates (ISO 8601 format), distinguishing original cosponsors from those that join later
- **Committee/subcommittee referrals** (using THOMAS committee IDs, w/ last committee action, and indication of primary referral where available)
- **Subject terms** (using a persistent unique identifier per term)
- **Subject term taxonomy** (a complete listing of terms, relations between terms and top-terms, when terms were added, changed, and removed, and historical terms)
- **Related bills** (related bill identified by ID, relation, e.g. identical, rule, identified by)
- CRS summary (summary date, “as of” bill status, summary text, and the complete set of past summaries for a bill as new summaries are written)
- Related resources (e.g. links to CBO reports, committee reports, identified by standard citation/SuDoc numbering where applicable)
- **Actions**
 - Whether the action is in the House or Senate (where appropriate)
 - Date or date & time of action (ISO 8601 format)
 - A sequence number to preserve the order of multiple actions on the same date w/o times
 - Committee/subcommittee associated with the action, if any (using THOMAS’s committee IDs)
 - Action text
 - Congressional Record references
 - Other links (e.g. roll call votes)
 - House Floor action type and/or cross-reference to House Floor XML
- List of amendments (see below)
- **Master list of all bills/resolutions and ability to efficiently get updated records.**
- All of this for 93rd Congress and forward.

Amendment Status (Library of Congress “bill summary & status”)

- **Congress/chamber**
- **Bill or amendment amended** (properly identified by ID)
- **The THOMAS ‘H.Amdt.’ number**
- The House Clerk ‘A_____’ number
- The House Rules committee report number
- The Congressional Record print number
- **Sponsor** (see notes for bills above)
- **Offered date**
- **Description/Purpose**

- **Actions** (see for bills above)
- Congressional Record references

Roll Call Votes (House Clerk)

- **Related amendment using THOMAS H.Amdt. number**
- Reason for missing Members (i.e. Speaker not included)
- Context within Congressional Record

Committees & Subcommittees (House Clerk)

- Membership using bioguide IDs
- Membership seniority, titles (chair, ranking member, etc.)
- Mailing address, website, etc.
- **Committee votes**

List of Members (House Clerk)

- Office mailing addresses, etc.
- Leadership, seniority information
- **Current vacancies**
- Sworn-in dates

Bill Text (GPO)

- **Improved sitemaps** (list downloadable files, include file hashes)
- Explain reasons for missing bills in FDSys
- Historical bill text in HTML (103rd Congress and forward)

Other Data:

- **Text of Amendments**
- Bioguide.congress.gov database (text, photo, citations)
- Congressional Record in HTML, XML
- Historical bills (American Memory collection, etc.)
- Historical Membership of Congress (dates of swearing in, resignations, etc.)
- Historical committees & caucuses (dates of formation, membership with dates of joining the committee/caucus and roles)
- Communications to the Senate and executive branch
- Franked materials and other miscellaneous filed disclosure forms