So what's left to do?

Joshua Tauberer (@JoshData)
GovTrack.us

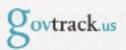
Legislative Data & Standards May 22, 2013



All legislative events are recorded in structured data.

All legislative artifacts are publicly available.

(How hard could that be?)



Legislative Data

- Bill Summary & Status
- Amendment Status & Text
- List of Members
- Committee Artifacts
- Historical Bill Text, Statutes, and so on.



[First prepared October 17, 2012. Updated May 19, 2013.]

What is Legislative Metadata?

The following is a comprehensive list of fields in House, Library, and GPO databases of primary legislative materials and, marked by asterisks, information that *ought to* be properly recorded.

Bills & Resolutions (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 (e.g. YYYY-MM-DD)
- · Primary title
- Title list (for each title, whether title is for a portion of a bill, the title type (e.g. 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 w/join and withdrawn dates,
 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 (a unique identifier per term if possible; w/ identification of which is the 'top term')
- Related bills (related bill properly identified by ID, relation (e.g. identical, rule))
- 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 (see below)
- List of amendments (for amendments, see below)
- Master list of all bills, resolutions, amendments
- Omit "reserved for speaker" bills if these are not actually introduced measures

Actions (Library of Congress "bill summary & status"):

- Whether the action is in the House or Senate (where appropriate)
- Date or date & time of action (ISO data/time 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)
- · Roll call vote data

Amendments (Library of Congress "bill summary & status"):

- · Congress/chamber/number
- Bill amended (properly identified by ID)

- Amendment amended, when relevant (with the amendment properly identified by ID)
- The sequence number of the amendment for the bill (i.e. 1st amendment for that bill, and so on)
- · Sponsor (see notes for bills above)
- Introduced date
- Whether the amendment was "submitted" or "offered" (if this is a real distinction)
- · Description/Purpose
- Actions (see above)

List of Members (Clerk):

- · Members identified by bioguide ID
- First name, last name, state, district, party, place of residence, etc.
- · Mailing addresses, etc.
- Leadership information, seniority information
- · Current vacancies

Committees (Clerk, et al):

- Committees/subcommittees identified with unique IDs
- · Membership using bioguide IDs
- Membership indicating seniority, titles (chair, ranking member, etc.)
- Mailing address, website, etc.

- · List of meetings scheduled
- *Committee votes
- *Committee prints
- *Committee reports
- Additional committee records such as transcripts, witness lists and testimony

Legislative text & documents:

- Current bill text
- Historical born-digital bill text (in GPO's existing structured data format called Locator Codes)
- *Historical bills (American Memory collection, etc.)
- *Amendment text
- Congressional Record (in GPO's existing structured data format called Locator Codes)
- Historical Congressional Record, Congressional Globe, etc.

Other Data:

- Bioguide.congress.gov database (text, photo, citations)
- Historical Membership of the House (dates of swearing in, resignations, etc.)
- Historical committees & caucuses (dates of formation, membership with dates of joining the committee/caucus and roles (e.g. chair))
- Communications to the Senate and executive branch



What to do first?

	Law	Services	Operations	Public Data
FOI				
Online & Free				
Open				
Structured				
Global IDs				
APIs				
Linked Data				

http://opengovdata.io/maturity/

	Law	Services	Operations	Public Data		
FOI	The legislative branch is not subject to the Freedom of Information Act.					
Online & Free	The US Code is online.	Constituent mail via web.	Voting records, disbursements online.	Where are congressionally mandated reports?		
Open Meets 7 Principles.	All records are open.	Service info is non- discriminatory.	All records are open.			
Structured Analyzable. Processable.	No legislative metadata.	No data for finding your rep or submitting constituent mail.	Voting records in XML. Disbursements are PDFs.			
Global IDs URLs & URIs.	Paragraph-level citation markup.		Bioguide IDs.			
APIS Provides random or write access.	What is 17 USC 1201?	Constituent mail submission system.				
Linked Data Lives on the semantic web.						