Bulk Data Task Force Meeting
April 27, 2016

GitHub repo

https://github.com/usgpo/bill-status

The Challenge

Legislative documents and its related data must be

- prepared,
- managed,
- distributed, and
- archived.

This includes paper and electronic means for handling the official documents

The Challenge (continued)

Today's challenge is not *whether* parliaments will have open sessions and publicly available documents, but rather *how* parliaments will modernize their systems and workflows to utilize new technologies.

110TH CONGRESS
1ST SESSION

H. R. 3712

To designate the Federal building and United States courthouse located at 1716 Spielbusch Avenue in Toledo, Ohio, as the "James M. & Thomas W.L. Ashley Customs Building and United States Courthouse".

IN THE HOUSE OF REPRESENTATIVES

OCTOBER 1, 2007

Ms. KAPTUR introduced the following bill; which was referred to the Committee on Transportation and Infrastructure

---

A BILL

To designate the Federal building and United States courthouse located at 1716 Spielbusch Avenue in Toledo, Ohio, as the "James M. & Thomas W.L. Ashley Customs Building and United States Courthouse".

1 *Be it enacted by the Senate and House of Representa-*
2 *tives of the United States of America in Congress assembled,*
3 **SECTION 1. JAMES M. & THOMAS W.L. ASHLEY CUSTOMS**
4 **BUILDING AND UNITED STATES COURT-**
5 **HOUSE.**
6 (a) DESIGNATION.—The Federal building and United
7 States courthouse located at 1716 Spielbusch Avenue in
8 Toledo, Ohio, shall be known as the "James M. & Thomas

1 W.L. Ashley Customs Building and United States Court-
2 house".
3 (b) REFERENCES.—Any reference in a law, map, reg-
4 ulation, document, paper, or other record of the United
5 States to the Federal building and United States court-
6 house referred to in subsection (a) shall be deemed to be
7 a reference to the "James M. & Thomas W.L. Ashley Cus-
8 toms Building and United States Courthouse".

○

The paper version is the official document of record.

Government Data Should Be:

- Accessible
- Accurate
- Complete
- Described
- Timely

- Usable
- Free
- Machine-readable
- Permanent
- Searchable

Policy Changes

[Print]   [Print selection]   [Close] He

§181. Program for exchange of information among legislative branch agencies

(a) On September 16, 1996, there shall be established a program for providing the widest possible exchange of information among legislative branch agencies with the long-range goal of improving information technology planning and evaluation. The Committee on House Oversight of the House of Representatives and the Committee on Rules and Administration of the Senate are requested to determine the structure and operation of this program and to provide appropriate oversight. All of the appropriate offices and agencies of the legislative branch as defined below shall participate in this program for information exchange, and shall report annually on the extent and nature of their participation in their budget submissions to the Committee on Appropriations of the House of Representatives and the Committee on Appropriations of the Senate.

(b) As used in this section-

(1) the term "offices and agencies of the legislative branch" means, the office of the Clerk of the House, the office of the Secretary of the Senate, the office of the Architect of the Capitol, the Government Accountability Office, the Government Publishing Office, the Library of Congress, the Congressional Research Service, the Congressional Budget Office, the Chief Administrative Officer of the House of Representatives, and the Sergeant at Arms of the Senate; and

(2) the term "technology" refers to any form of computer hardware and software; computer-based systems, services, and support for the creation, processing, exchange, and delivery of information; and telecommunications systems, and the associated hardware and software, that provide for voice, data, or image communication.

(Pub. L. 104–197, title III, §314, Sept. 16, 1996, 110 Stat. 2415 ; Pub. L. 108–271, §8(b), July 7, 2004, 118 Stat. 814 ; Pub. L. 113–235, div. H, title I, §1301(b), Dec. 16, 2014, 128 Stat. 2537 .)

Policy Changes: task force created

On June 1, 2012, House Report 112-511 that accompanied H.R. 5882 Legislative Branch Appropriations Act of 2013 directed the establishment of the House Bulk Data Task Force to examine increased dissemination of congressional information via bulk data download from non-governmental groups supporting openness and transparency in the legislative process.

Project Timeline: Bill Status Data

- December 7, 2015 - Sample Congress.gov Bill Status data at https://github.com/usgpo/bill-status

- December 15, 2015 - BDTF hosted Bill Status XML panel and fielded questions

- February 24, 2016 - Bill Status data at https://www.gpo.gov/fdsys/bulkdata/BILLSTATUS

- April 27, 2016 - BDTF hosted Bill Status XML panel and fielded questions

- July 5, 2016 – Projected THOMAS retirement date

Announcement:

http://www.loc.gov/today/pr/2016/16-004.html?loclr=ealn

Technology Timeline
http://xml.house.gov/resources/TechTimeline.htm

Following bills in Congress requires three major pieces of information:

″ the text of the bill,

″ a summary of what the bill is about, and

″ the status of the bill.

Thanks to Daniel Schuman with the Congressional Data Coalition for blogging about this project on 12/8/2015.

The original request: bulk data

https://www.gpo.gov/fdsys/bulkdata

Text of Bills

2013: House Bills were made available as XML bulk data on the FDsys Bulk Data Repository. Senate bills were added to this collection in January 2015.

Summaries of Bills

2014: House Bill Summaries were made available as XML bulk data on the FDsys Bulk Data Repository. Senate bill summaries were added to this collection in January 2015.
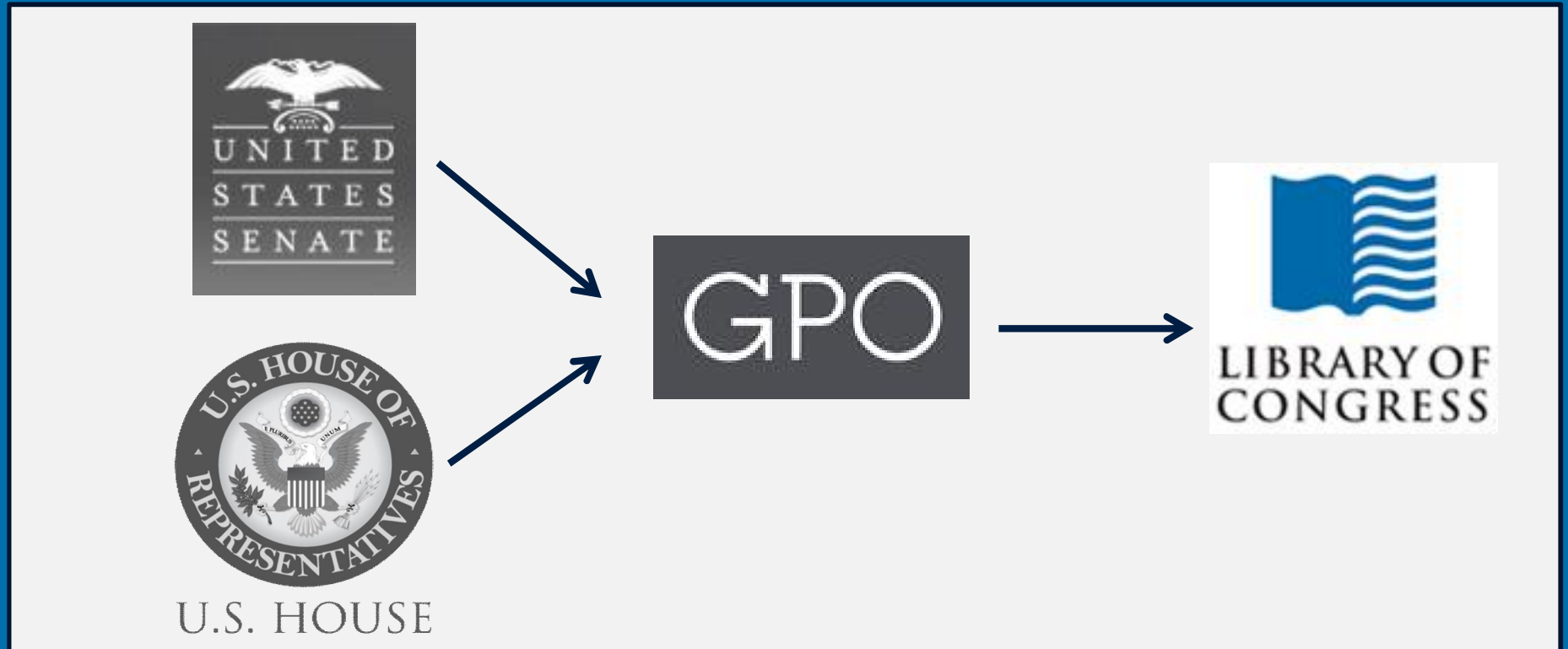
Status of Bills: Released February 2016
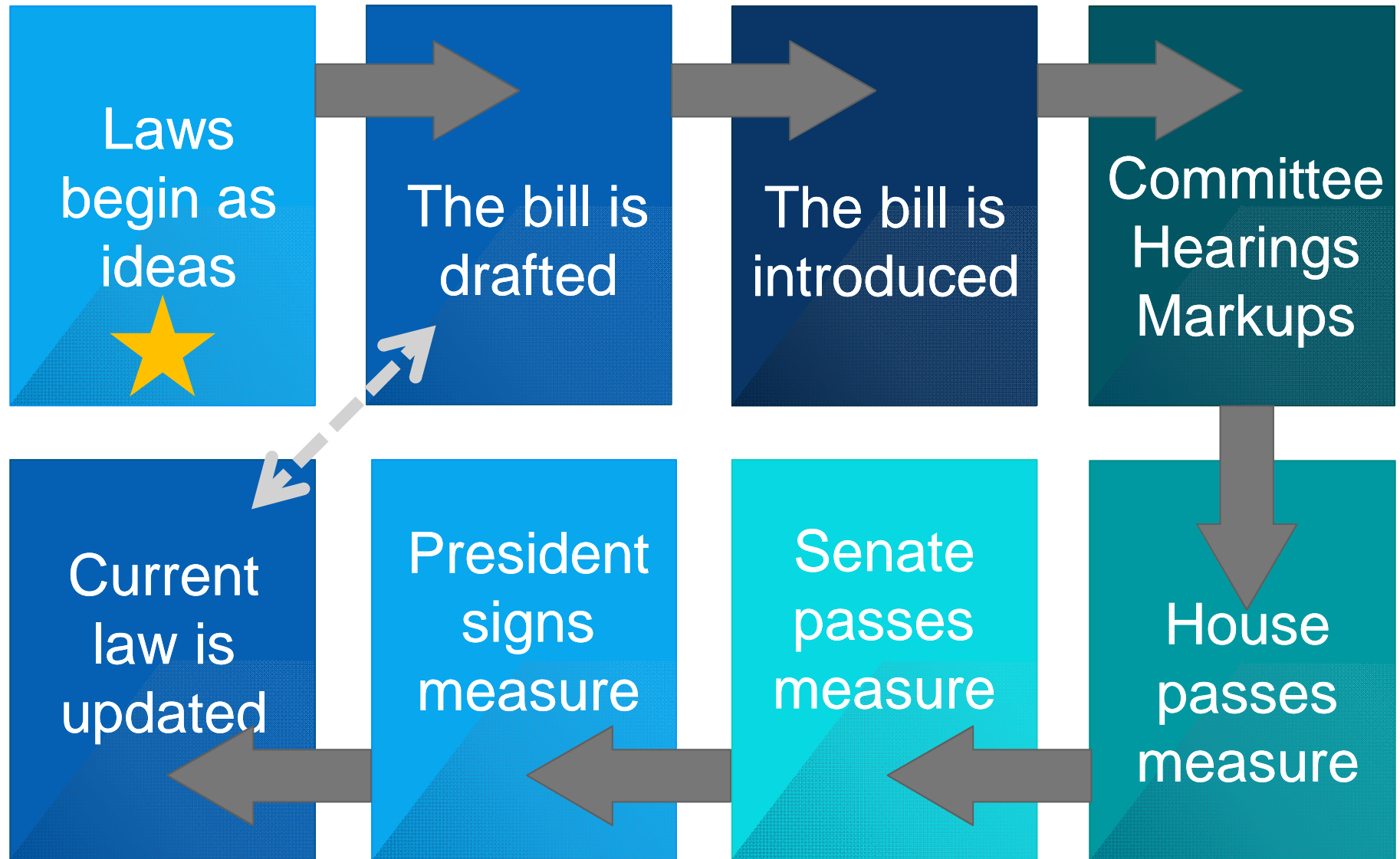
Bill Status in Bulk

The FY2014 legislative branch appropriations committee report (H.Rept.113-417 – page 5).

The Committee request that the Clerk of the House, the Librarian of Congress and the Public Printer work together to make available to the public through Congress.gov or FDsys bulk data downloads of bill status by the beginning of the next Congress.
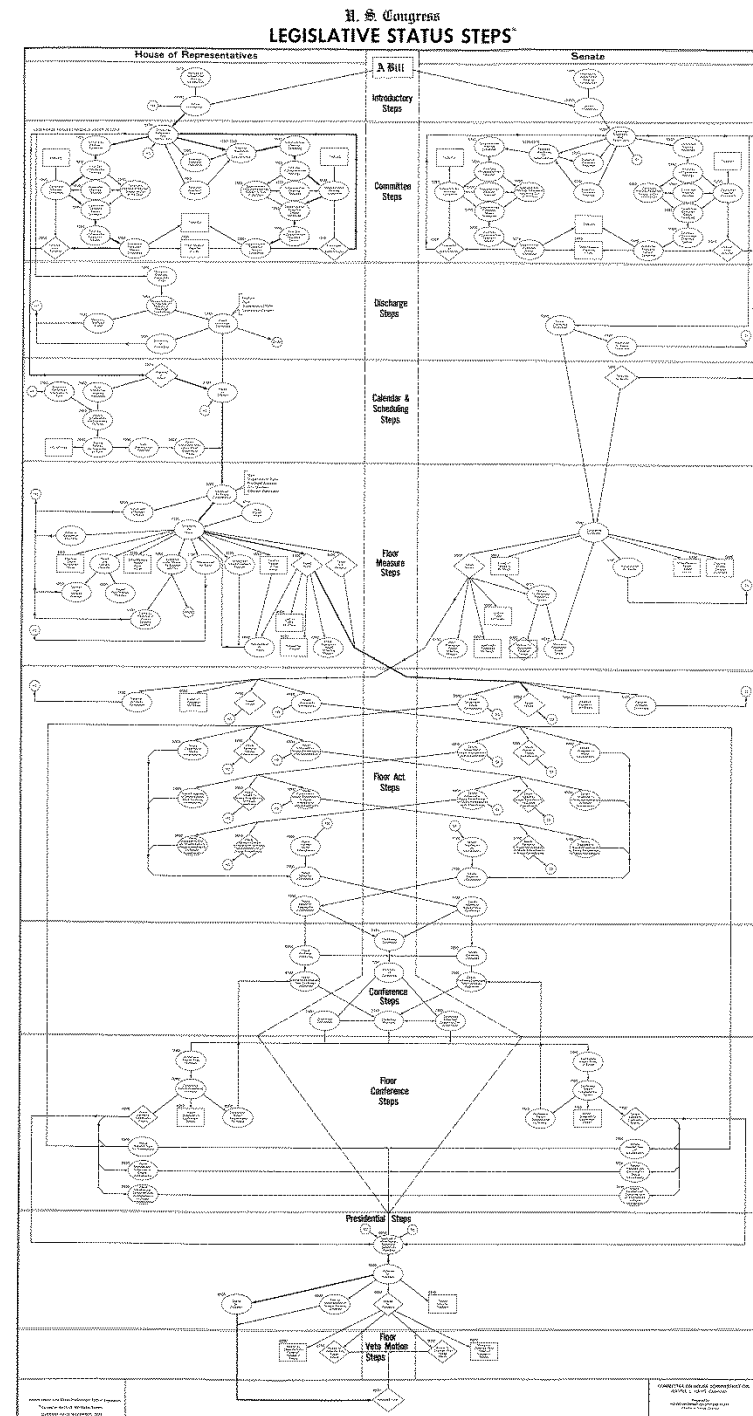
Simple Workflow for the documents
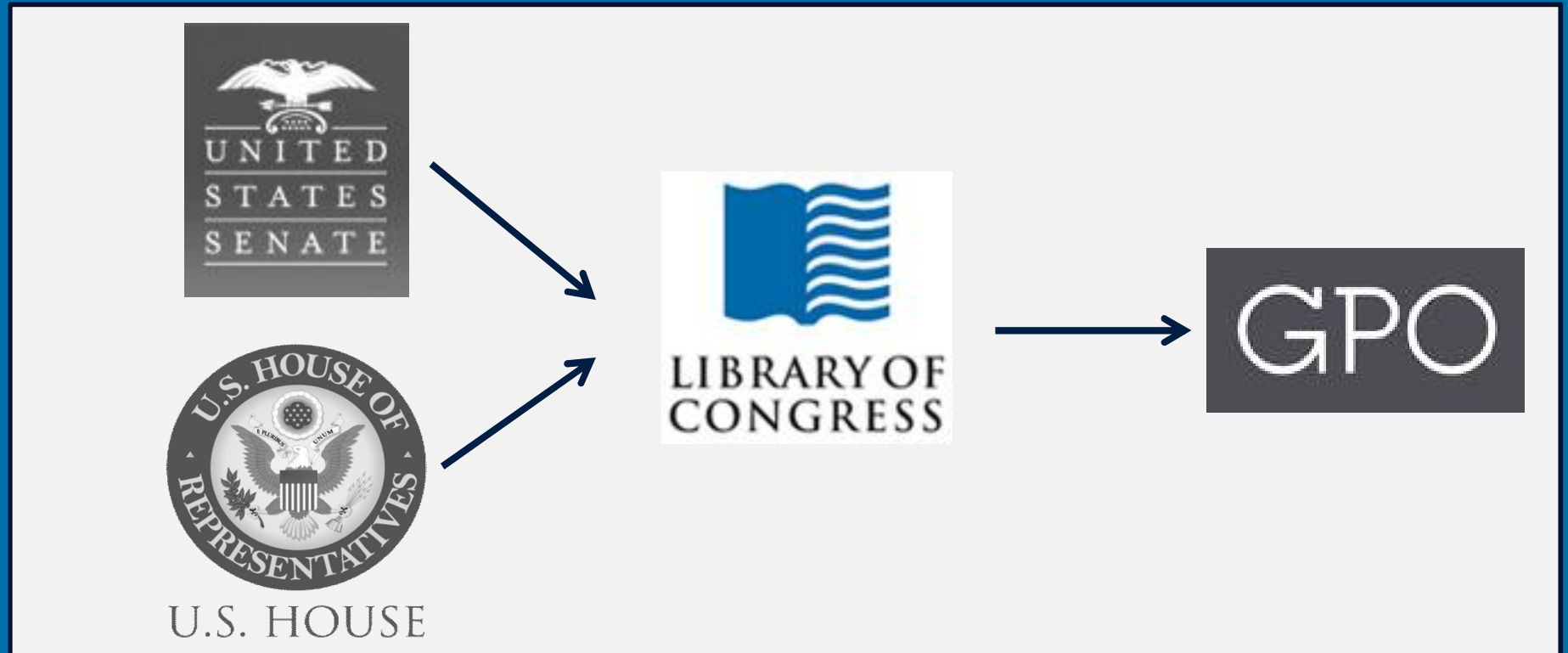
How a Bill Becomes the Law

Laws begin as ideas ⭐	The bill is drafted	The bill is introduced	Committee Hearings Markups
Current law is updated	President signs measure	Senate passes measure	House passes measure

The shortest
success path for a
bill to become law
is twenty decision
points

U. S. Congress
LEGISLATIVE STATUS STEPS

Simple Workflow for legislative status

Bill Status in Bulk

- The repository is organized by Congress and bill type

- Zip files for each bill type for a specific Congress

- One XML file for each bill or resolution

Index of /bulkdata/BILLSTATUS/113/hr

Parent Directory

Name	Last Modified	Size
BILLSTATUS-113hr83.xml	16-Sep-2015 12:06	0.15 M
BILLSTATUS-113hr3300.xml	16-Sep-2015 12:06	0.02 M
BILLSTATUS-113hr2748.xml	16-Sep-2015 12:06	0.03 M
BILLSTATUS-113hr26.xml	16-Sep-2015 12:06	0.07 M
BILLSTATUS-113-hr.zip	16-Sep-2015 12:06	0.04 M

Bill Status in Bulk

- At 8:00 AM, GPO retrieves a list of legislative measures that have been added or updated

- GPO processes the added and updated legislative measures to create each Bill Status XML file

- Processing time is generally 2 to 20 seconds per legislative measure

Bill Status in Bulk

- Usually less than 1,000 additions or updates per day, unless data set needs to be reprocessed

- Changes recorded in Bill Status Bulk Data sitemap

- Example for H.R. 1508 (114th Congress)

  - https://www.gpo.gov/smap/bulkdata/BILLSTATUS/sitemapindex.xml

  - https://www.gpo.gov/smap/bulkdata/BILLSTATUS/114hr/sitemap.xml

  - <url><loc>https://www.gpo.gov/fdsys/bulkdata/BILLSTATUS/114/hr/BILLSTATUS-114hr1508.xml</loc><lastmod>2015-12-01T11:56:00.266Z</lastmod><changefreq>monthly</changefreq><priority>1.0</priority></url>

Top Level Elements

- billstatus, bill, actions, amendments, billNumber, billType, calendarNumber, cboCostEstimates, constitutionalAuthorityStatementText, committeeReports, committees, congress, cosponsors, createDate, introducedDate, isByRequest*, isReserved*, lastAction, laws, recordedVotes, notes, originChamber, primarySubject, relatedBills, sponsors, subjects, summaries, title, titles, updateDate, dublinCore

* isByRequest is no longer a top element; isReserved will be removed.

Top Level Elements

H.R.1531 - Land Management Workforce Flexibility Act ‹title›

114th Congress (2015-2016) | Get alerts

‹constitutionalAuthorityStatementText›

| LAW | Hide Overview ✕ |

More on This Bill

Constitutional Authority Statement

CBO Cost Estimates ‹cboCostEstimates›

‹sponsors› ‹introducedDate›

Sponsor: Rep. Connolly, Gerald E. [D-VA-11] (Introduced 03/23/2015)

Committees: House - Oversight and Government Reform | Senate - Homeland Security and Governmental Affairs

Committee Reports: S. Rept. 114-123; H. Rept. 114-182 ‹committeeReports›

‹primarySubject› Subject — Policy Area:

Government Operations and Politics

Latest Action: 08/28/2015 By Senator Johnson from Committee on Homeland Security and Governmental Affairs filed written report under authority of the order of the Senate of 08/05/2015. Report No. 114-123. (All Actions) **The elements ‹calendarNumber›, ‹laws›, ‹recorededVotes›, and ‹originChamber› are in Actions.**

‹lastAction›

View subjects »

‹subjects›

Tracker:

Introduced › Passed House › Passed Senate › To President › Became Law

‹summaries› ‹actions› ‹titles› ‹amendments› ‹cosponsors› ‹committees› ‹relatedBills›

| Summary (6) | Text (7) | Actions (21) | Titles (7) | Amendments (0) | Cosponsors (16) | Committees (2) | Related Bills (2) |

‹billType› ‹billNumber› ‹congress›

Summary: H.R.1531 — 114th Congress (2015-2016)

Other top level elements are ‹billstatus›, ‹bill›, ‹createDate›, ‹isByRequest›, ‹isReserved›*, ‹notes›, ‹updateDate›, and ‹dublinCore›.

All Bill Information (Except Text)

🔊 Listen to this page ▶

There are 6 summaries for H.R.1531. | Public Law (08/07/2015) ▾ | GO | Bill summaries are authored by CRS.

Thank you for your great feedback!

General Resources

- https://www.gpo.gov/fdsys/bulkdata/BILLSTATUS/resources

- https://www.congress.gov/help

- https://www.congress.gov/about/coverage-dates

- Field Value Lists

- Glossary

The process by which a bill becomes law is rarely predictable.

Mapping Bills Summaries and Bill Text

Summaries may map to committee reports that are issued in multiple parts, in addition to bill texts.

Two examples:

H.R.2647 - Resilient Federal Forests Act of 2015

H.R.3490 - Strengthening State and Local Cyber Crime Fighting Act

The Bill Status bulk data set is DOWNSTREAM from multiple systems that manage data about a rarely predictable process.

Responding to feedback – Data Elements

√ Data flow schedule adjustments: Update job changed to 8:00 AM instead of 4:00 AM to capture more of the early morning updates

√ Amendment: actions, sponsors, sponsor BioGuide ID

√ Timestamp removed from <sponsorshipDate>

√ actionTime for House floor actions

√ Normalize values for Related Bill <billType>

√ Adjust CBO child node attribute names

FEEDBACK: What's a Display Title?

It's used to identify which title should be used as the <H1>

114hr1314, 113s25, 113hr5313, 111hr3590

28

FEEDBACK: Not all of the action codes are documented, e.g. H34200

See User Guide section 3. Action Code Element Possible Values

FEEDBACK: Would it be possible to have separate attributes for state, party, district?

Yes, it is possible to provide separate elements for party, state, and district.   This request is being discussed by the stakeholders.

We recommend using BioGuide IDs as control numbers for this data set, and using the authoritative sources for member data available at

http://clerk.house.gov/xml/lists/MemberData.xml and http://www.senate.gov/general/contact_information/senators_cfm.xml

NOTE: there is a separate project to modernize http://bioguide.congress.gov

Please continue to add questions and requests to GitHub.

We are using this as a tracking tool.

Change Management

GitHub will be the location to receive advanced notice on changes that will be published.

Label Assignments in GitHub:

-Data Issue
-Documentation Issue
-Bug
-Enhancement

Comments?

https://github.com/usgpo/bill-status

xml-comments@mail.house.gov

or to presenters directly