Bulk Data Task Force Meeting
December 15, 2015

WiFi: HousePublic

PW is the same

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.

Government Data Should Be:

- Accessible
- Accurate
- Complete
- Described
- Timely

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

Policy Changes

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.

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.
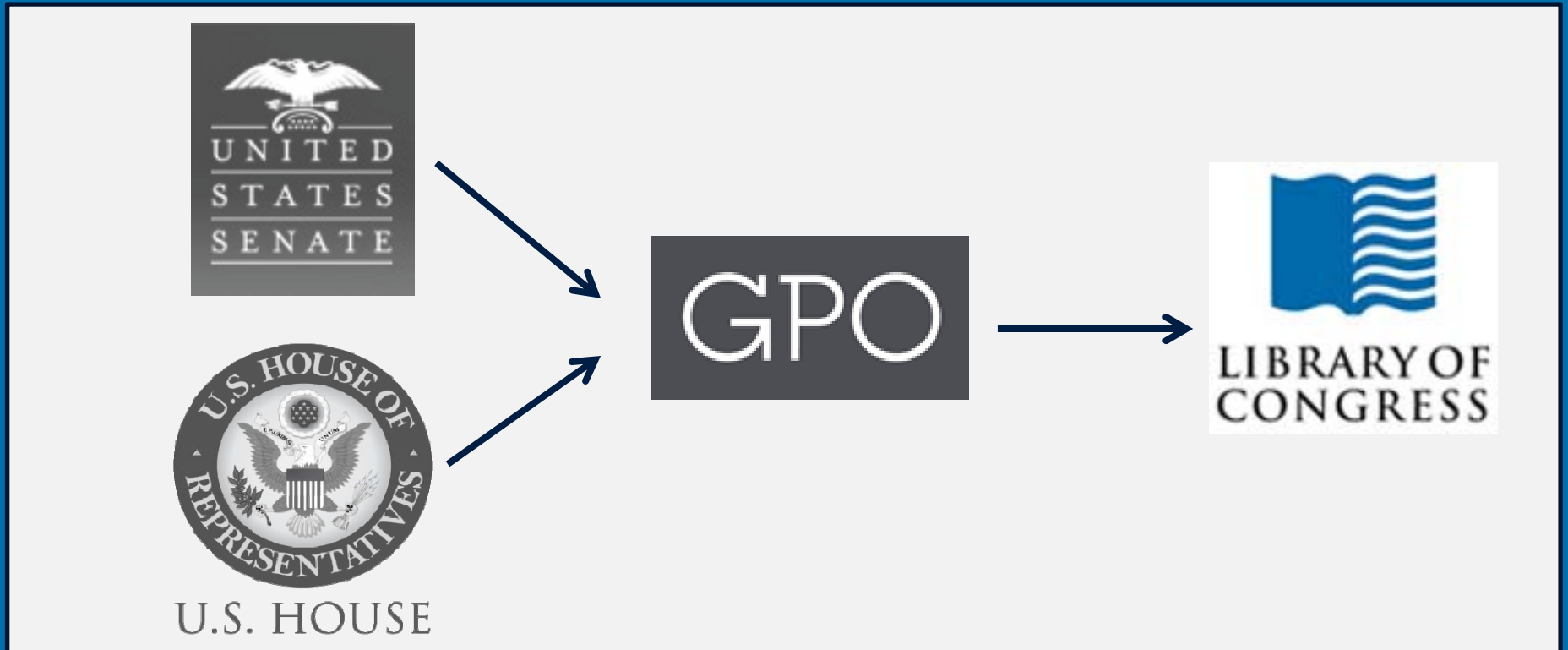
Tech Timeline: http://xml.house.gov/resources/TechTimeline.htm
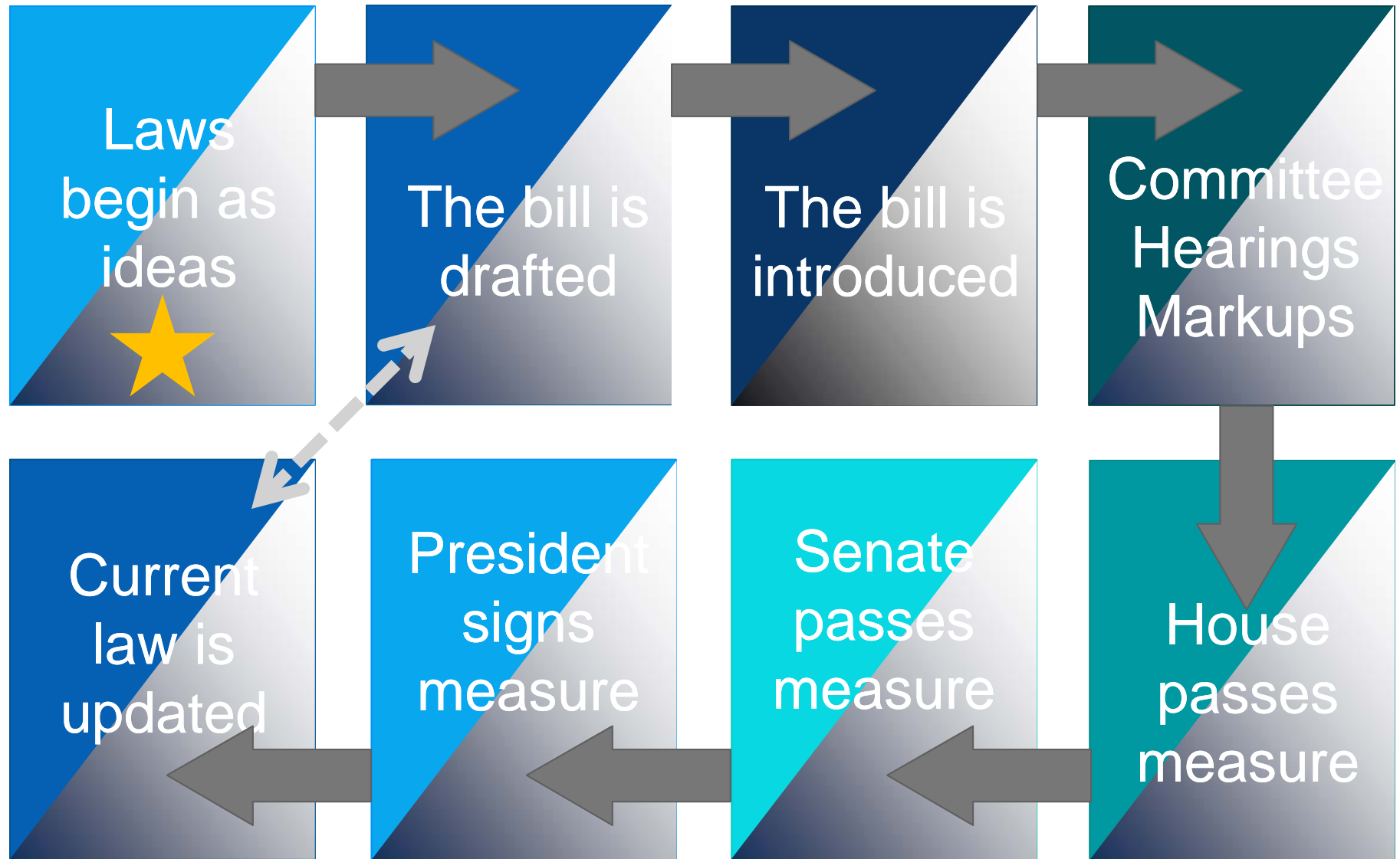
Bill Status in Bulk

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.
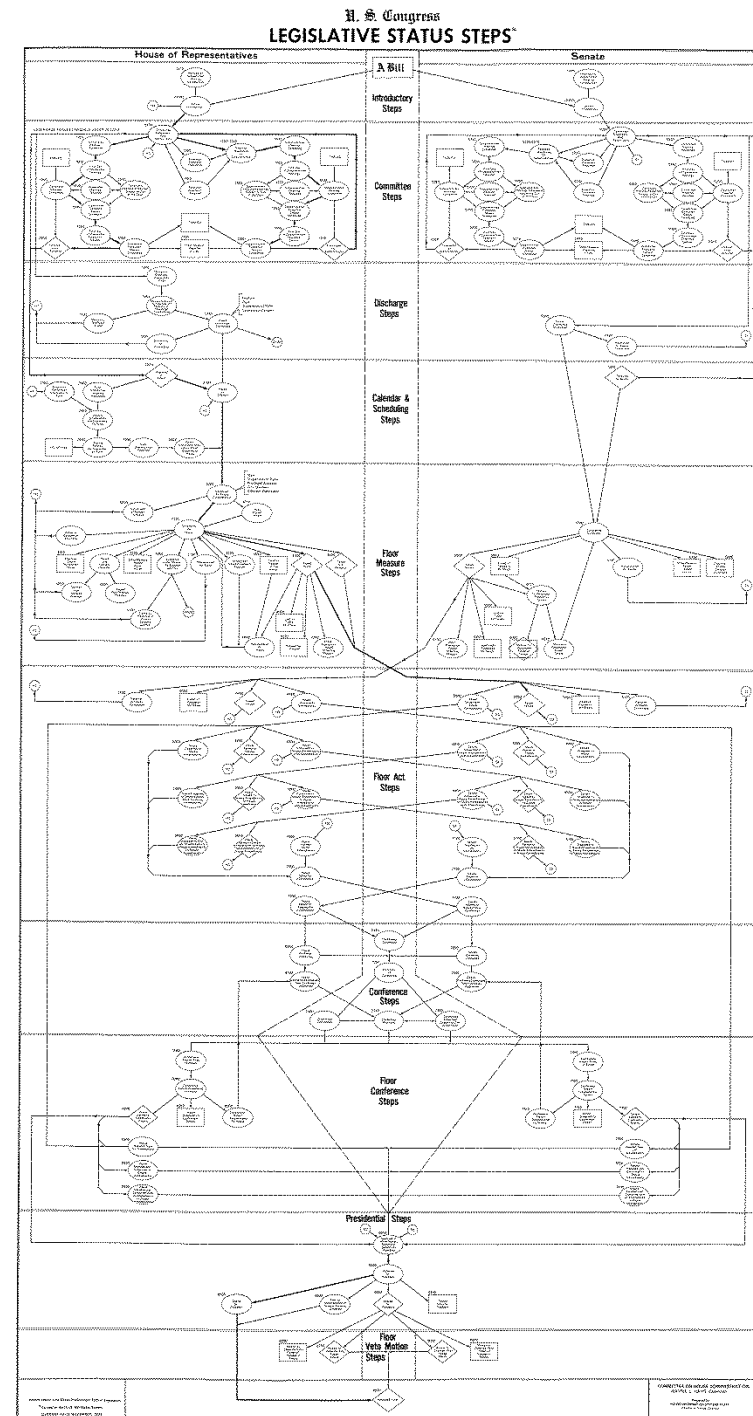
Simple Workflow for the documents
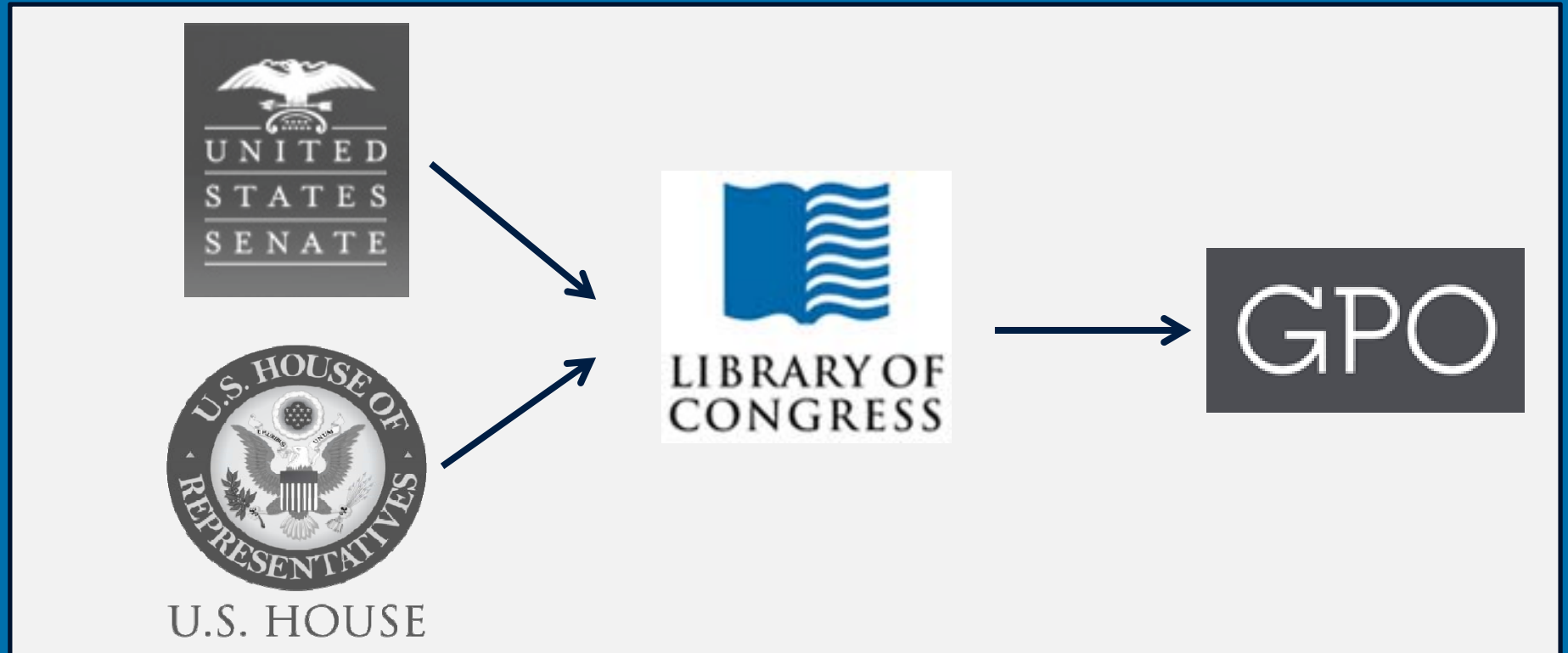
How a Bill Becomes the Law

U.S. Congress
LEGISLATIVE STATUS STEPS

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

http://www.loc.gov/pictures/resource/ppmsca.33996/

Simple Workflow for legislative status

Bill Status in Bulk

Where?

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

When?

Early 2016

Comments?

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

xml-comments@mail.house.gov or presenters directly

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 4:00 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

* This element will be removed in 2016.

Top Level Elements

H.R.1531 - Land Management Workforce Flexibility Act `<title>`

114th Congress (2015-2016) | Get alerts

`<constitutionalAuthorityStatementText>`

LAW   Hide Overview ✕

`<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>`

Latest Action: 08/28/2015 By Senator Johnson from Committee on Homeland Security and Governmental
`<lastAction>`   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.**

Tracker:

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

More on This Bill

Constitutional Authority Statement

CBO Cost Estimates `<cboCostEstimates>`

`<primarySubject>` Subject — Policy Area:
Government Operations and Politics

View subjects »
`<subjects>`

`<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.

When will THOMAS be retired?

A specific date for retiring THOMAS has not been determined, but the Library expects to be able to announce a date in early 2016. As part of the future announcement, an ample grace period will be allowed for users to adjust their processes and systems to the change.

For the past year, the Library has been actively encouraging THOMAS users to begin using the Congress.gov site. Most of the outstanding content and features of THOMAS have been incorporated into Congress.gov. Remaining items will be incorporated in early 2016.

Bills Summaries and Bill Text

Jim Harper
@Jim_Harper

Seeking rule for correlating bill text versions and summaries on @congressdotgov. @JoshData @danielschuman @GullicksonK @derekwillis Anyone?

9:15 AM - 13 Nov 2015

K Gullickson
@GullicksonK

@Jim_Harper @JoshData @danielschuman @derekwillis @congressdotgov remember loc summaries don't always map to a printed bill text one for one

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

Comments?

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

xml-comments@mail.house.gov

or to presenters directly

```xml
<floor_action act-id="H37300" update-date-time="20150707T14:44" unique-id=
"15433">
<action_time for-search="20150707T14:44:44">2:44:44 P.M. -</action_time>
<action_item>H.R. 1531</action_item>
<action_description>
On motion to suspend the rules and pass the bill Agreed to by voice vote.
</action_description>
</floor_action>
```

Floor summary XML July 7, 2015

```xml
<item>
  <actionCode>H37300</actionCode>
  <actionDate>2015-07-07</actionDate>
  <actionTime>02:44PM</actionTime>
  <committee />
  <links>
    <link>
      <name>H4782</name>
      <url>
      https://www.congress.gov/congressional-record/2015/house-section/page/H4782
      </url>
    </link>
  </links>
  <sourceSystem>
    <code>2</code>
    <name>House floor actions</name>
  </sourceSystem>
  <text>On motion to suspend the rules and pass the bill Agreed to by voice vote.
  (text: CR H4782)</text>
  <type>Final Passage Under Suspension of the Rules Results</type>
</item>
```

Bill status XML for H.R. 1531 114th Congress