The Congressional Data Task Force

Public Meeting

Thursday, September 10, 2026

Welcome

Kirsten Gullickson

Office of the Clerk
U.S. House of Representatives

Stephen Dwyer

Office of the Chief Administrative Officer
U.S. House of Representatives

The Congressional Data Task Force

The Task Force is a **partnership of staff** from across the Legislative Branch collaborating to modernize systems and improve access to congressional information.

Participating organizations include:

- House of Representatives
- Senate
- Library of Congress
- Government Publishing Office
- Government Accountability Office

The Congressional Data Task Force

Cross-organizational efforts and working groups

(ad-hoc, formal, and informal)

- XML Technical Working Group
- Data Interchange Working Group
- Video Preservation and Access Technical Working Group Working Group
- Digitization Of Congressional Documents Working Group
- Legislative Branch-wide Online Staff Directory Working Group

The Congressional Data Task Force

Since 2012*, the Task Force—

- Meets to advance transparency and modernization initiatives, and improve access to congressional information
- Holds public meetings
- Engages with civil society organizations

* On June 1, 2012, House Report 112-511 that accompanied H.R. 5882, the Legislative Branch Appropriations Act of 2013 directed the establishment of the Bulk Data Task Force

The Congressional Data Task Force

background material, meeting archives,
and the data catalog on
the Innovation Hub
https://usgpo.github.io/innovation/

email address
DataTaskForce@mail.house.gov

Agenda

1. Welcome and Background
2. Reports from working groups and legislative branch organizations
3. Report(s) from civil society organizations and others
4. Announcements/Closing

Meeting ends at 4:00 pm EST

Q&A will be held throughout the meeting

Housekeeping Items

The meeting will be recorded and posted online at https://usgpo.github.io/innovation/

Use the Chat feature for questions and comments; panelists will respond during the meeting.

Use the "raise hand" feature during discussion to be unmuted.

Ensure your name is displayed correctly.

Meeting will start at ~2:02 pm and end at 4:00 pm EST

The Congressional Data Task Force

The Task Force is a partnership of representatives from various Legislative Branch Agencies including the—

- House of Representatives,
- Senate,
- Library of Congress,
- Government Publishing Office,
- Government Accounting Office
- Congressional Budget Office.

Basic Data Exchange



More comprehensive workflows can be found on the Innovation Hub:
https://usgpo.github.io/innovation/resources/



PREPARE



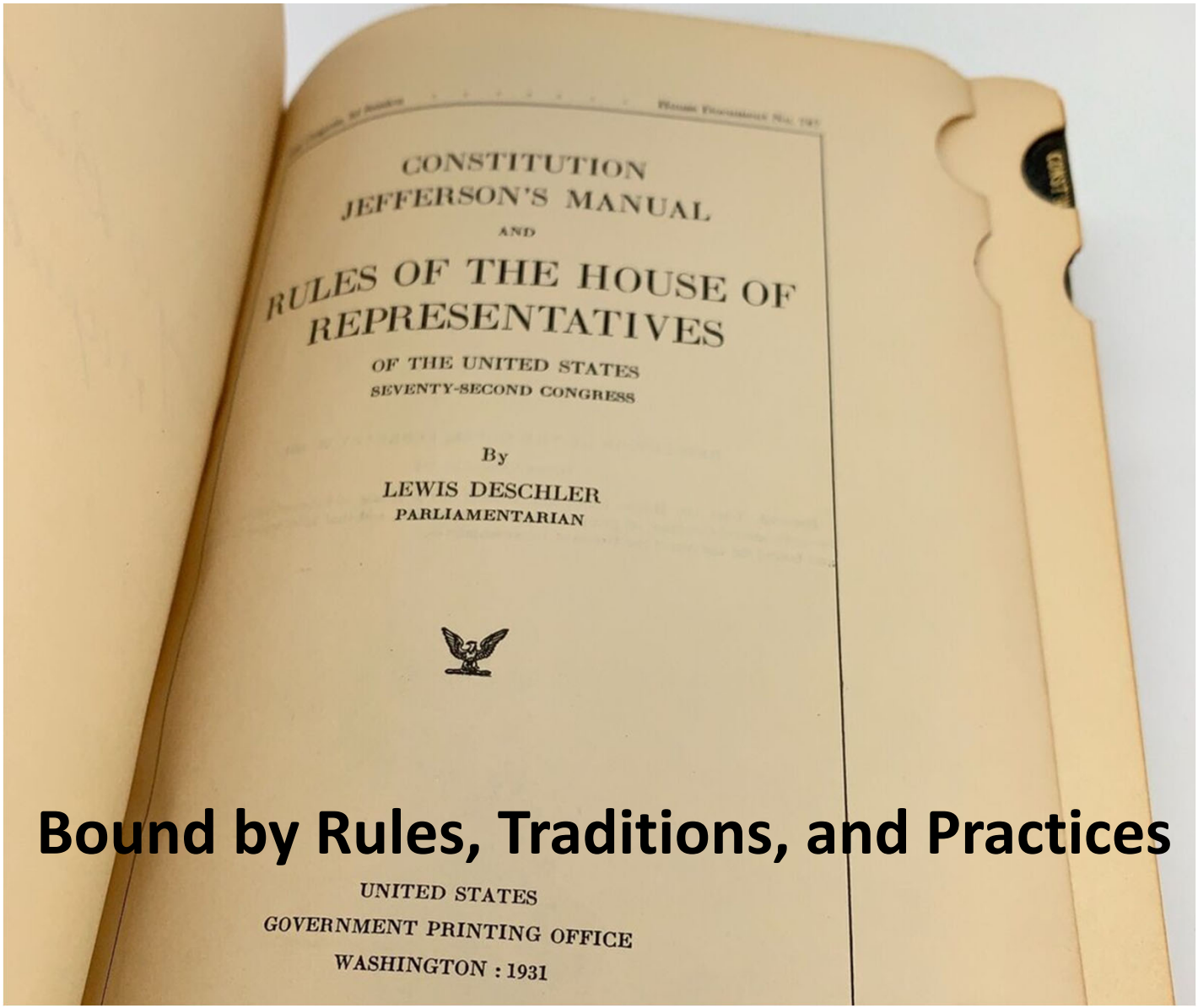
MANAGE



DISTRIBUTE

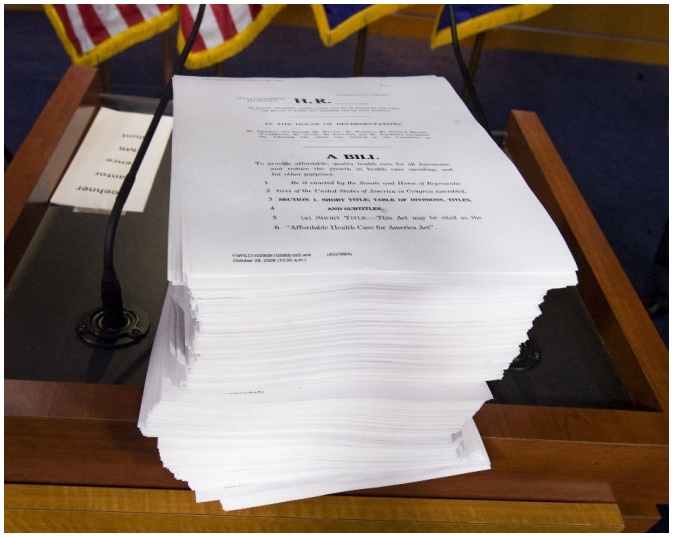


ARCHIVE/PRESERVE



Bound by Rules, Traditions, and Practices









Adding a digital layer to a centuries-old paper process

While the paper version remains the official document of record

Building and maintaining the digital layer

Common foundations preserve meaning and authority as congressional information moves.

STRUCTURE
How information is organized

SEMANTICS + METADATA
What information means and how it is described

STANDARDS
Common rules for creating and exchanging information

INTEGRATION + INTEROPERABILITY
Information moves reliably across systems and organizations

Information can be found, exchanged, understood, and reused while preserving context, provenance, and authority.

A SHARED LEGISLATIVE DATA ECOSYSTEM

Standards & Structure
USLM • XML • common identifiers • metadata

Exchange & Integration
APIs • data exchanges • interoperable systems

Discovery & Wayfinding
Legislative Branch Data Catalog • authoritative sources

Publication & Reuse
Congress.gov • GovInfo • APIs • bulk data

CONGRESSIONAL INFORMATION SHOULD BE:

- Accessible
- Accurate
- Trusted
- Complete
- Described
- Free
- Machine-readable
- Permanent
- Searchable
- Timely

AUTHORITATIVE INFORMATION FOR PEOPLE AND MACHINES

The Congressional Data Task Force

institutional knowledge • collaboration • trust

Next on the Agenda
September Events

- 9/18 at 2:00 PM: Constitution Day: The Life and Work of President and Chief Justice William Howard Taft.
- 9/24 at 1:00 PM: A Congress.gov Public Forum.

- 9/22 to 9/25

LegisTech Forum – 6th Edition, hosted by Bussola Tech in cooperation with the Organization of American States (OAS) and the American Governance Institute

Next on the Agenda

- Updates from the Government Publishing Office
  - Including DCAT Data Catalog

Data Catalog Vocabulary
DCAT-US



Structuring a way for developers to access machine readable congressional dataset





Creating a pathway for data efficiency for data stewards

- ✓ Know what data exists to help avoid data silos
- ✓ Identify data provenance

What is DCAT and why we choose this standard

- DCAT is a “Data Catalog Vocabulary” – started around 2014 - Providing a resource definition framework that allows for, discovery, exchange, and system interoperability of data catalogs and **datasets**.

- DCAT offers the foundation for an ontology in which the relationships between objects and properties are modeled. “Common language to describe **datasets** across various domains” allowing for better data discovery.

- DCAT can be used for its metadata standards with platforms like CKAN.

“Establishing a common vocabulary is the key to communication” (data.gov)

DCAT – US | Data Catalog Application Profile for the United States of America

The purpose of the DCAT-US 3.0 Profile is to improve data discoverability, accessibility, and interoperability among US government agencies. By adhering to the **FAIR** principles, the profile promotes more effective data sharing and reuse. The FAIR principles emphasize that data should be:

- **Findable**: Easy to discover by humans and machines.
- **Accessible**: Easily retrieved, with well-defined access protocols.
- **Interoperable**: Ready for use with other data sources and applications.
- **Reusable**: Clearly documented and licensed, facilitating reuse by others.

DCAT-US Required Fields

Field	DCAT Property	Why Developers Need It
Dataset ID	`dct:identifier`	Unique reference
Dataset Title	`dct:title`	Human-readable name
Description	**`dct:description`**	**What the dataset contains**
Publisher	`dct:publisher`	Which office owns it
Contact Point	`dcat:contactPoint`	Where developers ask questions
Keywords	`dcat:keyword`	Improves search
Theme	**`dcat:theme`**	**Grouping of datasets**
Last Updated	`dct:modified`	Shows freshness
Update Frequency	`dct:accrualPeriodicity`	Helps developers plan integrations

Next on the Agenda

- Demonstration

Q&A/Discussion

Use the “raise hand” feature so we can unmute your mic.

Use the Q&A feature to ask questions. The panelists will respond during the meeting.

Make sure the name displayed is you.

Next on the Agenda

- **Congressional staff modernization association**
  - **Ananda Bhatia**
  - **Paul Hurton**



MODERNIZATION STAFF ASSOCIATION

MSA is a nonpartisan, bicameral staff association working to improve the efficiency and effectiveness of Congress.

Become an MSA member today!

Next on the Agenda

Reports from

- Senate, Office of the Secretary
- House, Office of the Chief Administrative Officer, House Digital Service
- House, Office of the Clerk, Legislative Computer Systems

Modernization Reports

Another set was submitted in July, which included

- **Standardized Formats for Legislative Documents, USLM**
  - Currently modeling CBO costs estimate letters and reports
- **Comparative Print Suite**
  - Demo later in the meeting
- **Committee Activity Portal, Committee Votes as Data**
  - Pilot group for Committee Activity Portal
- **Unique IDs for Reports Filed by Registered Lobbyists**
  - a unified Congressional search website – lda.gov

https://cha.house.gov/modernization-reports

CHANGE OF CONGRESS

Looking ahead to the 120th Congress Members-elect, U.S. House

If elected, 119th Members who may serve in different districts in the 120th: 11

119SSDD	120SSDD	Last Name	First Name
CA03	CA06	Kiley	Kevin
CA06	CA03	Bera	Ami
CA38	CA41	Sánchez	Linda
CA41	CA40	Calvert	Ken
FL22	FL23	Frankel	Lois
FL23	FL25	Moskowitz	Jared
FL25	FL20	Wasserman Schultz	Debbie
TX35	TX37	Casar	Greg
UT01	UT02	Moore	Blake
UT02	UT03	Maloy	Celeste
UT03	UT04	Kennedy	Mike

- **Former Members running again**
  - Jerry Carl (2021–2025) – AL-01
  - Hilda Solis (2001–2009) – CA-38
  - Melissa L. Bean (2005–2011) – IL-08
  - Colin Allred (2019–2025) – TX-33
  - Ben McAdams (2019–2021) – UT-01
  - Elaine Luria (2019–2023) – VA-02
  - Tom Perriello (2009–2011) – VA-05

More of this type of information at the December CDTF meeting.

Unique Identifiers

Members

BioGuide IDs

– A000370

https://bioguide.congress.gov/search/bio/A000370
https://clerk.house.gov/members/A000370

House Committees/Subcommittees

Historic four-digit code

– AG00

https://clerk.house.gov/committees/AG00
https://clerk.house.gov/committees/AG14

https://clerk.house.gov/xml/lists/MemberData.xml

Release of Data

- Unofficial Members-elect data will be released when it is ready in mid-to-late December
- Official Member data will be released for the 120th Congress on opening day

---

- For duplicate names and salutations such as Ms/Mrs, use
  - https://clerk.house.gov/member_info/duplicate-names-119.pdf
  - *Has a change log on the back page that is helpful throughout the Congress.*

Q&A/Discussion

Use the “raise hand” feature so we can unmute your mic.

Use the Q&A feature to ask questions. The panelists will respond during the meeting.

Make sure the name displayed is you.

Next on the Agenda

Reports from civil society organizations and others

Next on the Agenda

Demonstration

by

Will Adler from the Bipartisan Policy Center

FloorVote, a new open-source platform for teams to discuss and collaborate on legislation.

It lets team members comment, vote, and set team positions; provides AI summaries; and keeps members up to date with email updates about prioritized bills.

Next on the Agenda

Demonstration

By

Dave Zvenyach

Two projects related to **appropriations** reports

- Extracting data from committee reports
- Comparing committee report language

Q&A/Discussion

Use the “raise hand” feature so we can unmute your mic.

Use the Q&A feature to ask questions. The panelists will respond during the meeting.

Make sure the name displayed is you.



Roadmap for Publishing Legislative Documents in USLM XML Format[1]

Completed	U.S. Code
Completed	Enrolled bills, public laws, and the Statutes at Large
Completed	HOLC/SOLC statute compilations
Completed	Remaining bill/resolution versions (Introduced, reported, engrossed, all amendments, et. al.)
Initial Deliverables Completed	Committee Reports (legislative and conference)
D	House portions of the *Congressional Record*, including the *Daily Digest*
E	Hearing Records
F	House Calendar
G	*Precedents of the U.S. House of Representatives*
H	*House Journal*

[1] For more information about this roadmap, please see page four of the initial report at https://usgpo.github.io/innovation/resources/reports/Clerk-Initial-Standardized-Formats.pdf.

Next on the Agenda

Reports from

- Library of Congress, Congress.gov Team

Next on the Agenda

Demonstration of two applications used inside the House and Senate

- The Clerk's Office, Legislative Computer Systems
- Library of Congress, Congressional Research Service



The Comparative Print Suite displays legislative changes in context, including how a bill might change current law and how two versions of a legislative proposal are different



The Text Analysis Program (TAP) is a tool that enables users to fin and compare measures with similar text.

Q&A/Discussion

Use the “raise hand” feature so we can unmute your mic.

Use the Q&A feature to ask questions. The panelists will respond during the meeting.

Make sure the name displayed is you.

AGENDA

- **Announcements**
- **Closing**

Employment Opportunities
usajobs.gov

- House: https://www.house.gov/employment/
  - *CAO: https://house.csodfed.com/ux/ats/careersite/1/home*
  - *Clerk: https://house.csodfed.com/ux/ats/careersite/5/home*
- Senate: https://www.senate.gov/visiting/employment.htm
- GPO: https://www.usajobs.gov/Search/?a=LP00
- LOC: https://loc.gov/careers/

September Events

- **9/18 at 2:00 PM**: Constitution Day: The Life and Work of President and Chief Justice William Howard Taft.
- **9/24 at 1:00 PM**: A Congress.gov Public Forum.

- **9/22 to 9/25**

LegisTech Forum – 6th Edition, hosted by Bussola Tech in cooperation with the Organization of American States (OAS) and the American Governance Institute

Future Events

CONGRESSIONAL DATA TASK FORCE MEETINGS

Thursdays | 2 pm to 4 pm

March 26, 2026

June 11, 2026

September 10, 2026

December 3, 2026

Hybrid meetings; DC (Capitol Complex) location announced closer to each date.

DataTaskForce@mail.house.gov

Email address for the
Congressional Data Task Force

Legislative Branch Innovation Hub

https://usgpo.github.io/innovation/

Thank you for attending!

Congressional Data Task Force – Public Meeting