

The Congressional Data Task Force

Public Meeting

Tuesday, December 19, 2023

Welcome

Kirsten Gullickson

Office of the Clerk
U.S. House of Representatives

Kirsten.Gullickson@mail.house.gov

For more information:
USGPO.GITHUB.IO/INNOVATION/

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

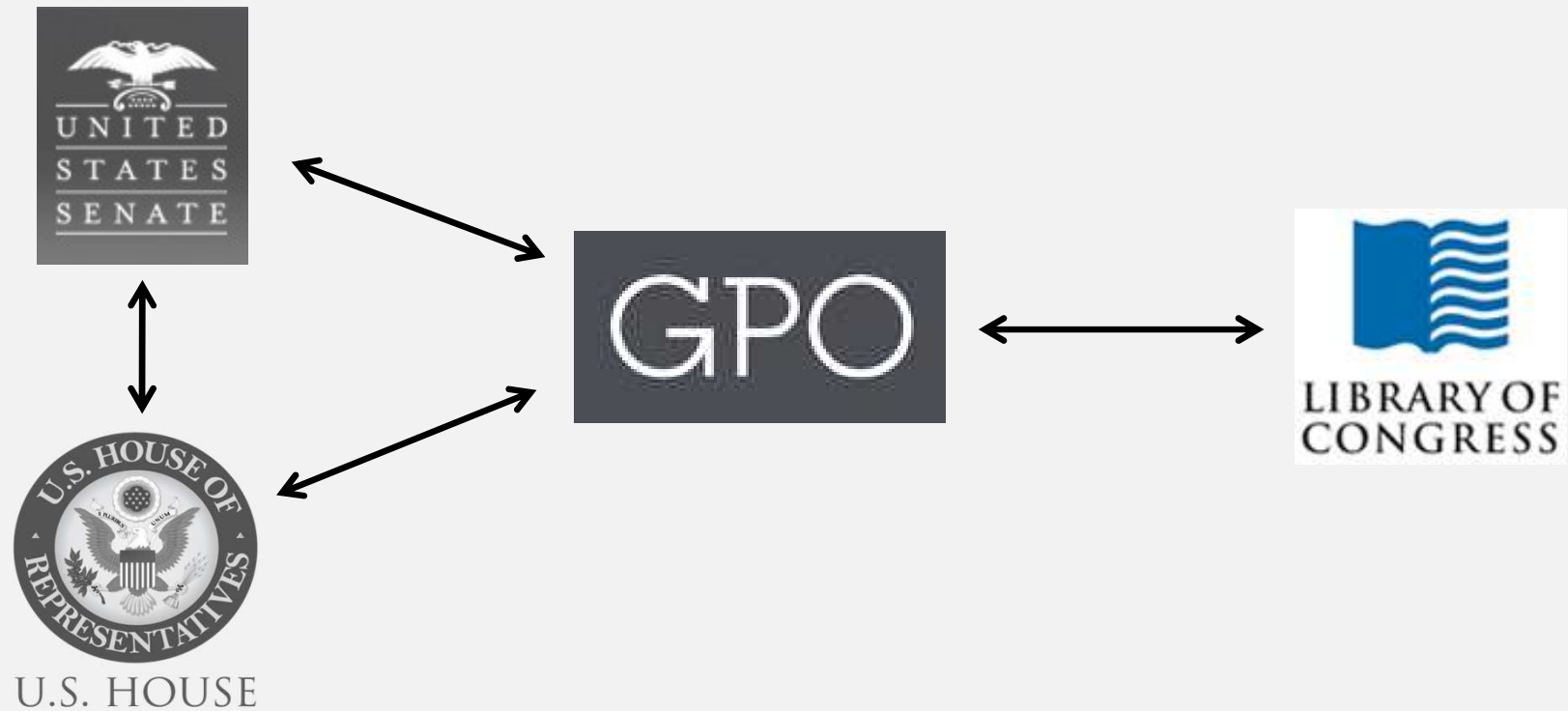
that work in a cooperative manner to coordinate and accomplish openness and transparency goals.

The Congressional Data Task Force

- The Bulk Data Task Force was created in 2012 to focus on increased dissemination of congressional information via bulk data
- *current work goes beyond bulk data*
 - meets internally with Legislative Branch organizations to coordinate projects and other transparency and modernization initiatives
 - meets with civil society organizations

The name changed to the Congressional Data Task Force on June 21, 2022

Data Exchange: Simple



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

Legislative Branch Innovation Hub

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

AGENDA

1. Welcome and Background
2. Report from GPO on Congressional Mandated Reports Collection
3. Panel on Artificial Intelligence
4. Reports/Updates from CDTF Working Group and Legislative Branch organizations
5. Time for Questions and Answers
6. Announcements/Closing

Meeting ends at 4:00 pm EST

Housekeeping Items



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



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



Use the “raise hand” feature during the Q&A portion so we can unmute your mic.



Make sure the name displayed is you. Using the chat, please share a little about yourself. Suggestions: your name, where you work, and expectations for this meeting.



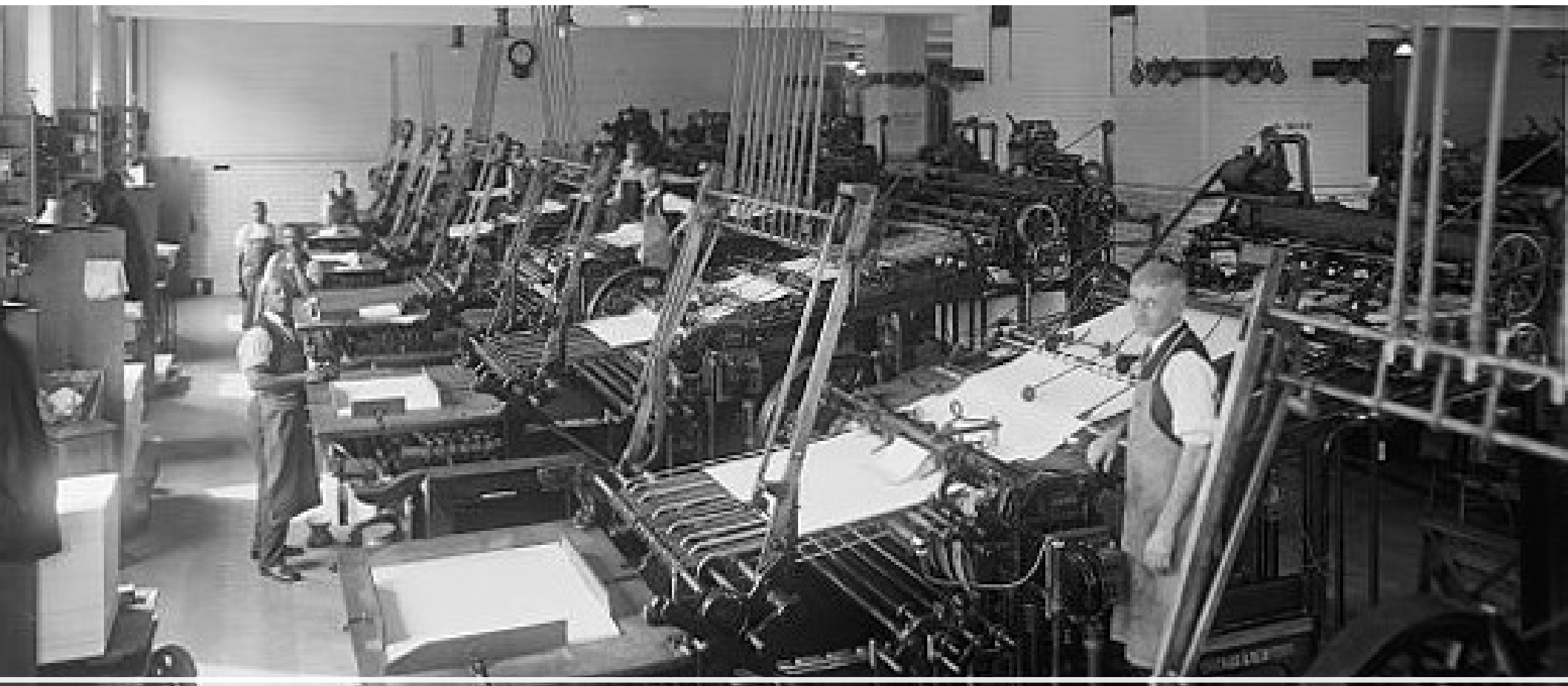
Meeting will start at ~2:02 pm and end at 4:00 pm EST

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

that work in a cooperative manner to coordinate and accomplish openness and transparency goals.



PREPARE



MANAGE



DISTRIBUTE



ARCHIVE/PRESERVE

CONSTITUTION
JEFFERSON'S MANUAL
AND
RULES OF THE HOUSE OF
REPRESENTATIVES
OF THE UNITED STATES
SEVENTY-SECOND CONGRESS

By
LEWIS DESCHLER
PARLIAMENTARIAN



UNITED STATES
GOVERNMENT PRINTING OFFICE
WASHINGTON : 1931

Bound by Rules, Traditions, and Practices

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”.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. JAMES M. & THOMAS W.L. ASHLEY CUSTOMS

BUILDING AND UNITED STATES COURT-

HOUSE

(a) DESIGNATION.—The Federal building and United States courthouse located at 1716 Spielbusch Avenue in Toledo, Ohio, shall be known as the “James M. & Thomas

2

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.

Large bills and the ecosystem

116 HR 133 Consolidated Appropriations FY21

- Engrossed Amendment in House (EAH): 5,432 pages
- Enrollment (ENR): 2,124 pages

116 HR 8352

- Introduced House (IH) : 4,435 pages

117 HR 2617 Consolidated Appropriations FY23

- Engrossed Amendment Senate (EAS2): 4,126 pages

117 HR 4521 COMPETES Act of 2022

- Engrossed House (EH): 3,610 pages

117 HR 7776 NDAA FY23

- Engrossed House (EH) – 3,470 pages
- Engrossed Amendment House (EAH) – 4,312 pages
- Enrolled (ENR) – 1,772 pages

118 HR 2670 NDAA FY24

- Engrossed Amendment Senate (EAS) – 2,330 pages

H. R. 133 ENR on its way to the Senate (for signature): 12/24/2020

Photo courtesy of the staff in Legislative Operations, Office of the Clerk





**Adding a digital layer to a
centuries-old paper process**

88TH CONGRESS
2d Session

H. R. 7152

IN THE HOUSE OF REPRESENTATIVES

FEBRUARY 10, 1964

Ordered printed as passed the House

AN ACT

To enforce the constitutional right to vote, to confer jurisdiction upon the district courts of the United States to provide injunctive relief against discrimination in public accommodations, to authorize the Attorney General to institute suits to protect constitutional rights in public facilities and public education, to extend the Commission on Civil Rights, to prevent discrimination in federally assisted programs, to establish a Commission on Equal Employment Opportunity, and for other purposes.

- 1 *Be it enacted by the Senate and House of Representa-*
- 2 *tives of the United States of America in Congress assembled,*
- 3 That this Act may be cited as "The Civil Rights Act of

.....

**Online means more
than just the
electronic
representation of
the paper document**

The Solution

Structured Data

Formats: XML, JSON, HTML...

GPO Developer Hub at GovInfo.gov/developers

The three characteristics that documents share:

- **Presentation:** how the content looks
- **Structure:** how the content is organized
- **Semantics:** what does it mean



Roadmap for Publishing Legislative Documents in USLM XML Format¹

| | |
|---------------------------------|--|
| Completed | U.S. Code |
| Completed | Enrolled bills, public laws, and the Statutes at Large |
| Completed | HOLC/SOLC statute compilations |
| IN PROGRESS <i>Initiated</i> | Remaining bill/resolution versions (introduced, reported, engrossed, all amendments, et. al.) |
| C | Committee Reports |
| D | House portions of the <i>Congressional Record</i> , including the <i>Daily Digest</i> |
| E | Hearing record |
| F | House Calendar |
| G | <i>Precedents of the U.S. House of Representatives</i> |
| H | <i>House Journal</i> |

The Solution: XML USLM schema

Legislative Branch Innovation Hub

[About the Hub](#)[Leadership Support](#)[Data Standardization](#)[Reports](#)[Events](#)[Resources](#)

Reports

Standardized Formats for Legislative Documents, USLM

- **Initial Report** - Required by Section 501 of House Resolution 756 of the 116th Congress - [PDF](#)
- Quarterly Report 1 - October 15, 2020 - [PDF](#)
- Quarterly Report 2 - January 15, 2021 - [PDF](#)
- Quarterly Report 3 - April 15, 2021 - [PDF](#)
- Quarterly Report 4 - July 15, 2021 - [PDF](#)
- Quarterly Report 5 - October 15, 2021 - [PDF](#)
- Quarterly Report 6 - January 15, 2022 - [PDF](#)
- Quarterly Report 7 - April 15, 2022 - [PDF](#)

[//usgpo.github.io/innovation/reports/](https://usgpo.github.io/innovation/reports/)

Congressional Data Task Force

ACTIVITIES SINCE THE JUNE MEETING

Next on the Agenda

The Government Publishing Office

- on the implementation of the provisions of Public Law 117-263 relating to the Access to Congressionally Mandated Reports Act

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

Panel on Artificial Intelligence

2023 Words of the Year

AI
HALLUCINATE
AUTHENTIC

AI: a term that describes **‘the modelling of human mental functions by computer programs’** is Collin’s dictionary word of the year

HALLUCINATE: the Cambridge Dictionary’s word of the year and describes **“when an artificial intelligence hallucinates, it produces false information”**

AUTHENTIC: Merriam-Webster’s Dictionary says, “*Authentic* has several meanings including **“not false or imitation,” a synonym of *real* and *actual*; and “true to one’s own personality, spirit, or character.”** A high-volume lookup most years, authentic saw a substantial increase in 2023, driven by stories and conversations about AI, celebrity culture, identity, and social media.

Sources:

<https://www.collinsdictionary.com/woty>

<https://dictionary.cambridge.org/us/editorial/woty>

<https://www.merriam-webster.com/wordplay/word-of-the-year>

FIRST QUESTION (current/past activities)

To start, I would like invite each of you to take two minutes to share what your organizations are currently doing (or have done) related to emerging technologies and artificial intelligence.

ORDER: GPO, LOC, House, CBO

SECOND QUESTION (forward thinking)

Given the increasing presence of AI in various sectors, technologists supporting the U.S Congress, as well as our colleagues in parliaments around the world, are witnessing the gradual integration of these technologies in the workplace. As technologists who support Congress, what are the **most significant opportunities that AI presents for Congressional work**, the work we do supporting the Institution, and for the legislative process?

- ORDER: LOC, GPO, CBO, HOUSE

THIRD QUESTION (data and cross-organizational considerations)

We are here because of data. It is well known that data is the foundation for any AI system, particularly when large language models are used. Governance and risk management of such data are critical in the legislative and lawmaking environment. In your roles, how do you approach the challenge of ensuring **transparency** and **authenticity** in the use of AI and other emerging technologies in Congress? What strategies could be implemented to maintain and enhance **public trust**? As the Congressional Data Task Force, what cross-organizational considerations such as **data governance**, **shared technologies**, and **upskilling**, do we need to consider as we do our work?

ORDER: House, GPO, LOC, CBO

CLOSING CONSIDERATIONS/REMARKS

- ORDER: SENATE, CBO, LOC, GPO, HOUSE

Next on the Agenda

Cross-organizational efforts and working groups
(ad-hoc, formal, and informal)

- the Legislative Branch XML Technical Working Group
- the working group on a Legislative Branch-wide Online Staff Directory
- the Congressional Video Preservation and Access Technical Working Group
- *the Legislative Branch Data Interchange Working Group*
- *the working group on Digitization Of Congressional Documents*

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

- The Senate (Secretary's Office)
- The House (Clerk's office and CAO's House Digital Services)
- The Congressional Budget Office
- Library of Congress

CBO's Use of GitHub

December 19, 2023

Presentation to the
Congressional Data Task Force

Kevin Perese
Senior Adviser, Data Science and Transparency

**CBO prioritizes
objectivity,
analytical soundness,
responsiveness,
transparency, and
accessibility.**

CBO's Transparency Efforts

- Testifying and publishing answers to questions
- Explaining analytic methods
- **Releasing data and computer code**
- Analyzing the accuracy of its estimates
- Comparing current estimates with previous estimates
- Comparing its estimates with those of other organizations
- Estimating the effects of policy alternatives
- Characterizing the uncertainty of estimates
- Creating data visualizations

Guidance in CBO's Administrative Manual on Posting Code

Section 7.17. Posting CBO Code

To promote transparency, Division Directors are encouraged to post data, computer code, and documentation associated with models and projects on CBO's GitHub page (<https://github.com/us-cbo>) to the **extent practicable**. When a Division Director has determined that there is value in releasing code, they should contact the Research Director for a code review to test for replicability of results and to ensure that no sensitive information is released. To maximize the quality of the underlying analyses that support CBO's written work and to help the release process go smoothly, employees should strive to follow **CBO's Coding Conventions** and should refer to **CBO's Code Review Checklist** before submitting their code for review.




Congressional Budget Office · G x +

← → ↻ 🏠


github.com/us-cbo

🖨️ 🔗 ☆ 🗄️ K ⋮

 Product ▾ Solutions ▾ Open Source ▾ Pricing

🔍 Search or jump to... /

Sign in Sign up



Congressional Budget Office

Nonpartisan Analysis for the U.S. Congress

👤 18 followers 🔗 https://www.cbo.gov ✉ communications@cbo.gov

🏠 Overview

📁 Repositories 4

📦 Packages

👤 People

Popular repositories

means_tested_transfer_imputations

Public

● Stata ☆ 11 🍴 4

debtwelfare

Public

● Python ☆ 2

captax

Public

● Python ☆ 7 🍴 6

eval-projections

Public

● Python ☆ 1

People

This organization has no public members. You must be a member to see who's a part of this organization.

Top languages

● Python ● Stata

eval-projections

- **Project:** Evaluating CBO's projections of components of the federal budget
- **Summary:** Code that replicates CBO's analysis of its projection errors, which can be found in the following four publications:
 - [*The Accuracy of CBO's Budget Projections for Fiscal Year 2023*](#)
 - [*An Evaluation of CBO's Projections of Outlays From 1984 to 2021*](#)
 - [*An Evaluation of CBO's Past Revenue Projections*](#)
 - [*An Evaluation of CBO's Past Deficit and Debt Projections*](#)
- **Programming language:** Python
- **Data:** CBO's baseline projections going back to the mid-1980s
- **Updates:** When CBO releases a new baseline or evaluation report

captax

- **Project:** CBO's CapTax model
- **Summary:** A large-scale model that produces CBO's estimates of effective marginal tax rates on capital from new investments; those estimated tax rates are included in CBO's data on [tax parameters and effective marginal tax rates](#) in each baseline
- **Programming language:** Python
- **Data:** Extensive collection of comma-separated values (*.csv) files containing model parameters
- **Updates:** When new baseline data become available

means_tested_transfers_imputations

- **Project:** CBO's adjustments for survey underreporting of receipt and income from selected means-tested transfer programs
- **Summary:** Code to replicate CBO's adjustments that correct for the undercount of receipt of selected means-tested transfer benefits in the Annual Social and Economic Supplement (ASEC) of the Census Bureau's Current Population Survey; those adjustments are used in the analyses underlying CBO's [*Distribution of Household Income*](#) reports
- **Programming language:** Stata
- **Data:** ASEC extracts from 1980 to 2022
- **Updates:** When new ASEC data files become available

debtwelfare

- **Project:** CBO's December 2022 working paper [*The Welfare Effects of Debt: Crowding Out and Risk Shifting*](#)
- **Summary:** An overlapping generations (OLG) model with simulations that include the decomposition of the effects of government debt on welfare into the crowding out of capital and the shifting of risk from current to future generations
- **Programming language:** Python
- **Data:** Model parameters and calibration targets
- **Updates:** No regular updates currently planned

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

- **Time for Discussion/Questions and Answers**
- **Announcements/Closing**

Meeting ends at 4:00 pm EST

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.

Legislative Branch Innovation Hub

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

Future Events

Tentative meeting dates – Congressional Data Task Force

March 19, 2024

June 6, 2024

December 12, 2024

Employment Opportunities usajobs.gov

- House: <https://www.house.gov/employment/positions-with-other-house-organizations>
- Senate: <https://www.senate.gov/visiting/employment.htm>
- GPO: <https://www.usajobs.gov/Search/?a=LP00>
- LOC: <https://loc.gov/careers/>

Thank you

Kirsten Gullickson

Office of the Clerk
U.S. House of Representatives

Kirsten.Gullickson@mail.house.gov

For more information:
USGPO.GITHUB.IO/INNOVATION/



Thank you for attending!

Congressional Data Task Force – Public Meeting
Tuesday, December 19, 2023