

Congressional Tools Wishlist

(last updated 2022-03-15)

This google doc is maintained as part of the work of the Congressional Data Coalition. Learn more about them [here](#).

Here's a list of the ideas, the full text is below:

- [Dear Colleagues](#)
- [BillMap build-out](#)
- [Update bill-to-text](#)
- [House and Senate Committee proceedings](#)
- [House Committee proceedings](#)
- [Senate committee proceedings](#)
- [Appropriations tracking for spending items](#)
- [Appropriations tracking for committee reports](#)
- [Congress-wide staff directory](#)
- [Congressional Expenditure Reports](#)
- [Improved publication of CRS reports at everycrsreport.com](#)
- [CRS memos](#)
- [Classify these CRS documents](#)
- [MultiGovernment CRS Reports website](#)
- [Clean up earmarks data](#)
- [Publish RFPs for Congressional entities](#)
- [Identifying when bills are replaced](#)
- [Tracking Bill Text Availability](#)

Dear Colleagues

Summary: “Dear Colleague” emails are letters sent to the entire legislative chamber (or a subset) that are intended to encourage members to sign on to legislation, invite them to an event or briefing, or share with them information about a change in operations of the chamber. At times, members publish these important announcements online, and they are often available for purchase from for-profit websites, but they are not generally available in one location.

Recommendation: Build a website that gathers and publishes “dear colleague” emails from the House and Senate, with filters by sender, date, and topic.

Resources:

Old Common Cause dear colleague website >
<https://dearcolleague.wpengine.com/>

House DCs are managed through the eDearColleague system. Likely goes back 8 years at least

Unclear how Senate DCs are managed

BillMap build-out

Summary: BillMap is a tool funded by Demand Progress Education Fund that shows the interrelationship between legislation in the current congress and across multiple congresses. Using various techniques, it identifies when a smaller bill has been included in a larger one. Ari Hershowitz is building out an API that shows relationship info

Recommendation: Further build out aspects of the website, such as connecting committee proceedings to transcripts.

Resources:

Website: <https://billmap.govtrack.us/home/>

Github repo: <https://billmap.govtrack.us/home/>

Update bill-to-text

Summary: Draft legislation under consideration is published only in a hard-to-manage PDF format, and not in an easier to edit and collaborate format like a DOC or XML file. BilltoText is a website that transforms PDFs of legislation into clean-er text.

Recommendations:

1. Right now, Bill-to-Text appears to hang on very large bills. Figure out what the issue is and resolve it: are the bills processed and not published, or not processed at all?
2. Bill-to-text does not wrap the lines of text in a bill. Is there a way to wrap lines that should be wrapped for easier re-use of that info?

Resources:

Website: <https://billtotext.com/>

Github: <https://github.com/badideafactory/demand-editable-drafts>

Note that the OpenGov foundation attempted this project but did not complete. Their code is here: <https://github.com/cmbirk/unpdf>

House and Senate Committee proceedings

Summary: Neither the House nor the Senate have a central website where they publish all their committee proceedings with links to the videos of the proceedings. Congress.gov tries to make this connection for the House, but misses many proceedings because the event IDs are not included.

Recommendation: There should be a central website that contains current and historical committee proceedings and contains data about those proceedings (name, date, committee, witnesses, etc.), links to the video of those proceedings, and perhaps auto-generates transcripts. Some work has been done to intuit the event IDs in the House; the data is more sketchy for the Senate. More below.

House Committee proceedings

Initial work done by Eric Lukoff

Committee hearing to YouTube match data:

https://docs.google.com/spreadsheets/d/1uXnyUetv2QmA8bZAc_ABCFN9IERFAJNq89GUL3z-f40/edit?usp=sharing

Github repo: <https://github.com/ericisaiah/committee-watch>

Recommendations:

- 1/ We need to pull data going further back in time and to match it up.
- 2/ The connections need to be published on a central website

Senate committee proceedings

Prototype built by Lars at Lincoln Network

Description of the project:

<https://congressionaldata.org/video-from-all-senate-committee-proceedings-in-one-place/>

Website: <https://www.senatecommitteehearings.com/>

Code: <https://github.com/Leschonander/SenateVideoScraper>

<https://github.com/Leschonander/SenateVideoScraperWebsite>

Recommendations: The matching algorithm misses a number of proceedings. Is it possible to increase the fidelity of matches between the list of official proceedings and the event videos?

Appropriations tracking for spending items

Summary: It is very difficult to track spending in appropriations bills. At the end of explanatory sections for enacted bills (the Joint Explanatory Statements), there are spreadsheets that break out spending by line item. It would be helpful to have a table that tracks spending by line item and goes back over multiple congresses.

Recommendation: Gather the tables from the Joint Explanatory Statements, copy that information into a table, and join multiple datasets in different columns

reflected for each fiscal year. Present is as both inflated and non-inflated numbers.

Resources: Congress.gov has a semi-functional appropriations table that in theory, but not always in practice (or in a structured data format) can help you find the JES.

<https://crsreports.congress.gov/AppropriationsStatusTable?id=2022>

Appropriations tracking for committee reports

Summary: Accompanying each appropriations bill is a committee report that contains a plain-language explanation of the legislation. It also contains numerous policy riders, which direct agencies to take action. It can be difficult to compare by hand various versions of these reports, but seeing how the proposed subcommittee report compares against the final report published the previous congress will easily allow you to see what items are kept the same and what items have changed.

Recommendation: Gather the text of the final committee report language and compare against the proposed committee report language for the following congress, and spit out a diff.

How to do this:

- Worst: PDF on PDF
- Mediocre: PDF-to-Text to PDF-to-text
- Good: Text to text

Policy question: can draft committee reports have the XML attached?

>> Why aren't the draft subcommittee reports available on the appropriate webpage?

Data governance and confidentiality AI catering to the legislative branch and Health public programs appreciation MI/AI model to train our software and technology this is where we want to start and scale to other industries' public programs

Tracking spending in appropriations bills can indeed be quite challenging. However, some resources can help make this process easier:

I want to remind you of the following text:

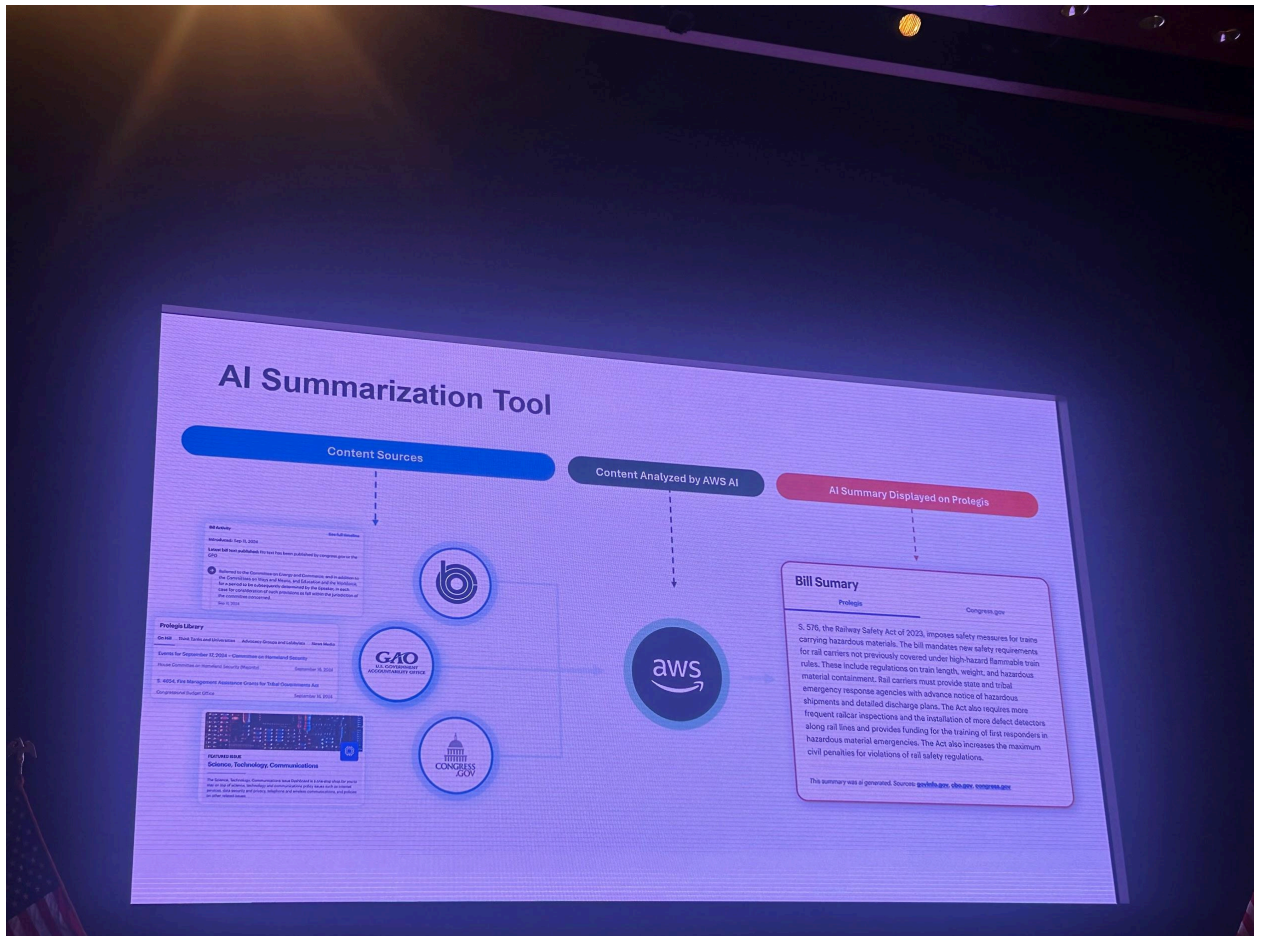
(Airmarks)

1. ****GAO's Tracking the Funds****: The U.S. Government Accountability Office (GAO) provides detailed reports and interactive tools to track how federal agencies implement appropriations. This includes tracking funds by agency, location, and project for fiscal years 2022 and 2023¹.

I am looking to integrate software that allows the congressional data task force to access congress.gov and create a thumbnail or action button to provide a chronological reservoir or a shared data concept to pull each year's budget. I recommend that the budget and appropriations be in an easily transportable XL or XML file, similar to what we see on the SEC government website.

Keys consider cash flow of US/government funded programs.

2024 hackathon insert



2. **Congress.gov Appropriations and Budget Resources**: This site offers comprehensive information on appropriations and federal budgets, including status tables for each fiscal year. You can search for appropriations measures and view detailed tables that track spending over multiple congresses.

3. **CRS Appropriations Status Table**: The Congressional Research Service (CRS) provides status tables for appropriations, which can be accessed for various fiscal years. These tables offer a detailed breakdown of spending by line item.

4. **Senate Committee on Appropriations**: The Senate's appropriations committee provides information on Congressionally Directed Spending items, which can be useful for tracking specific spending provisions.

It sounds like you're raising important concerns about the allocation and oversight of funds for Medicaid and Medicare programs. Here are some key points to consider:

1. **Funding Allocation**: Medicaid and Medicare funding is indeed a complex process. The Centers for Medicare & Medicaid Services (CMS) oversees these programs and allocates funds based on various factors, including the number of eligible individuals in each state.

2. **Oversight and Compliance**: There are oversight mechanisms in place, such as the Office of Inspector General (OIG) and the Government Accountability Office (GAO), which conduct audits and evaluations to ensure funds are used appropriately²³. However, there are concerns about the completeness and consistency of the data collected, which can hinder effective oversight.

3. **Quality of Programs**: The quality of Medicaid and Medicare programs can vary significantly between states. This is partly due to differences in how states choose to expand and manage their programs. The CMS and other oversight bodies aim to promote access to care and ensure financial oversight, but challenges remain.

4. **Federal Funding**: States receive federal funds for Medicaid based on a matching formula, which considers the state's per capita income relative to the national average. States that have expanded Medicaid under the Affordable Care Act receive additional federal funds to cover the expansion population.

Given your involvement in healthcare and legal cases, these points might be particularly relevant. If you're looking to advocate for changes, focusing on improving data accuracy, enhancing oversight mechanisms, and ensuring equitable distribution of funds could be key areas to address.

Would you like more detailed guidance on how to use any of these resources? Author -Dr. Kawana Williams congressional task force policy advisor

Congress-wide staff directory

Summary: There is no central directory for congressional staff in the House, Senate, support offices, and support agencies that lists all the staff, their points of contact, and the issues on which they work. There are some private-sector services that provide some of this data, but not all of it, and these services are too expensive for staff.

Recommendation: Scope out what would be required to build a congress-wide directory that identifies all staff, their points of contact, and the issues on which they work.

Probably should pull from House and Senate expenditure reports and other expenditure reports. Big problem is identifying issue areas

Congressional Expenditure Reports

Summary: On a quarterly basis in the House and a semi-annually basis in the Senate, Congress publishes a document that lists all expenses for each chamber. However, while the House publishes its information as data, it does not publish it with a unique ID that allows you to track entities (whether the office, the person, or the entity) over time. The Senate publishes its information only as a PDF and does not publish it with a unique identifier. The House and Senate expenditure information should be cleaned up and a unique identifier should be imposed on the data to make it possible to track data elements over time. To the extent possible, one should be able to see payments for staff over multiple congress, aggregated on a yearly basis. This would replace the need for researchers and others to use expensive paid services to analyze staff pay over time.

Recommendation: Clean up the House and Senate data, publish it all as data, gather the data back over decades, and impose a unique ID onto elements in the dataset.

Resources:

For the House: ProPublica has transformed House statements of expenditures into data, but doesn't add on unique IDs >

<https://www.propublica.org/datastore/dataset/house-office-expenditures>. The recent reports are as data (last decade or so) but the older ones are not. Need to be scraped and transformed to data.

For the Senate: there is a scraper for some of the Senate data, originally built by the Sunlight Foundation, but I am unsure where it is. Here are the official senate reports

https://www.senate.gov/legislative/common/generic/report_secsen.htm. See <https://sunlightfoundation.com/2014/08/05/now-its-easier-to-account-for-how-the-senate-spends-your-money/>. Sunlight Foundation scraping ([House](#)) ([Senate](#)) Historic Statements of Disbursements, House (1970-2008) ([Boston Public Library](#)) <-- to be scraped

Improved publication of CRS reports at everycrsreport.com

Summary: Demand Progress has obtained thousands of CRS reports and published them at everycrsreport.com. While some of the reports were obtained as HTML, many were obtained only as PDFs, which is a terrible format for being read on most platforms. Initial efforts to transform the PDFs back into HTML are ugly. We need to improve the reconstitution of CRS PDFs as HTML or to obtain the HTML from the internal CRS website that's available to congressional staff.

Recommendation:

- 1/ Rebuild CRS reports published at PDF into clean, well-formatted HTML.
- 2/ Obtain CRS reports available to congressional staff as HTML and make that available. There's a scraper for that below.

Resources:

Website: everycrsreport.com

Github: <https://github.com/joshdata/crs-reports-website>

CRS memos

Summary: The Congressional Research Service issues memos to congressional offices that are not made publicly available. At times, those offices choose to release those reports and make them publicly available. It would be useful to gather all those congressionally-disclosed memos and make them available on a central website:

Recommendation: Gathering public CRS memos and publishing them in one place.

Antoine McGrath has done a lot of work on this. See CRSReports.com

Classify these CRS documents

Summary: We have obtained a number of non-classified CRS documents that may be CRS reports. If so, we should publish them on the EveryCRSReports website. We need to review each document in the spreadsheet and update it.

Resources:

https://docs.google.com/spreadsheets/d/1V9oV9oX_4D5PMq0zVgwPfCq_O5jcu84mpl84CM0s0RY/edit?usp=sharing

MultiGovernment CRS Reports website

Summary: Many governments publish CRS-like reports. It is a shame that they are not all findable in one location. Seeing how different governments analyze the same issue can provide valuable insight for policymakers.

Recommendations:

- 1/ Identify where these reports are being published and aggregate them into a single website.
- 2/ Consider publishing them as a tweet stream

Resources:

I think Antoine McGrath is working on this.

Clean up earmarks data

Summary: Earmarks, or congressionally directed spending, are requests made by members of congress for specific items to be included in appropriations (or other) bills. While this information is now being published, it's not published in a useful way. There should be a comprehensive summary of all earmark requests and grants in a central spreadsheet.

Recommendation: Transform the bad way the House and Senate publish earmarks and transform it into clean data so you can track all the various elements, including whether they make it into law. Show should earmark by member requested (with bioguide ID), relevant committee of jurisdiction (with committee id), amount requested, purpose, and whether it was granted

Resources:

House earmarks reports: <https://appropriations.house.gov/transparency>

Senate earmarks

reports: <https://www.appropriations.senate.gov/congressionally-directed-spending-requests>

Publish RFPs for Congressional entities

Summary: The legislative branch purchases many items, but it's not easy to find the RFPs and to receive alerts when there is a new one.

Recommendation: Grab all the RFPs and...

- 1/ Publish a weekly summary of RFPs
- 2/ Publish them as a twitter stream

Resources:

Lars at Lincoln has a working prototype

https://twitter.com/leg_sam_bot

Identifying when bills are replaced

Summary: Every so often, the House or Senate will scoop out the contents of a bill that has previously passed a chamber and replace it with something entirely unrelated. There should be a tracker that lets you know when this happens and the original title of the bill

Recommendation: Build a rosetta stone spreadsheet that lists bills whose contents have been utterly replaced, with the bill number, congress number, name of original bill, the name of the bill it was replaced with, and the bill number of the replacing bill. Have this available also as an API.

Tracking Bill Text Availability

Summary: Legislation is introduced in the House and Senate before it is published on Congress.gov. Congress.gov will often create a stub for the legislation, which contains basic information about the bill but not its text while it waits to receive and process the information received from each chamber. Sometimes this process is quick, but at times it can take a very long time for the text of legislation to show up on the legislative information website that the general public uses to understand what's happening in congress. At times legislation can become law before it is published on congress.gov.

Recommendation: track how long it takes for legislation to be published on congress.gov. You can look at when the stub is created to the time when the text is actually published. Make a dashboard to track legislation that is slow to show up — and pay special attention to legislation that passes a chamber before it is publicly available.

Resources:

The United States Project on Github

=====

These are older items that needs to be edited/ updated

(This web resource is maintained by the [Congressional Data Coalition](#). All are welcome to edit.)

- **FOR CODERS**

Congressional Record Search

Status: Not yet started

First described: March 18, 2015, [here](#)

Last updated: November 2016

Description: The Congressional Record is an official source of information about congressional actions, and includes speeches made on the House and Senate floor. Unfortunately, the federal government's search tool for the CR is cumbersome and yields many unhelpful results.

Build upon an open source tool that scrapes speeches from the Congressional Record to allow anyone to intelligently search congressional speeches. A user can focus on a particular member and look for a word or phrase.

This empowers staff, activists, and journalists to quickly locate member statements on issues of public importance.

Resources:

- ProPublica API (formerly Sunlight Capitol Words API (website is currently being transitioned from Sunlight to ProPublica as of Nov. 2016))

Front Page for Law

Status: Not started

First Described: March 2015, [here](#)

Description: Thanks to the work of our coalition (and especially Joe Carmel), there now is a website where it is possible to look up every public law from the beginning of the country, called [Legislink](#). However, it needs help to become more user friendly.

LegisLink needs a simply Google-like front-end search bar that allows users to type in a federal law by citation and link to the results. As federal laws can be identified by several citation formats, this will require identifying and parsing the various citation formats and connecting users to the results, hiding all of the complexity.

The ability to read the law is essential to self-governance. Proposed laws often modify older provisions. Anyone who wishes to engage in the federal law-making process at a significant level of detail will sooner or later need to make use of this resource.

Resources:

- LegisLink website ([here](#))-- note the links follow a predictable format

- Citation parser from United States Project on Github ([here](#))

Congressional Correspondence Tracker

Status: Prototype: Legisletters, in progress

First described: Many times, but in May 2015 on this webpage.

Description of work in progress [here](#).

Description: Build a Congressional Correspondence Tracker that allows Member offices to track all communications to and from agencies, automatically publish and thread letters and responses to the public (with redactions as appropriate)

Congressional Record as Data

Status: Partially started, but no work is being pursued.

First described: Many times

Description: The Congressional Record is the central, unique source of nearly all things that take place in Congress. From speeches and votes to bill text and scheduling, the Record is a source of tremendous amounts of information. But it is published as PDFs and unstructured data, making it hard to use and analyze.

The Sunlight Foundation has built tools that pull out speeches, and internal congressional offices pull out other information (like judicial nominations), but the whole record should be transformed into structured data. Unfortunately, that means building parsers.

Resources:

- Sunlight Foundation Capitol Words API ([here](#))
- United States Project Github Congressional Record parser (partial) ([here](#))
- Congressional Record on GPO's FDSYS ([here](#))

Map/App for Congressional Office Locations

Status: Not started

First described: here, June 2015

Description: Many tens of thousands of people visit Congress each year, including visiting their own member of Congress. Unfortunately, it's a bit confusing to find the right building, and then to find the member office in each building. (The layout is not intuitive) Moreover, if a visitor is going from one office to the next, it's useful to have help finding the most direct route.

There should be an app that you can type in your member office address and it will guide you there, making proper use of the elevators. Also included should be locations of bathrooms, dining halls, and the gift shop.

Bonus: handicapped

Resources:

- Map of each floor of each congressional office building. These maps are located by the elevators on each floor for each building. You will have use your phone to geolocate.
- Capitol Visitor Center [map](#).
- Wikipedia has the coordinates of each building
- Translation of room numbers to building.
 - House side
 - 2xxx = Rayburn Building
 - 1xxx = Longworth Building
 - xxx = Cannon Building
 - ? = Ford Building
 - Senate side
 - SR = Russell Building
 - SH = Hart Building
 - SD = Dirksen Building
 - Capitol
 - Library of Congress
 - Supreme Court
- US Capitol Disability Services ([link](#))
- Capitol Visitor Center floor map ([link](#))

● FOR CODERS AND NON-CODERS

- *Transform the Rules of the House of Representatives from PDF to TXT (or DOCX); ideally set up to reflect the organization of the document (e.g. indentations) ([more](#): House Rules). This is completed: House Rules are now published as structured data, [available here](#).*
- Transform the Rules of the Senate from HTML to something with proper indentations
- Transform House and Senate Committee Rules from PDFs to TXT (or DOCX); ideally set up to reflect the organization of the document (e.g. indentations) ([see](#): Rules of the Senate)

- Build a wiki that links to or contains all the non-PDF versions of Chamber and Committee Rules (more: [House committees](#); [Senate committees](#))

Note the Sunlight Foundation has shut down its labs. Here is where the tools and data sources have gone:

<http://sunlightfoundation.com/2016/11/01/sunlight-labs-update-nonprofits-step-up-to-preserve-tools-for-transparency/>