

## Citation

Use of LobbyView data for academic purposes (presentations, working and conference papers, theses and dissertations, journal articles, software packages) requires that you cite the following paper:

Kim, In Song (2018). "LobbyView: Firm-level Lobbying & Congressional Bills Database." Working paper available from <http://web.mit.edu/insong/www/pdf/lobbyview.pdf>

## Issue Text Data Export

This CSV export contains a complete copy of all report issue text data. A row is a single issue from a single report. Reports are filed by registrants (lobbying firms), who employ lobbyists and are hired by clients (firms). We provide a series of identifiers designed to make it simple to connect clients to firm-level data and to other reports, and registrants to lobbyists and to other reports. The issue text data is supplied by registrants, and is the main mechanism through which we can connect lobbying efforts to specific congressional bills.

The below table specifies the columns provided in this CSV export. The fourth column, "Linked Datasets", describes how a column can be used to link this data to external data, including our other CSV files.

### Issue Text Level Variables

Column Name	Description	Data Type	Linked Datasets
report_uuid	Unique identifier for each lobbying filing (report), officially provided by the U.S. Senate LDA Office. Can be used to link to any other metadata about the report and through the report to other data about the client, registrant, or lobbyist.	TEXT	Report Level Data, Issue Level Data
general_issue_code	3-character code used to categorize the issue area covered by the lobbying activity. See below for more information.	TEXT	—
issue_ordinal_position	Ordinal number distinguishing the order the issue appeared in the original report (sequential, starting from 1). A row in this dataset is a single report-issue, but reports can contain many issues.	INT	—
bill_id_agg	Array of bill_ids mentioned in issue_text, formatted in the same manner as bill_id in our bill level dataset. Examples include: "hres967-93", "hjres1060-93", "sres319-93". <b>NOTE:</b> bill_id is not supplied in a structured format in LDA reports. We use automated processes to extract them. While overall accuracy is high, we recommend researchers check issue_text and original filing documents to ensure accuracy. See discussion below for how to convert ARRAY data to individual observations in R.	ARRAY	Bill Level Data

Column Name	Description	Data Type	Linked Datasets
issue_text	Text provided by the registrant in their LDA filing. Text is provided in an unstructured format, and has varying levels of detail depending on the registrant. Text may describe public policy matters in general (e.g. 'wildlife protection') or in specific, and may clarify the registrant's position on the issue or may not. This field is the source for the bill_id_agg data.	TEXT	–

## Issue Codes

Here we replicate the official Lobbying Disclosure Act list of issue codes, originally available from the [U.S. Congress](#)

Issue Code	Issue Name	Issue Code	Issue Name
<i>ACC</i>	Accounting	<i>HOM</i>	Homeland Security
<i>ADV</i>	Advertising	<i>HOU</i>	Housing
<i>AER</i>	Aerospace	<i>IMM</i>	Immigration
<i>AGR</i>	Agriculture	<i>IND</i>	Indian / Native American Affairs
<i>ALC</i>	Alcohol & Drug Abuse	<i>INS</i>	Insurance
<i>ANI</i>	Animals	<i>LBR</i>	Labor Issues / Antitrust / Workplace
<i>APP</i>	Apparel / Clothing Industry / Textiles	<i>INT</i>	Intelligence and Surveillance
<i>ART</i>	Arts / Entertainment	<i>LAW</i>	Law Enforcement / Crime / Criminal Justice
<i>AUT</i>	Automotive Industry	<i>MAN</i>	Manufacturing
<i>AVI</i>	Aviation / Aircraft / Airlines	<i>MAR</i>	Marine / Maritime / Boating / Fisheries
<i>BAN</i>	Banking	<i>MED</i>	Medical / Disease Research / Clinical Labs
<i>BNK</i>	Bankruptcy	<i>MIA</i>	Media (Information / Publishing)
<i>BEV</i>	Beverage Industry	<i>MMM</i>	Medicare / Medicaid
<i>BUD</i>	Budget / Appropriations	<i>MON</i>	Minting / Money / Gold Standard
<i>CAW</i>	Clean Air & Water (Quality)	<i>NAT</i>	Natural Resources
<i>CDT</i>	Commodities (Big Ticket)	<i>PHA</i>	Pharmacy
<i>CHM</i>	Chemicals / Chemical Industry	<i>POS</i>	Postal
<i>CIV</i>	Civil Rights / Civil Liberties	<i>RRR</i>	Railroads
<i>COM</i>	Communications / Broadcasting / Radio / TV	<i>RES</i>	Real Estate / Land Use / Conservation
<i>CPI</i>	Computer Industry	<i>REL</i>	Religion
<i>CSP</i>	Consumer Issues / Safety / Protection	<i>RET</i>	Retirement
<i>CON</i>	Constitution	<i>ROD</i>	Roads / Highway
<i>CPT</i>	Copyright / Patent / Trademark	<i>SCI</i>	Science / Technology
<i>DEF</i>	Defense	<i>SMB</i>	Small Business
<i>DOC</i>	District of Columbia	<i>SPO</i>	Sports / Athletics
<i>DIS</i>	Disaster Planning / Emergencies	<i>TAR</i>	Miscellaneous Tariff Bills
<i>ECN</i>	Economics / Economic Development	<i>TAX</i>	Taxation / Internal Revenue Code
<i>EDU</i>	Education	<i>TEC</i>	Telecommunications
<i>ENG</i>	Energy / Nuclear	<i>TOB</i>	Tobacco
<i>ENV</i>	Environmental / Superfund	<i>TOR</i>	Torts

Issue Code	Issue Name	Issue Code	Issue Name
<i>FAM</i>	Family Issues / Abortion / Adoption	<i>TRD</i>	Trade (Domestic & Foreign)
<i>FIR</i>	Firearms / Guns / Ammunition	<i>TRA</i>	Transportation
<i>FIN</i>	Financial Institutions / Investments / Securities	<i>TOU</i>	Travel / Tourism
<i>FOO</i>	Food Industry (Safety, Labeling, etc.)	<i>TRU</i>	Trucking / Shipping
<i>FOR</i>	Foreign Relations	<i>URB</i>	Urban Development / Municipalities
<i>FUE</i>	Fuel / Gas / Oil	<i>UNM</i>	Unemployment
<i>GAM</i>	Gaming / Gambling / Casino	<i>UTI</i>	Utilities
<i>GOV</i>	Government Issues	<i>VET</i>	Veterans
<i>HCR</i>	Health Issues	<i>WAS</i>	Waste (hazardous / solid / interstate / nuclear)
		<i>WEL</i>	Welfare

## Dealing with Array Data

In this section we describe a basic workflow to unpack ARRAY-encoded data into individual rows using the **R** programming language and the popular **tidyverse** family of software packages. We provide this example code strictly for user convenience, and do not offer support for data cleaning or manipulation tasks.

```
library(tidyverse)
```

```
# Read the data
```

```
long_data <- read_csv("issue_text.csv") |>
```

```
# Remove the braces that are at the beginning and end of the ARRAY and the quotation marks
```

```
mutate(bill_id_agg = str_replace_all(bill_id_agg, "[{}\\\"]", "")) |>
```

```
# Split the array by comma
```

```
separate_longer_delim(bill_id_agg, delim=",")
```

## Data and Version Notes

This data was generated on 2024-10-25. For detailed version history, please see the `version.pdf` file. Versions of the data beginning with v2024.10.01 contain a new algorithm for extracting bill matches (used to generate `bill_id_agg`) from raw filings. This has expanded the number of bills matched from existing issues and improved the quality of matching. We cannot guarantee that all bills deemed matched in previous versions of the data remain.

## Reporting Data Issues

Our team has invested an enormous amount of effort to ensure that the LobbyView data is of the highest possible quality. If you detect an error in our data, we would love to hear from you. Please contact [lobbydata@gmail.com](mailto:lobbydata@gmail.com) and include as much information as possible so we can understand the error you are reporting.

## Connecting LobbyView Data

The below diagram provides a quick guide explaining how to connect each LobbyView data download to each other, and to external datasets.

Connecting LobbyView Data

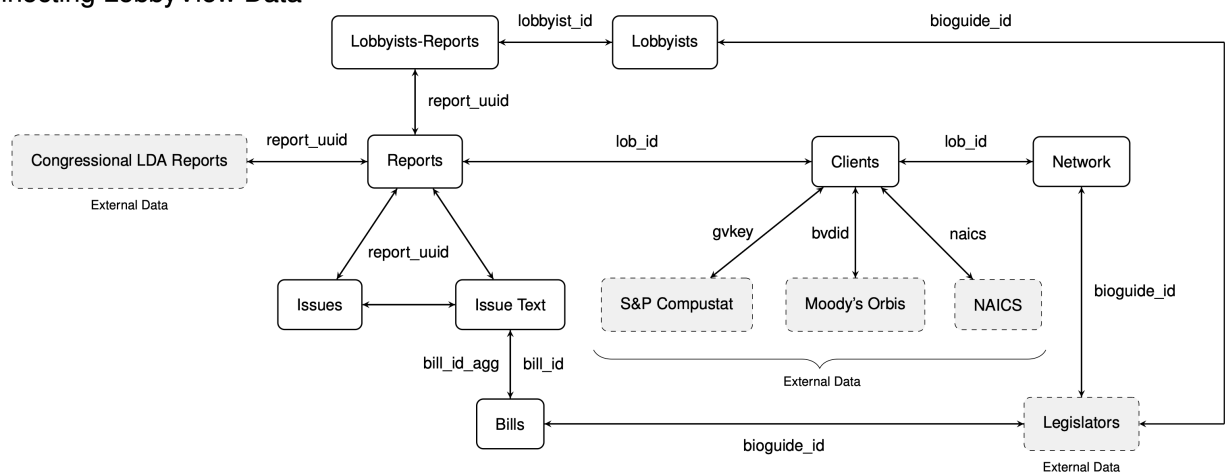


Figure 1: Connection Diagram