



BICAM Data Codebook

Citation

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

Committee Reports

We include 106,498 committee reports from the 16th to the 118th Congress. Committee reports are the summary of the investigation, findings, and disposition of the relevant committee towards legislation, and are believed to shape congressional voting. This data is linked to other main components and related sub-components via `report_id` (comprised of report type, report number, and Congress number). Sub-components include associated bills, associated treaties, and texts.

Each CSV file in this component is described below, one per page.



BICAM Data Codebook

Committee Reports

Row Count: 106,498

Variables

Column Name	Description	Data Type
report_id	Primary key, unique identifier capturing an individual Committee Report across all Committee Report files.	CHAR
report_type	Type of report published, which is tied to congressional chamber ("hrpt" for House Committee Reports, "srpt" for Senate Committee Reports, and "jrpt" for Joint Committee Reports).	CHAR
report_number	Ordinal number associated with Committee Report within Congress. This is the 2nd portion of report_id	NUMERIC
report_part	Ordinal number associated with a section of a multi-part Committee Report within Congress. This is the 3rd portion of report_id	NUMERIC
congress	The term of U.S. Congress for a record	NUMERIC
session	Session of a given Congress term for a record	NUMERIC
title	Text title associated with record	CHAR
subtitle	Secondary title for a record	CHAR
chamber	Chamber of Congress for a record. Most often "house" or "senate", but may also be "joint" depending on component type.	CHAR
citation	Internal citation for reference for a record	CHAR
is_conference_report	Boolean value for if the given Committee Report was a conference report (report to resolve differences between two chambers of Congress)	BOOLEAN
issued_at	Timestamp for when a record was officially released by Congress	DATE
pages	Integer value of how many pages consist of text from a given record	NUMERIC
texts_count	Number of text versions associated with a given record	NUMERIC
updated_at	Date associated with last update of document (from Congress, not to be confused with BICAM version dates)	DATE



BICAM Data Codebook

Committee Reports-Associated Bills

Row Count: 38,694

Variables

Column Name	Description	Data Type
report_id	Primary key, unique identifier capturing an individual Committee Report across all Committee Report files.	CHAR
bill_id	Primary key, unique identifier capturing an individual Bill across all bill files.	CHAR



Committee Reports-Associated Treaties

Row Count: 325

Variables

Column Name	Description	Data Type
report_id	Primary key, unique identifier capturing an individual Committee Report across all Committee Report files.	CHAR
treaty_id	Primary key, unique identifier capturing an individual Treaty across all treaty files.	CHAR



BICAM Data Codebook

Committee Reports-Members

Row Count: 18,620

Variables

Column Name	Description	Data Type
report_id	Primary key, unique identifier capturing an individual Committee Report across all Committee Report files.	CHAR
bioguide_id	Primary key, unique identifier capturing an individual Member across all member files. This comes from Congress's Bioguide Directory	CHAR



BICAM Data Codebook

Committee Reports-Metadata

Row Count: 34,952

Variables

Column Name	Description	Data Type
report_id	Primary key, unique identifier capturing an individual Committee Report across all Committee Report files.	CHAR
govinfo_package_id		CHAR
su_doc_class_number		CHAR
migrated_doc_id		CHAR
govinfo_collection_code		CHAR



BICAM Data Codebook

Committee Reports-Metadata Granules

Row Count: 19,013

Variables

Column Name	Description	Data Type
report_id	Primary key, unique identifier capturing an individual Committee Report across all Committee Report files.	CHAR
govinfo_package_id		CHAR
govinfo_granule_id		CHAR



BICAM Data Codebook

Committee Reports-Texts

Row Count: 83,637

Variables

Column Name	Description	Data Type
report_id	Primary key, unique identifier capturing an individual Committee Report across all Committee Report files.	CHAR
raw_text		CHAR
formatted_text		CHAR
formatted_text_is_errata		BOOLEAN
pdf		CHAR
pdf_is_errata		BOOLEAN