

Congressional Hearings Data Codebook

Policy Agendas Project
2012

Contact Information

Visit www.policyagendas.org or contact us at policyagendas@gmail.com.

Citation Guidelines

Please visit www.policyagendas.org/page/how-cite for guidelines on how to cite the use of this or other Policy Agendas Project datasets.

Congressional Hearings Data Codebook

An Overview of the Data Set

The Agendas Project Hearings Data Set tabulates information on all congressional hearings conducted between 1947 and 2005.¹ The data was tabulated from the hearings sections in the *Congressional Information Service /Annual: Abstracts of Congressional Publications and Legislative History Citations*. The hearings included are those of committees, subcommittees, task forces, panels and commissions, and the joint committees of Congress. Not all entries in the CIS hearings section are actually hearings, and they are not coded and inputted in this dataset. Committee reports, publications, supplementary materials, and declassified material are examples of entries in the source that are excluded in the database. **Note:** Research making use of the congressional hearings dataset should bear in mind that the hearings for the last year available on our website are incomplete. This is because there are commonly hearings for a given year that are not archived in the CIS for that year. The remainder of the hearings appears in the following CIS Volume. For example, the hearings for the year 1999 are not complete until we have collected and coded the CIS Volume for the year 2000. This means that the hearings for any given year are incomplete until we have collected the data for the CIS Volume that follows.

Content Coding

We have coded all hearings by the policy content of the proceedings, along with other critical variables. We employed 19 major topic and 225 minor topic codes in our policy content coding, and each hearing was assigned one and only one content code. This allows researchers to trace hearing activity on a particular topic across the period. This approach avoids the oversampling problem characteristic of word searches in data bases such as the CIS hearings files. The full description of our content categories may be found in the major topics codebook at <http://www.policyagendas.org/page/topic-codebook>.

Variable Names and Descriptions:

KeyID

This column records the unique identifier for each observation. It has no substantive application.

Source

The source number is a unique identifier used in the CIS Abstract Volume. It can be used via LexisNexis Congressional to reference full text.

CISYear

The CIS Abstract volumes are divided by year from 1970 to current. This column is used to indicate the year of the volume for the record listed.

Chamber

This column records in which chamber the hearing took place. The codes are:

- 1= House of Representatives
- 2= Senate
- 3= Joint Committees and Commissions

In the volumes for 1970 to present, the serial number to the hearing indicates this. Hearing H181-50 is a hearing that took place in the House. If a "J" precedes the serial number, it took place in a joint committee, and "S" indicates Senate.

CISCommittee

The column is used to record the three digits that immediately follow the letter of the serial number. For example, the number "181" to the serial number H181-50 would be recorded.

Sequence

This column is used to record the number that immediately follows the three digits designating committee. For example, in H181-50, the number to code is "50."

Month

The month column records the month on which the hearing started. Instead of writing in the month, a numeric code designating January as "1" and December as "12" is used. If a hearing lasted more than a month, the month on which the hearing began is recorded only.

Congress

This column denotes the Congress that held the hearing.

Year

The year column records the year on which the hearing started. This information is found following the title of the hearing.

FiscalYear

This column records the fiscal year in which the hearing was held.

Days

The column is used to keep record of the number of days mentioned in the abstract.

Sessions

This column is used to record the number of sessions within a hearing. This number can be found in the number of the subsection after the summary.

Committee1

The column records the committee that held the hearing. This code is different from the committee column. A coding scheme independent of the CIS Abstract is used and can be found in the committee codebook on the Policy Agendas Project website.

Subcommittee1

This column is used to more accurately describe the body that held the hearing. Hearings can either be held by a full committee or by any of its subcommittees. If no subcommittees belonging to the parent committee are mentioned, it is presumed that the hearing took place before a full committee.

Committee2

The committee 2 column is similar to the committee 1 column. This column is used if:

1. A second subcommittee within the committee was present
2. If a second subcommittee or an outside subcommittee was present.

In the event that there are two subcommittees of a committee "A" and one subcommittee of committee "B," the code for committee "B" is recorded in this column.

Subcommittee2

This column is used whenever more than one subcommittee holds a hearing. This can happen when more than one subcommittee of a committee meet or if two subcommittees from different committees meet.

Refer

This column records whether the hearing was a legislative or a non-legislative hearing.

- 1=legislative
- 0=non-legislative

Approp

This column is used to indicate if a hearing considered, authorized, made or reauthorized appropriations or budgets.

Agency

This column is used to record if the hearing subject matter involved the creation of a new government agency. The codes for this column are:

- 0=No mention of agency creation
- 1=Agency creation proposed

Program

This column records whether the hearing involved the proposal for a new program. This column is independent of the new agency column. Codes are:

- 0=No mention of new program
- 1=New program is discussed

Admin

This column is used to record hearings on issues or legislation proposed by the administration of the President. The appropriations, oversight or referral columns do not affect this column.

- 0=No administration proposal within Congress is mentioned
- 1=Administration action within Congress

MajorTopic

This column records the Policy Agendas Project's major topic code that corresponds to the information found in the Description column. The codes and examples for this column are found in the Major Topics Codebook on the Policy Agendas Project website.

SubTopicCode

This column denotes the Policy Agendas Project's subtopic code that corresponds to the information found in the Description column. The codes and examples for this column are found in the Major Topics Codebook on the Policy Agendas Project website.

Description

This column contains a short description of the topic entry/record.

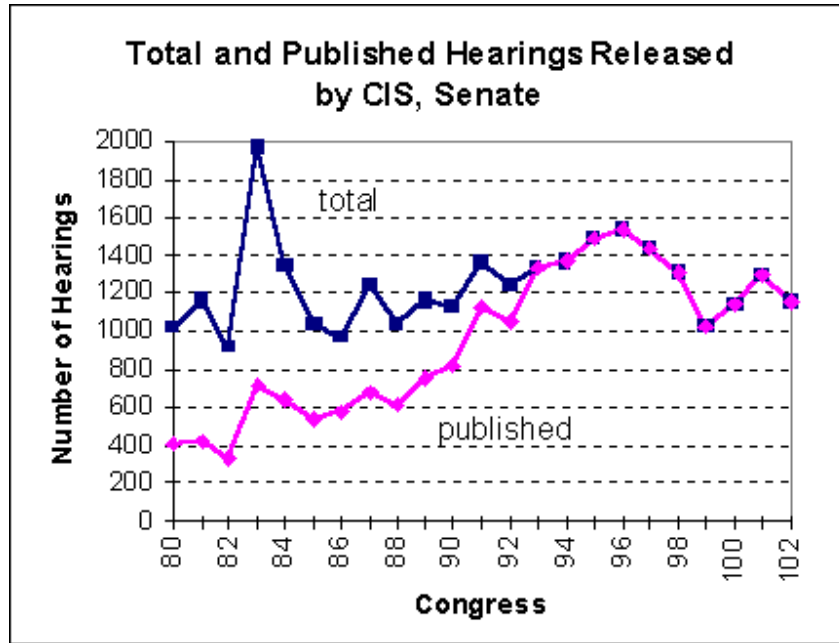
Congressional Hearings Data Issues:

The Problem of Unpublished Hearings

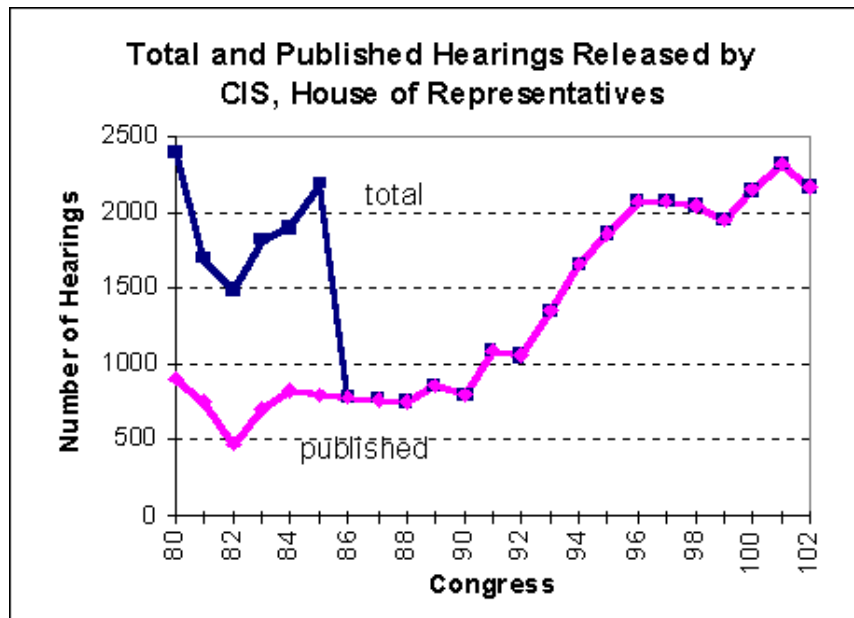
Many congressional hearings were not available for public record when the Congressional Information Service (CIS) initially began compiling hearing transcripts for publication due to the following reasons: hearings were held in executive sessions (this includes most committee business meetings and Senate nomination hearings); hearings dealt with investigations involving matters of individual privacy; hearings involved matters of national security; or hearings were simply not released at the discretion committee chairs. Many of these hearing transcripts were subsequently released.

¹All hearings have been coded from CIS volumes through the 1993 volume. Some hearings conducted in 1993 were released in the 1994 volume and CIS has published them in supplemental volumes. Supplemental volumes cover the years 1823 to 1958 for the House and 1833 to 1972 in the Senate. Thus there remain potentially large numbers of hearings (5,000-10,000) after these dates that have not yet been published by CIS and are therefore not included in our dataset. Although in recent years a much larger proportion of hearings transcripts have been published, Congress continues to withhold or delay publication of some hearings (albeit at a much reduced rate), so that the problem to a slight extent plagues the data in later years.

In the Senate, the problem of unreleased unpublished hearings is relatively minor, because CIS has processed these documents through 1972, and subsequent to this period most hearing transcripts have been published. For the Senate, CIS tabulates 28,679 total hearings, but only 21,736 of these were originally published. The figure below graphs the number of hearings by Congress for each category. It seems that there are few hearings in the Senate that have not been released; hence we recommend that analyses be based on the total number of hearings.



In the House of Representatives, the problem is more acute because CIS has released unpublished hearings only through 1958. Of 37,201 total hearings released, 30,192 were originally published. However, the problem of release is acute for a period of time from the 85th Congress to the 92nd Congress.



We suggest the following alternatives. First, one might interpolate. If one were tracing a policy topic or a committee, one would want to base the interpolations on that topic or committee, rather than the totals. Second, if one is studying a transformed variable, such as jurisdictional concentration, one could compare differences between the published and unpublished hearings for the period of overlap. Third, any comparisons between congresses will be valid except for the

85th through 92nd. Fourth, as we note below, the vast bulk of unpublished hearings occurred in a very limited number of categories. One might restrict analysis to the categories less affected. Finally, in tracing change in attentiveness over time, it is probably best to use percentage measures rather than counts, to adjust for the differences in release of unpublished hearings.

Variation by Topic and Committee

There is considerable variation in the frequency of unpublished hearings by topic and committee. The topics with the greatest percentage of unpublished hearings are defense, public lands and water management, foreign affairs and foreign aid, and government operations.

CIS notes that the House Judiciary, Interior and Insular Affairs, and Armed Services Committees have held the most hearings not released for public record, while the Senate Judiciary has held a large number of its hearings in closed session. In addition, hearings that were held in executive session were not originally published; this includes nomination hearings in the Senate and general committee business meetings in both chambers.

In the House, topics including public lands management, government operations, and defense include a substantial percentage of hearings that were not originally made public. In the Senate, the topics of government operations, housing and community development, public lands management, foreign affairs and foreign aid, and defense have 20 percent or more unpublished hearings. Thus our figures for these topics may change as CIS publishes more supplemental volumes.

Within each of the major topics that have a significant percentage of unpublished hearings there are specific subtopics that account for most of these hearings. In the House, most of the unpublished hearings in the topic defense involve subtopics such as war reparations, military installations and land transfers, and military personnel issues. In government operations, most of the unpublished hearings are on subtopics pertaining to congressional operations, District of Columbia Affairs, and the relief of claims against the U.S. government brought by private individuals and corporations. Within the topic public lands management most of the unpublished hearings are on the subtopics national parks and recreation, Indian affairs, public land transfers, and water resources management. All these subtopics are consistent with the issues that CIS notes are the areas most likely to be unpublished (CIS 1994; 1995).

In the Senate, the percentage of originally unpublished hearings by topic is a bit more spread out but again a large number of the unpublished hearings fall within specific subtopics in each the categories shown. In community development and housing issues, most of the unpublished hearings were from an investigation over a period of several months of the Federal Housing Administration, during which hundreds of confidential hearings were held on various elements of fraud and abuse. Of the 3,600 unpublished hearings in the topic government operations, almost 2,100 involved nominations, while another 800 were devoted to District of Columbia affairs. The story for the topic public lands management is much the same as in the House.

For more details, see "New Issues and Old Committees: Jurisdictional Change in Congress: 1947-93"; Frank R. Baumgartner, Bryan D. Jones and Michael C. Rosenstiehl. Paper presented at the Midwest Political Science Association, April, 1997.