

Announcing New Databases for U.S. Federal Government Policy Activities since 1947.

Frank R. Baumgartner, Pennsylvania State University
Bryan D. Jones, University of Washington
John Wilkerson, University of Washington
E. Scott Adler, University of Colorado

With substantial support from the National Science Foundation, we have compiled and make available on-line comprehensive databases concerning US federal government activities since 1947. The Policy Agendas and Congressional Bills Projects web-sites include extensive search and interactive analysis tools. They are valuable in teaching and research and are used in scores of universities across the country. They should be of interest more generally to anyone interested in analyzing trends in government activities over time, getting background information about a policy area or a particular public law, or in locating original government documents. Except where indicated below, all records are selected, not a sample. That is, the projects compile an exhaustive record of all government activities across all policy areas since World War Two.

Datasets currently available include the following:

The Policy Agendas Project (www.policvagendas.org)

- U.S. Congressional hearings, 1947 to 2000 (>70,000 hearings). Coded from the original *Congressional Information Service* abstracts.
- Public Laws, 1948 to 2002 (17,044 records).
- *Congressional Quarterly Almanac* stories, 1948 to 2002 (12,583 records). Coded from the annual *CQ Almanacs* available in many government documents libraries.
- U.S. Budget Authority, 1947 to 2003. Coded from the annual Presidential *Budget*, adjusted for inflation, using consistently defined OMB spending categories.
- Executive Orders of the President of the United States, 1945 to 2001 (3,675 records).
- Presidential State of the Union Addresses, 1947-2005.
- Public Opinion data on the Most Important Problem facing the nation, 1939 to 2001. Annual and quarterly data for 19 major topic categories, revised from original Gallup polls.
- A Random Sample of entries from the *New York Times Index*, 1946 to 2001 (36,403 records).

The Congressional Bills Project (www.congressionalbills.org)

- All public bills (H.R., S.) introduced in Congress, 1947 to 2000 (379,000 bills)
- Title, Subject area, and progress of the bill (e.g., where it was referred, whether it was reported, passed, or enacted).
- Institutional and Biographical Information about each bill's Sponsor.
- Can be linked to original Library of Congress (www.thomas.loc.gov) data for the full text for bills after 1989.
- Full text for bills before 1989 will be available in the future when digitizing is complete.

Both projects are ongoing and will be updated as resources allow. As they were funded by the National Science Foundation, data are freely available to all users at no charge. We expect to be adding the following datasets to the Policy Agendas website in the near future:

- Supreme Court Decisions, 1947-
- Congressional Roll Call Votes, 1947-
- Entries from the *Encyclopedia of Associations*, 1959-

Organization of the project: Each database is freely down-loadable and can also be analyzed interactively over the web. Databases include complete identification materials and codebooks allowing a user easily to locate the original source material. Congressional hearings are identified by their CIS numbers, allowing anyone with access to the Lexis-Nexis Congressional Universe to locate the originals. Similarly the bill numbers, public law numbers, Congressional Quarterly date and page references are systematically included. So one use of the datasets is as a sophisticated index and retrieval mechanism for publicly available sources. One can easily identify, for example, all congressional hearings, CQ stories, bills, and laws primarily focused on water pollution, international trade, immigration, or virtually any other topic.

A common and exhaustive topic coding system: The projects use a common set of topic and more detailed subtopic codes, assigning each document to one and only one policy topic category based on the predominant focus of the document. There are 19 major topic categories (agriculture, health care, energy, environment, defense, international affairs, etc.) and each of these is further subdivided into more detailed subtopics for a total of 226 subtopics in all. Budget data are coded into 62 OMB classifications. For the New York Times datasets, only the major topics are coded, and there are additional topic codes for such areas as fires, cultural reviews, and sports results where there are no government activities. The result of this comprehensive coding system is that one can analyze, with the click of a mouse, levels of government attention to any of 226 pre-defined topics.

An interactive web interface: The Policy Agendas Project web site allows users full customization options relating to downloading all or parts of each database. Without downloading anything, and from any web connection worldwide, users can quickly produce graphs and reports tracing budget or policy trends across time for any combination of the 226 topic categories. Drill-down tables offer textual descriptions of each record. These quick and interactive graphing tools allow a user quickly to understand the general trends of public policy over time in any issue area. Further analysis can be conducted with the complete source materials downloaded for easy use in any spreadsheet or statistical program.

A teaching tool: Students at all levels from high-school through graduate programs can easily learn the web interface and conduct original research-based projects on public policy developments over time. A teaching section of the web site includes syllabi and teaching materials currently being used by instructors in major universities nationwide.

A resource for the general public: With greater dissemination and knowledge of these resources, particularly through libraries, we believe that the Policy Agendas and the Congressional Bills Projects could be of wider interest among the general public as well. Besides the ability to trace public and government attention to issues over time, the projects

allow research into the background and development of particular policies. Links to original source materials, including congressional hearings, bills, and statutes facilitate public knowledge about the origins and purposes of federal policies in all areas. Our budget tool, providing consistently defined and inflation-adjusted spending totals across all areas of the federal budget over the entire post-war period, is in itself an invaluable tool for understanding public spending.

A high quality, systematic tool for consistent historical search: In contrast to the vast bulk of computer-searchable databases, the Policy Agendas and Congressional Bills Projects do not rely on keyword searches, though these are also possible in both projects. Rather, all the datasets associated with the projects are constructed with historically consistent topic categories that do not change over time. Keyword-based indices of government activities suffer from the tendency in government to use different words to describe similar programs during different historical periods. This is often not a problem when the historical period covered is relatively short. But over the past 50 years, dramatic changes have indeed taken place in the language used to describe individual public policies. All the databases associated with these projects were created by trained human coders achieving the highest levels of inter-coder reliability and applying consistent definitions of the content categories across the entire historical period and across all the component databases. The result is a series of comparable and useful databases to public record sources of the highest quality, available for free, courtesy of the National Science Foundation.

Related Publications:

Bryan D. Jones and Frank R. Baumgartner. 2005. *The Politics of Attention: How Government Prioritizes Problems*. Chicago: University of Chicago Press.

Frank R. Baumgartner and Bryan D. Jones, eds. 2002. *Policy Dynamics*. Chicago, University of Chicago Press.

Visit our web sites:

www.policyagendas.org
www.congressionalbills.org

December, 2005