

Parliament and Government Composition Database (ParlGov)*

Holger Döring[†] Philip Manow[‡]

Version 11/07 – 26 July 2011

The Parliament and Government Composition Database (ParlGov) is a new type of data infrastructure that overcomes the shortcomings of many contemporary approaches towards data collection on political institutions. With ParlGov, we aim to provide an infrastructure that systematically combines information on party positions, election results and government composition. ParlGov offers: a database and data tables for the empirical study of parties, elections and governments; a web interface with a comprehensive and accessible way of presenting the coded information, and feedback mechanisms to continuously improve the data collected over time.

Currently, ParlGov contains data on elections and governments for all EU and most OECD members for the entire post-war period. The database combines approximately 1400 parties, 640 elections with 5500 election results and 2100 governing parties in 890 cabinets. This information is linked to various external data sources. Releases at regular intervals will provide up-to-date empirical information including recent elections or changes in government, coding suggestions from country experts as well as extensions to more countries and time periods.

*Work on this project started in 2005 at the Max Planck Institute for the Study of Societies (MPIfG) in Cologne. It was later based at the University of Konstanz (2007–2009), the European University Institute (2009/10), the University of Heidelberg (2009/10) and the Centre for Social Policy Research, University of Bremen (2010–today). We would like to acknowledge the support of these institutions. The ParlGov data and the website (www.parlgov.org) contain a list of all people that have supported the project.

[†]University of Bremen – Centre for Social Policy; doering@zes.uni-bremen.de

[‡]University of Bremen – Centre for Social Policy; manow@zes.uni-bremen.de

A new infrastructure for data on parties, elections and governments

Students of political institutions are in need of a modern replacement for data handbooks to provide them with empirical data needed for their analyses in digital form, data on political institutions comparable to election survey archives and economic data as provided by the OECD. To date, the discipline lacks integrated empirical information on election results, government composition and party positions that can be easily used to study political institutions with quantitative methods. The existing sources are either available in printed form only, or focus exclusively on elections, or governments or parties. Moreover, the data available is not updated regularly and is difficult to combine. Currently, much valuable research time is spent on finding, digitalising and merging the various and heterogeneous data sources that exist online and in data handbooks. Integrated data on elections, parliaments and governments will allow researchers to focus more quickly on their substantive questions and research design. This is what ParlGov is about.

The Parliament and Government Composition Database (ParlGov) overcomes contemporary shortcomings of data on political institutions by making use of new software techniques such as databases, internet presentations and online feedback systems. There is no international organization or administrative agency that – like the OECD, Worldbank, IMF or ILO with respect to economic data – would count it among its responsibilities to collect and publicly provide data on elections, parliaments and governments. Hence, the discipline itself has to endeavor to provide scholars with such data.

ParlGov is a data infrastructure that combines existing information in a non-trivial way, presents the empirical information more comprehensibly, allows for systematic updating and steady data improvements. In addition, it incorporates feedback mechanisms to integrate the detailed knowledge of users and country experts. With ParlGov, we hope to provide a new form of data infrastructure for political scientists that significantly reduces the shortcomings of contemporary approaches to data collection. In what follows, we present ParlGov and its components in more detail.

A database to link empirical information more efficiently

Databases are widely-used information technologies, but have only rarely been used in political science. Databases enable users to store information non-redundantly, to provide consistency checks and to combine information from database tables in a non-trivial way. Powerful merging techniques can generate virtual tables by combining database tables. Our empirical information is collected in a relational database by making use of a systematic database model for parties, elections and governments. Using this data model, we can store information non-redundantly,

link tables through unique keys and make it possible to easily extract specific information for empirical studies through database queries and virtual tables. Using a database allows to avoid entering empirical information into a huge data matrix with many variables and observations. Coding all information into only one data matrix is prone to errors and difficult to maintain in the long run. Making use of a database, however, makes it possible to integrate and easily combine observations in ParlGov with other data sets. In the database we distinguish four types of observations:

- primary data: coded information (database tables)
- secondary data: virtual tables (database views based on primary data)
- external data: data sets linked to ParlGov (merge tables)
- calculated data: calculated parameters from software (R and Stata) scripts

Primary observations are the information we collect according to our database model. Currently, we focus on parties, elections, and governments. In order to collect this information we have set up multiple tables. Elections, for example, are coded in two tables: one table collects information at the election level, such as the election date and turnout data, another table assembles the election information at the party level, such as the number of seats won by and votes cast for a party. By making use of unique keys for parties, elections and governments we can combine the information in our tables through database queries. These queries can form the basis of virtual tables. The latter combine stored information and are very likely to form the basis for empirical studies, but store no information on its own. Based on the primary data, we can generate new variables and observations. We call this information secondary data. For instance, we have not collected specific information on the electoral performance of government parties in the database. However, combining the information on election results with a table that codes parties in government makes it possible to generate a virtual table that combines this information into a data matrix on the electoral performance of government and opposition parties. Information such as previous electoral performance, governmental experience and party positions may be added to the data matrix in this automated way.

Empirical information included

Information on parties, elections and governments has been collected for all EU member states and for most OECD members over the post-war period. Information on *election results* has mostly been gathered from national official statistics. We have coded all parties that won seats in national parliamentary and European Parliament elections. For some countries we have also

coded all parties that have won more than 0.5% of the votes in an election and the full official electoral results. Second, we collect detailed information on all *parties* that are included in our database. The party information is connected to a wide set of information about party positions that we will describe in the next paragraph. For the parties, we provide information on their names, changes of names and related parties. Finally, we collect information on *governments* and in particular on the parties participating in them. For these observations we gather starting and ending dates of governments and a list of the parties with ministers in cabinets. For some countries, we have started to combine this data with information on political careers, such as ministerial and parliamentary appointments. However, individual data on members of parliament and ministers will only be included in future versions of ParlGov.

In addition to the data collected for ParlGov, we combine our information with other existing data sets. Our current focus has been on linking various data sets on party positions with the party data that we have collected for ParlGov. We have linked our parties to all of the major party expert surveys, the Comparative Manifestos Project (CMP) and the EU Profiler data. These external data sources are not part of the ParlGov data set but can be combined with the ParlGov data through ‘merge tables’ that are part of ParlGov. To indicate its different origins, we call this information ‘external data’.

Currently, our database includes information on approximately:

- 1400 parties including renamings and changes
- 640 elections with 5500 party level election results
- 890 governments with about 2100 participating parties
- links to parties of all major expert surveys, CMP and EU Profiler

Finally, there is some information based on primary information in ParlGov, but cannot be generated with database queries only. Most of this information consist of analytical constructs such as the median party, the effective number of parties, pivotal players and the disproportionality of electoral systems. These parameters are calculated from primary ParlGov data with the help of software scripts (in R and Stata) that we have written. We calculate this information at regular intervals and include it in the ParlGov database (tables starting with ‘viewcalc_’). The computer scripts that perform these calculations are also part of the ParlGov data infrastructure and are available to users.

Users of ParlGov have two ways of working with the empirical information gathered in the database. They can rely on the virtual tables (secondary data) that we generate. These tables combine primary data from ParlGov into a data set. Currently, there are three major tables. A first table gives the main information on parties and links the parties in ParlGov to various

external data sets with party positions. A second table links the ParlGov information on elections at the party level. Finally, a virtual table provides information on all parties in parliament at the time of government formation and their seat strength in parliament. Most of the applications of the ParlGov data can be started by means of these three tables. More advanced users might want to combine the ParlGov tables to generate data sets for their analyses by setting up their own database queries, or by merging the primary ParlGov tables with their preferred statistical software package.

An online presentation and feedback mechanisms

Let us now briefly describe the online presentation and the feedback mechanisms that we have set up to improve and to update the empirical information that is stored in ParlGov. To help users to better understand the data-coding in ParlGov, we have developed a web interface that presents the empirical observations on various webpages. These webpages list all information in a structured way. To give an example: we generate one webpage for each party and list the entries of the party in the election result data and its government participation. In addition, we list all the party names in the external data sets that we have linked to ParlGov. We provide similar pages for election results and governments. This alternative approach of presenting our empirical data allows us to make coding decisions more transparent and to quickly find potential coding errors. Presenting data in this way makes it easier to understand the structure of the data while working with the empirical information from ParlGov for a research project.

The webpages of the ParlGov data are generated dynamically, so that corrected entries, as suggested by users, appear instantly. The software used makes it possible to enter data into the database with the help of an online interface. Project members log into the system and are given certain editing privileges to update and revise the information collected in ParlGov. This administration interface is only available for registered users with a certain set of privileges. Our current technological approach also allows us to quickly generate new webpages that may present ParlGov and new data in alternative forms. As a consequence, suggestions, error reports and data contributions from users can be incorporated into ParlGov's content quickly.

Presenting empirical information in ParlGov should allow country experts to give coding recommendations and to easily report errors. To foster feedback from experts, we have integrated different techniques that allow users to comment on the ParlGov data. There is the opportunity to enter comments on to the data webpages. These comments are sent to us via e-mail. However, we have also set up a more sophisticated way of providing recommendations and of organising error reports, as well as suggestions for improvements.

We make use of techniques that are widely applied by software developers to write computer programs. For ParlGov, we have set up a wiki and an issue management system. With the wiki,

users can make comments and provide example scripts. In the wiki, we can also more easily collect links to external data sources and different views on controversial cases. An online issue management system is also integrated into the wiki. In the issue system we collect information about planned future extensions of ParlGov and more extensive error reports. These reports can be assigned to planned future versions and their development status can be documented. Working with an issue management system allows ParlGov developers as well as its users to improve the empirical information and its presentation systematically over several versions.

Data archiving, release cycles and future developments

So far, we have focused on the online version of ParlGov and its feedback mechanisms. The empirical information in ParlGov is updated and corrected regularly. Thus, we have to be sure to provide stable versions of the data that can be replicated and to document changes between versions. To provide these stable versions of the ParlGov data, we plan regular releases at least once a year. A release provides a stable data set that can be used for empirical work in political science. In the release, we store the database, webpages, scripts that generate calculated parameters, and basic documentation. We plan to submit every released version to a social science data archive to secure an external backup of the ParlGov data. It will also be possible to access a development version that includes the latest revisions, updates and extensions and which constitutes the basis for the next official release. A news section on the pages of the development version summarises information about recent updates and changes.

ParlGov draws on new computer techniques and encourages political scientists to overcome their reliance on the data matrix. Data sets in ParlGov are generated dynamically and can be updated quickly. Using databases makes it possible to create empirical information for research on political institutions more quickly. We believe that ParlGov can also provide a more general data infrastructure that fosters collaboration among political scientists and country experts. Hopefully, ParlGov will help to overcome many of the contemporary shortcomings and redundancies in collecting, managing and creating empirical information on parties, elections and governments. We envision a data infrastructure that enables students of political institutions to focus on their empirical questions and a future where the days spent typing information from data handbooks into spreadsheets will be gone.

Reference

- ParlGov database (Döring and Manow 2011)
- Döring, Holger and Manow, Philip. 2011. Parliament and government composition database (ParlGov): An infrastructure for empirical information on parties, elections and governments in modern democracies.
- Stable version: www.parlgov.org
- Development version: dev.parlgov.org
 - user: guest
 - password: ParlGovGuest

Version history

- Summary of version changes – more details at www.parlgov.org
- ParlGov 11/07 – 26 July 2011
 - tracking ParlGov relevant events on Twitter – www.twitter.com/parlgov
 - new data: Japan elections and cabinets from 1945 to 1958 added; Northern Ireland added (parties only); Austria, Cyprus, Poland updated from seats to votes level data
 - recent elections and cabinets until mid July 2011 and data corrections
- ParlGov 10/11 – 16 November 2010
 - complete database refactoring – recoding of table/variable names and all id variables
 - new web design: Jens Hoffmeister – underline webdesign Berlin
 - new data: Swiss cabinets since 1945 added; European Election Study (2009) party ids added; Belgium, United Kingdom, Portugal, Latvia updated from seats to votes level data
 - recent elections and cabinets until October 2010 and data corrections
- ParlGov 10/02 – 25 February 2010
 - initial public release