Parliaments daybyday

Turner-Zwinkels, Tomas; Huwyler, Oliver; Frech, Elena; Manow, Philip; Bailer, Stefanie; Goet, Niels D.; Hug, Simon

Published in:
Legislative Studies Quarterly

DOI:
10.1111/lsq.12359

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

LSQ
LEGISLATIVE STUDIES QUARTERLY
Published by the Department of Political Science at Washington University in St. Louis

TOMAS TURNER-ZWINKELS
Tilburg University
OLIVER HUWYLER
University of Basel
ELENA FRECH
University of Bamberg
PHILIP MANOW
University of Bremen
STEFANIE BAILER
University of Basel
NIELS D. GOET
University of Geneva
SIMON HUG
University of Geneva

Parliaments Day-by-Day: A New Open Source Database to Answer the Question of Who Was in What Parliament, Party, and Party-group, and When

Reliably answering questions about representation and parliamentary behavior requires data about which parliamentarian was where, and at what time. However, parliament membership is not stable over time. For example, it is common for politicians to change office (we find up to 40% turnover between elections). Consequently, parliament membership, as well as party and party group composition change on a daily basis. To address the challenges that these fluctuations present, we introduce a new open-source database:' 'Parliaments Day-By-Day" (PDBD). PDBD currently contains demographic and day-by-day membership data for the national parliaments of Germany, Switzerland, and the Netherlands, covering the period between 1947 and 2017, and comprising a total of 21 million parliament-legislator-day observations. We demonstrate the usefulness of this high-resolution data in a concise study of the day-by-day development

LEGISLATIVE STUDIES QUARTERLY, 47, 3, August 2022
DOI: 10.1111/lsq.12359

of parliaments in terms of gender and seniority. This reveals hitherto unknown patterns of early turnover, gendered replacement, and seniority.

Work in legislative studies often relies on "who-is-who"; data from parliaments. Students of representation, for example, need to know who was in which parliament on what day. Similarly, when studying parliamentary behavior, such as party unity, speech-making, or legislative attention, it is crucial to know who was aligned with which political party and party group. However, despite its key role in legislative research, there is still a lack of reliable cross-country data on who was a member of which parliament, party, and party group on what day.

With the aim of filling this gap, the Parliamentary Careers in Comparison (PCC) project team[1] presents a new set of guidelines for the collection of parliamentary membership data, to advance comparative research of parliaments. The result of this effort is an open-source database: the "Parliaments Day-By-Day" (PDBD) database. The current version of this database has complete day-by-day membership data for *parliaments*, *parties*, and *party groups* for all the members of the national parliaments of Germany, Switzerland, and the Netherlands between 1947 and 2017. In total, these data include 21 million parliament-legislator-day observations. Researchers can query these data using statistical software like R, STATA, or SPSS. The data are also available on Dataverse (https://doi.org/10.7910/DVN/PYGBDO) in several ready-to-use aggregated formats.

The remainder of this article is structured as follows. In the first part, we reflect on the need for an(other) open-source parliamentary who-is-who database. In the second part of this article, we present the features of our open-source relational database that tackle these challenges. We also present the data in this database and explain how to use it. In the third part, we use these new data to present descriptive graphs that highlight hitherto unknown patterns of early turnover, gendered replacement, and sudden declines in the average seniority of its members in the parliaments currently included in PDBD.

Part 1: Does Parliamentary Research Need an(Other) Open-Source Parliamentary Who-is-Who Database?

Parliamentary who-is-who data can be defined as *time-stamped information about the roles and relations of elected*

[1]As part of the PCC project, we collected extensive political career data. The data presented here are the first part of this data collection effort.

representatives. The most important who-is-who data answers the question of who was a member of which *parliament*, *party*, and *party group* on what day.

Previous efforts to collect parliamentary who-is-who data

Given its central importance, we are not the first to be interested in parliamentary who-is-who data, nor are we the first to try to develop such data sets. Previous efforts to collect and make available parliamentary membership data tend to fall in one of three broad categories.

The first category comprises relatively **rich, country-specific data sets with relatively long time frames** (i.e., 20+ years). Some notable examples of such data sets are: the "Roster of United States Congressional Officeholders and Biographical Characteristics of Members of the United States Congress"[2] (1798–1997), and the British Political Development database (1802–2010) Eggers and Spirling (2014). These data sets offer high detail and quality.

In the second category, there are **multi-country projects that focus on a specific time frame and include a somewhat reduced feature set (i.e., fewer variables)**. The most prominent of these is the "Global Leadership project," which covers parliamentarians in 162 countries (2010–2013) Gerring et al. (2014). Two other projects in this category are the EveryPolitician-project[3] and the Comparative Legislators Database Göbel and Munzert (2021). Both projects use "wiki-data"[4]. These projects have an impressive global scope, yet the breadth and depth of information available per politician is less extensive: the current offer for most countries includes a list of the names and genders of politicians per parliament for recent[5] years, and does not include, for example, daily membership data.

The third and final category constitutes parliamentary membership information collected and (often) **shared** (online) **by the**

[2]https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/7803.

[3]https://www.wikidata.org/wiki/Wikidata:WikiProject_every_politician.

[4]A sister project to Wikipedia that offers an open knowledge base of structured data

[5]As a notable exception the lists for the United Kingdom go back until 1945.

administrative services of parliaments. These data range from printed books with "who-is-who" information (e.g., Biografisches Handbuch der Mitglieder des Deutschen Bundestages 1949–2002) and publicly available spreadsheets (e.g., the Bundestag's master data sheet, well-known among German political scientists[6]), all the way to fully-fledged API's[7] (e.g., the Irish Oireachtas[8]). These parliamentary archives offer a wide range of raw information. However, they are typically not designed to be used as research data. Hence, time-intensive data restructuring and cleaning are required to render the information they contain suitable for further analysis.

All in all, although the projects in each of these three categories serve an important purpose, it is fair to say that reliable parliamentary who-is-who data remains either disconnected, limited in either temporal resolution or depth, or time-consuming to collect or transform into a format that is suitable for further analyses. In the next section, we outline the challenges of using politician level parliamentary who-is-who data and how we dealt with them.

Challenges When Using Politician Level Parliamentary Who-is-Who Data

Politician level parliamentary who-is-who data data comes with several challenges.

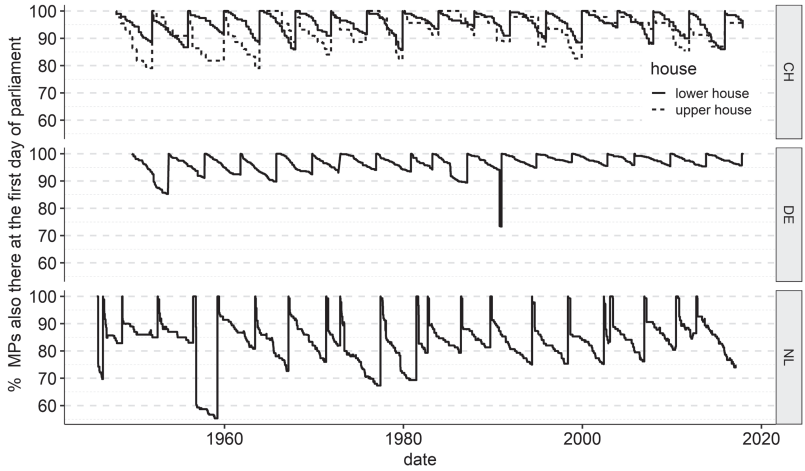
Challenge 1: Population Instability. The first challenge facing many parliamentary researchers is to identify the exact population, that is, a parliament, party, or party group's *actual* members at the time of interest. However, doing so is not easy, because—as we illustrate in figure 1—elected politicians frequently move between offices. Because of this "population instability" (defined as the "frequency of change in the target population and its information points," see Turner-Zwinkels (2020b)), the answer to a simple question like "who was in what parliament" differs depending on the day for which one asks the question. However,

[6]https://www.bundestag.de/services/opendata

[7]Application Programming Interfaces that can be used to request and download data.

[8]https://api.oireachtas.ie/

FIGURE 1
Stability of Parliamentary Membership Within Parliamentary Terms



almost all previous efforts to collect and share parliamentary who-is-who data lack sufficient detail to generate daily information. As a result, much parliamentary research currently relies on data that are at risk of being, to some degree, incomplete, misspecified, or both.

The amount and frequency of movement in and out of our research populations between elections tells us the temporal resolution that our membership data need to have. The *temporal resolution* of data specifies the smallest time frame between which changes can be detected. If it were the case that every politician who enters parliament stayed for their whole mandate, there would be no need to measure membership every day. Between elections, who was a member of what parliament would be the same no matter *when* you looked. However, Figure 1 reveals that this is **not** the case.

Figure 1[9] shows a "survival curve" of parliamentary membership on a day-by-day basis[10] for the countries currently in the PDBD database. The *x*-axis shows time. The y-axis shows the percentage of representatives in parliament that also held a seat on the first day of the first session of that particular parliament.

[9]The R-scripts used to generate all the graphs presented in the article are available via https://github.com/TomasZwinkels/PCC_daybyday

[10]Online supporting information Appendix B includes these graphs for party- and party group membership.

We learn three key things from Figure 1. First, there is considerable population **in**stability, even when we discount a number of historical shocks[11]. This is particularly the case in the Netherlands, where MPs are legally required to leave parliament when they are selected for cabinet positions. Second, we observe that movement in and out of parliament happens frequently (every couple of days) and continually (all throughout the parliamentary term). Third, we see that these small daily instabilities accumulate to large differences over time. At the end of a term, only between 60% and 85% of the initial MPs are still in parliament. These three factors underline the value of using data with a sufficiently high (e.g., daily) temporal resolution.

Challenge 2: Lacking Temporal Resolution. The second challenge is that current data often lack **temporal resolution**. Typically, data are collected in a "parliamentary mandate" format, with one observation per politician for every time they obtained a seat in a parliament. For some research questions, these (roughly) four-year snapshots lack sufficient granularity. For example, when combining information on parliamentarians with other data such as surveys, interviews, or social media data, the researcher will, for instance, need information on the party a politician was affiliated with when they posted a Tweet, not when the parliamentary term began. Similarly, in studies that focus on parliamentary activities, it is often important to know how long MPs were in parliament. Thus, Sciarini et al. (2019), used the data presented in this article to assess how the seniority of Swiss MPs affects their success in making legislative proposals. In the absence of fine-grained information, several existing studies, for example those explaining speech-making (e.g., Proksch and Slapin 2012; Bäck, Debus, and Müller 2014), do not adequately control for the number of speech opportunities an MP had. Using the data presented in this article, Frech, Goet, and Hug (2020) were able to control for the days individual parliamentarians spent in office. However, this practice is not yet common.

Challenge 3: The Lack of Standardized Flexible Data. The third challenge is a lack of flexible data in standardized formats. Current

[11]In 1953 the Netherlands House of Representatives increased in size from 100 to 150 MPs. A similar downward spike for Germany in 1990 is due to the integration of the members of the East German "Volkskammer" on October 3, 1990.

parliamentary data has typically been collected in "flat" data files (e.g., one Excel file with one unit of analysis per row) and are suitable for answering a limited set of research questions. These data often translate poorly to other applications and research questions. The Best and Cotta (2000) data, for example, are aggregated at the level of parliaments. This level of aggregation was well suited for the purpose of the authors' volume—understanding long-term trends in the representational layout of European parliaments (Best and Cotta 2000, 3)—but is, for example, less suitable for studying individual-level political behavior. The lack of a shared set of guidelines to collect and record these data also means that existing who-is-who data sets can rarely be automatically merged in a reliable fashion. The time investment required to manually exchange (i.e., "look up") politician identifiers (i.e., the number a politician was given in somebody else's data set) creates substantial hurdles for data-sharing.

Challenge 4: Data Quality Concerns. The fourth and final challenge is data quality. Obtaining a precise population or population sample is difficult. Currently, parliamentary researchers typically work towards a good sample by themselves. However, this approach is rather inefficient and error-prone. Reliable who-is-who data requires extensive triangulation and checks, so when data production is not cumulative (i.e., researchers do not reuse and incrementally improve on each others' data by making successive additions), mistakes in the data are likely to never be caught.

Part 2: The PDBD Database: What is in it and How to Use it

Having outlined the challenges involved with using parliamentary who-is-who data, we can shift our attention to how the PDBD database can be used to tackle these challenges.

Tackling the Parliamentary Who-is-Who Challenge: Database Design and Technical Features

Below we outline six important features (i.e., design decisions and features) of the PDBD database that help it to: (a) be flexible (i.e., to facilitate a broad array of possible research questions); (b) respond to the demands of modern quantitative research practices; and (c) address the challenges outlined above.

Feature 1: An Open-Source Data Repository. To alleviate data quality concerns, we have designed an open-source collective MariaDB database for legislative scholars. This database can be used to generate daily parliamentary who-is-who data (i.e., who was in what *parliament*, *party*, and *party group* on what day). Its current version (April 2021) contains data from 1947 until 2017 for Germany, Switzerland, and the Netherlands[12]. These data have been checked extensively, and are—as far as can reasonably be expected without the help of other researchers—a complete and correct representation of the exact day-by-day state of these parliaments over the last 70 years. In the future, the maintenance and extension of this database can be a community effort: if a researcher updates one incorrect—by design non-redundant—cell in the data, all researchers using this shared resource will automatically benefit.

	pers_id	res_entry_start	res_entry_end	political_function	res_entry_raw
1	NL_Rutte_Mark_1967	30jan2003	27may2003	NT_LE-LH_T3_NA_01	lid Tweede Kamer
2	NL_Rutte_Mark_1967	28jun2006	14oct2010	NT_LE-LH_T3_NA_01	lid Tweede Kamer
3	NL_Rutte_Mark_1967	20sep2012	05nov2012	NT_LE-LH_T3_NA_01	lid Tweede Kamer
4	NL_Rutte_Mark_1967	23mar2017	26oct2017	NT_LE-LH_T3_NA_01	lid Tweede Kamer

Feature 2: A Relational Database. Different research agendas often require different data structures. This renders committing to one data structure suboptimal in the long run. The PDBD data address this issue by working with a relational database structure that serves as the central repository that researchers can use to generate data at any unit of analysis they require.

Feature 3: Daily Temporal Resolution. Another sticking point is the lack of temporal resolution. Therefore, all membership information (who was a member of what parliament, and party (group)) in the PDBD database contains daily timestamps. For example, one of our data frames contains politician-specific parliamentary membership "episode data" (i.e., an event with a time-stamped start and end) in the following format:

Feature 4: Automated Detection of Data Anomalies. The storage of parliamentary who-is-who data with high temporal resolution in

[12]Data from Ireland, Scotland, Canada, and Austria are currently being processed and will be added in due course.

a central relational database comes with further benefits for data quality. First, the non-redundant storage of information in one database requires *triangulation of data points* when sources overlap. Non-redundant storage means that facts (i.e., a politician' date of first entry to parliament) are only stored once. This implies that when two sources contain competing information (i.e., different dates), the user is forced to further investigate and pick the correct value. Second, having data on a day-by-day basis allows for detailed automated day-by-day checks to ensure that data conform to several (institutional) regularities. We checked—and when necessary, corrected—the data to ensure that: (1) parliamentarians have, with few exceptions, only one party affiliation on a given day; (2) the number of parliamentarians on a given day does not exceed the number of available seats; and (3) that roles co-occur in accordance with institutional constraints (e.g., seats on parliamentary committees require parliamentary membership). We used a variety of R scripts to detect these anomalies (available upon request), and consulted additional sources to establish the correct data points for cases with contradictory or inconsistent information. For a more detailed reflection on the issues faced when collecting such data, as well as potential solutions, see Turner-Zwinkels (2020a).

Feature 5: pers_id; One Politician Identifier to Rule Them All?. To solve the issue that many data sets use a *different* identification number for the *same* politician, we suggest using "naturally occurring primary identifiers" instead of numerical identifiers. Naturally occurring primary identifiers are constructed from one or more unit-specific information points that uniquely identify a unit or person. For example, the *numerical primary identifier* for the current prime minister of the Netherlands in the archive of the Dutch Parliamentary Documentation Center is "02396." The *naturally occurring primary identifier* that we suggest to use instead is "NL_Rutte_Mark_1967." We believe that our field would benefit greatly if the personal identifier, "pers_id," whose use the PDPD data-standard revolves around, were to become common practice. This standardized "universal politician identifier" consists of the country, last name, first name, and birth year. For 99.93% of the cases included in PDBD this combination provided an identifier unique to parliamentarians across levels in the political system. For those exceptions for which this is not the case, we suggest adding the birth date (e.g., "NL_Rutte_Mark_1967feb14").

Using naturally occurring information instead of a numerical identifier might seem trivial, yet it can have profound effects. When two data sets need to be merged with numerical primary identifiers a—typically manual—look-up of the primary identifier is needed. This is not necessary with a naturally occurring primary identifier like the pers_id. Indeed, when naturally occurring primary identifiers are used, the primary identifier is constructed from data-points in the "to be added" source itself. This renders the need to look up this unit of analysis by its numerical primary identifier unnecessary. This saves time and minimizes false negatives (failure to match when there should have been a match), as well as false positives (producing a match when there should *not* have been a match).

Feature 6: links with other data sets. Whenever possible, the PDBD data include links (identifiers) to external data sets that are maintained by other researchers. The database is, for example, matched with the "Dutch Parliamentary Voting Data set" Louwerse, Otjes, and van Vonno (2018), with party-level data from "Parlgov," (http://www.parlgov.org/ Döring 2013), and the data from the the "Manifesto Project" Volkens et al. (2019a).

Taken together, these six features contribute to mitigating the challenges outlined above. The *open-source nature* and automated detection of data anomalies mitigates inefficient spending of resources and data quality concerns. Membership data with a *daily temporal resolution* alleviate the problem of population instability, as data can simply be queried for the exact day they are needed for. Furthermore, the storage of data in a flexible database structure means that collected data can be used for a wide variety of (future) research topics. Finally, the usage of the *pers_id* and the included *links with other data sets* means that it is relatively easy to merge the PDBD data with other data sets.

Available Data and How to Use it

Having described the PDBD database and its main features, we now turn our attention to the data that the database offers, and how they can be used. In a relational database, the data are defined not only by the available variables, but also by the relationships between data frames. Politicians, for example, belong to (i.e., are "cross-nested in") certain parliaments and parties at specific

points in time. Hence, when describing what the PDBD database contains, we need to specify its primary content (the data structure and variables at its most fine-grained level), as well as the universe of less fined-grained aggregated data that the database can generate. We do both below.

PDBD Primary Content: Data Frames and Variables. Table 1 summarizes the data structure at its most fine-grained level.[13] Please also see online supporting information appendix D for a complete list of currently included variables. Figure 2 furthermore illustrates the relational organization of the database, for example, showing how individual politicians (see "POLI"[14]) are matched to parties ("PART") via membership episodes ("MEME"). This relational data structure can be used to generate data at any unit of analysis required. Doing so involves four steps which we outline below.

Using the Data(Base). The PDBD data are available in two key formats: as a **queryable database** and in several **pre-generated datafiles**. To query the database server directly from commonly used statistical software like R, STATA, and SPSS, an access keys is required. These can be requested via https://parlcc.net/ or mailto: tomas.turner-zwinkels@uvt.nl. The script needed to do so is available via our website at https://parlcc.net/. For less database-minded users, we also offer the data in several pre-aggregated formats (see Figure 3) via Dataverse (see https://doi.org/10.7910/DVN/PYGBDO).

Four Steps From Primary Database to Aggregated Formats Ready for Analysis. Figure 3 outlines the four steps necessary to turn primary database data into aggregated data tables ready for analysis (an R script can be found on the data-verse repository at https://doi.org/10.7910/DVN/PYGBDO and on our website at https://parlcc.net/. Regardless of the statistical software used, first the unit of analysis needs to be defined, both on the *actor* and on the *time* dimension (see Figure 4 for an overview).

[13]The MariaDB uses mapping tables to connect some of the data frames. For clarity, these mapping table are not included in this diagram.

[14]Each data frame in the PDBD database has a four-letter abbreviation.

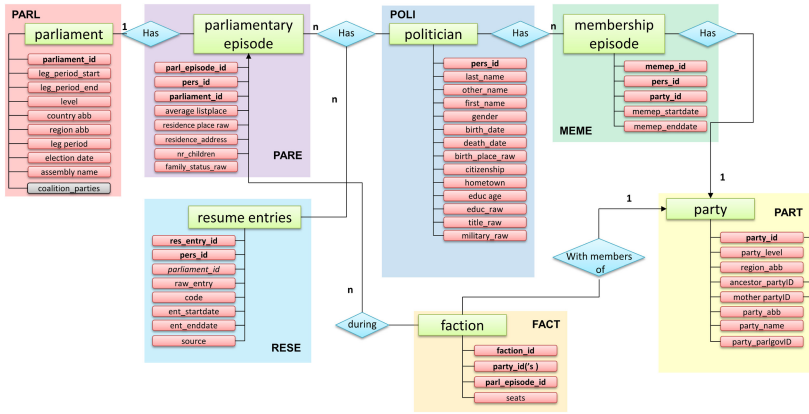
TABLE 1
Overview of Main Information and Key Variables in Each Data Frame

Dataframe	Country	N*	Main Information Include	Featured Variables
POLI	NL	1384	Static individual characteristics	*pers_id*
	DE	4073		*name*
	CH	1714		*gender*
				date_of_birth
				id_de_parliament
				id_ch_parliament
PARE	NL	3497	Episodes in parliaments	*pers_id*
	DE	13933		*parliament_id*
	CH	5649		
PARL	NL	22	Parliament term characteristics	*parliament_id*
	DE	19		*leg_period_start*
	CH	36		*assembly_name*
				coalition_parties
MEME	NL	1537	Episodes in parties	*pers_id*
	DE	4308		*party_id*
	CH	1830		*memep_startdate*
				memep_enddate
PART	NL	42	Political party characteristics	*party_id*
	DE	35		*party_name*
	CH	226		*party_parlgov_id*
RESE	NL	7984	Resumé entries (e.g., political job, educational and professional job episodes)	*pers_id*
	DE	23750		*res_entry_raw*
	CH	10111		*res_entry_start*
				res_entry_end
				*political_function_code***
				isco08
FACT	NL	225	Faction (i.e., party group) level characteristics	*faction_id*
	DE	1223		*faction_name*
	CH	1060		*faction_party_ids*
				parliament_id
				faction_start
				faction_end

**Rows of data currently in this data frame for this country.

***A five-part code specifying the geographical level, institutional domain, organization level, policy area, and position for political jobs.

FIGURE 2
Entity Relationship Diagram of the PDBD Database [Colour figure can be viewed at wileyonlinelibrary.com]



Given that the database offers information on the legislator-day level as its highest resolution, this will likely require aggregating information, particularly on the time dimensions.

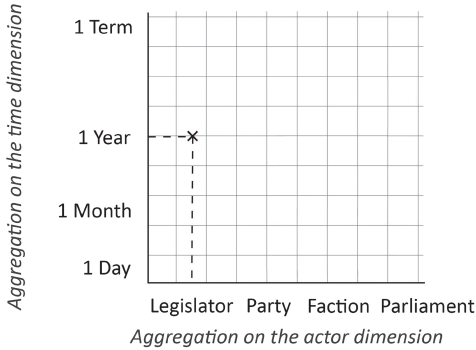
Second, the core data structure needs to be generated. This includes deciding which observations to include. For example, when employing legislator-year as the unit of analysis, a decision must be made as to what kind of legislators to incorporate into the sample: those present all 365 days of the year, those present at least half of the year, etc.?

Third, users will enrich their custom data set with (internal) variables. This requires another aggregation decision, this time for the variables. For example, legislators' party membership is observed on 365 days every year, but the researcher might need only a single value per year. This requires a decision on what party label to use for legislators who changed their party affiliation in a given year: should the party be used to which the legislator belonged on 1 January or 31 December, or perhaps the party to which they have belonged the longest?

Lastly, the custom data set needs to be expanded with extra (external) data. To that end, the necessary identifiers (e.g., pers_id) need to be added to the external data set. Alternatively, one of the external numerical identifiers already included in the PDBD database can used (check our website at https://parlcc.net/ for a current overview) to merge in additional variables.

FIGURE 3
The Four Steps Required to Construct an Analysable Data set on the Basis of the PDBD Database

1 **Define the unit of analysis**



2 **Generate the core data structure** ***(scope of the sample)***

pers_id	year
CH_Abate_Fabio_1966	2016
CH_Abate_Fabio_1966	2017
CH_Allemann_Evi_1978	2016
CH_Allemann_Evi_1978	2017
CH_Amaudruz_Celine_1979	2016
CH_Amaudruz_Celine_1979	2017
CH_Ammann_Thomas_1964	2016
CH_Ammann_Thomas_1964	2017

3 **Query the database to construct measures on the level of the unit of analysis**

	Time-invariant variable	Time-variant categorical variable	Time-variant continuous variable
Choose measure	Gender *(from POLI)*	Party membership *(from MEME)*	Tenure *(from RESE)*
Set aggregation rule, if needed	No aggregation required	Use modal party membership	Use mean (or modal) tenure

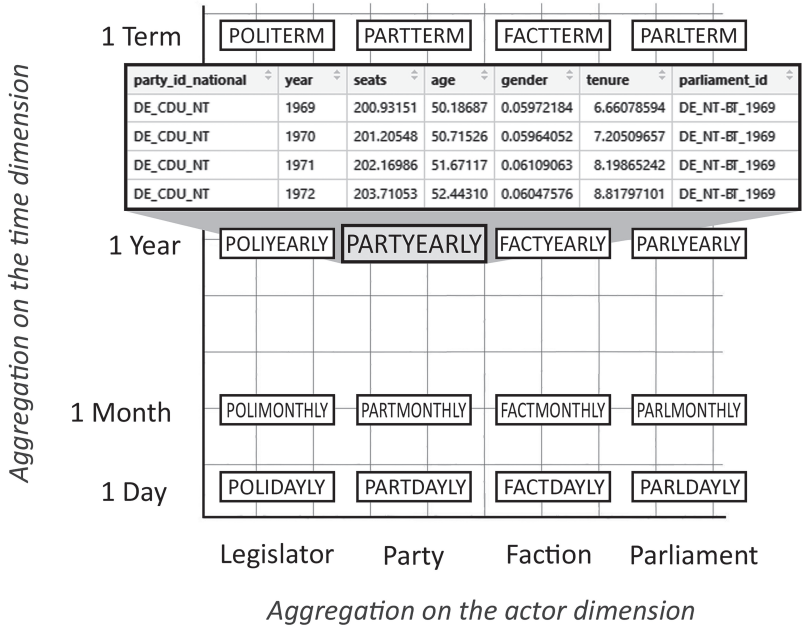
4 **Enrich the dataset with additionally collected data**

pers_id	year	gender	party_id	tenure
CH_Abate_Fabio_1966	2016	m	CH_FDPLib_RE-TI	15
CH_Abate_Fabio_1966	2017	m	CH_FDPLib_RE-TI	16
CH_Allemann_Evi_1978	2016	f	CH_SP|PS_RE-BE	12
CH_Allemann_Evi_1978	2017	f	CH_SP|PS_RE-BE	13
CH_Amaudruz_Celine_1979	2016	f	CH_SVP|UDC_RE-GE	4
CH_Amaudruz_Celine_1979	2017	f	CH_SVP|UDC_RE-GE	5
CH_Ammann_Thomas_1964	2016	m	CH_CVP|PDC_RE-SG	0
CH_Ammann_Thomas_1964	2017	m	CH_CVP|PDC_RE-SG	1

pers_id	year	speeches
CH_Abate_Fabio_1966	2016	57
CH_Abate_Fabio_1966	2017	61
CH_Allemann_Evi_1978	2016	53
CH_Allemann_Evi_1978	2017	27
CH_Amaudruz_Celine_1979	2016	47
CH_Amaudruz_Celine_1979	2017	23
CH_Ammann_Thomas_1964	2016	66
CH_Ammann_Thomas_1964	2017	82

Add additional data by matching on the unit of analysis

FIGURE 4
Sixteen pre-generated PDBD data formats offered on Dataverse



Pre-Generated Data Extracts. To facilitate usage by all, we also offer the PDBD data in several "out of the box" formats on Dataverse (see https://doi.org/10.7910/DVN/PYGBDO) so that steps 1–3 may be skipped. Figure 4 gives an overview of all of the formats provided. Currently, the table-structure of the database contains information about four discrete actor: legislators, parties, factions, and parliaments. These are displayed on the y-axis of figure 4. In contrast, the time dimension is continuous. The smallest available time resolution is daily, with any further aggregation being at the discretion of users (e.g., 15-day periods). However, as out of the box time aggregations, we currently offer four discrete choices: *day*, *month*, *year*, and *term*. These are shown on the x-axis of Figure 4. Together, these two dimensions result in 16 pre-generated data frames. For example, the previewed PARTYEARLY dataframe (in Figure 4) offers party information aggregated to the year level. We use the arithmetic mean as the method of aggregation for the variables (e.g., a party's mean seat

share in a certain year). Table 5 in the online supporting information contains three example projects for additional inspiration.

Part 3: New Data, New Insights?

Having established the rationale behind the database and its content, we now utilize the increased temporal resolution offered by the PDBD data to recreate some well-known descriptive statistics.

Female Representation Day-by-Day

Figure 5 depicts the percentage of female MPs in the national parliaments of Germany (Bundestag), the Netherlands (Tweede Kamer), and Switzerland (Nationalrat and Staenderat) since the 1950s on a day-by-day basis. Doing so reveals an interesting hitherto unknown empirical pattern: in all three countries the percentage of women in parliament increased steadily **between** elections. In fact, as can be seen by the relative lack of clear "jumps" in Figures 5(a) and 5(c), in the Netherlands and Switzerland, more women enter parliament *gradually* between elections than *suddenly* on election day[15]. Figure 5(c) furthermore reveals that the PDBD membership data (compared to the IPU data, represented by crosses) has lower levels of measurement error and offers substantial additional historical data points.

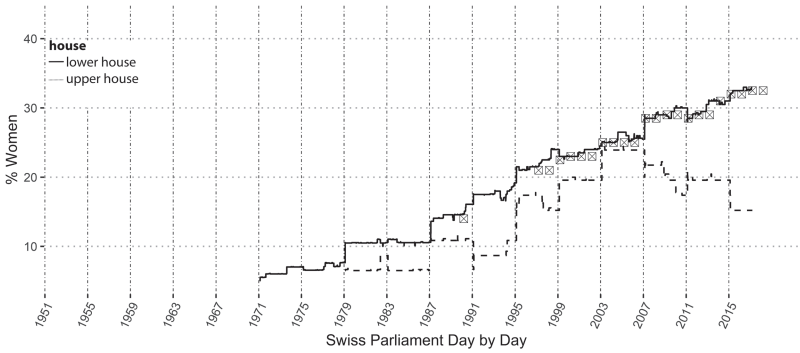
Tenure Day by Day: Seniority, But Not as We Know It?

The PDBD data also facilitate the study of parliamentary tenure, that is, the average number of years MPs spent in parliament at a certain point in time (e.g., Saalfeld 1997). Western representative democracies are believed to have professionalized with increasingly tenured members. However, when we use the PDBD data (Figure 6), we only see an increase for Germany.

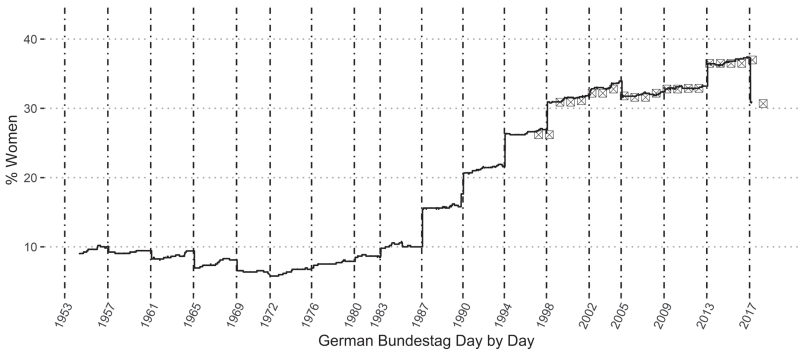
To scrutinize this trend, we utilize two unique features of the PDBD database: **party membership episode data** and its **high**

[15]Why female MPs have a greater likelihood of entering parliament between the terms receives further attention by Turner-Zwinkels and Frech (2020), McAllister and Studlar (2002), Studlar and McAllister (2002), and Salmond (2006).

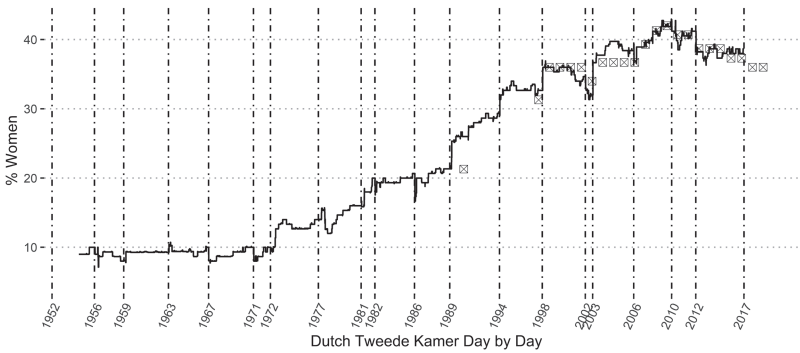
FIGURE 5
Graphs Showing the Day-by-Day Share of Female Representatives in the National Parliament Over Time. The Black Line Represents the Number Calculated on the Basis of Our Data. The Crosses Represent the—Seemingly Somewhat Biased—Yearly Data-Points Offered by the Inter Parliamentary Union



(a) Switzerland

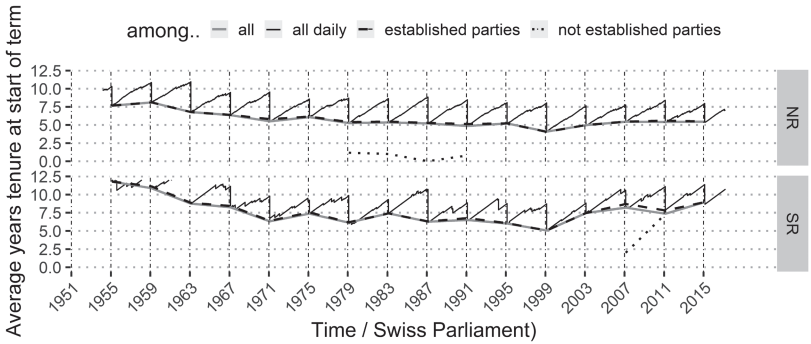


(b) Germany

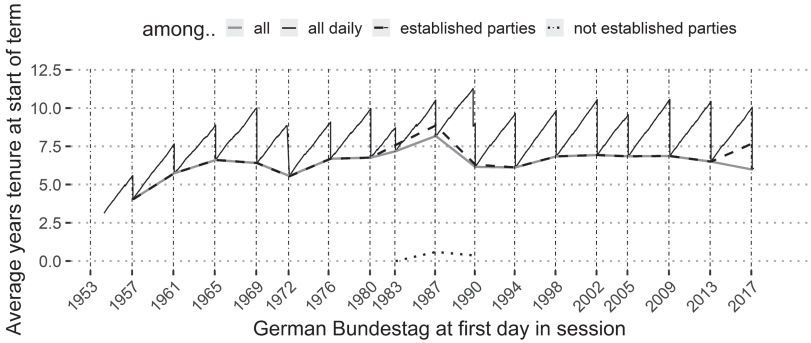


(c) Netherlands

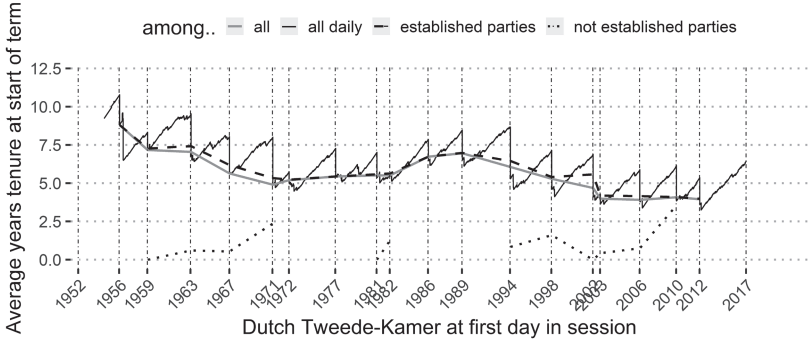
FIGURE 6
Graphs Showing the Development of the Average Tenure (Years that Members Have Been in Parliament) on the First Day the Parliament is in Session. We Show the Average for All Parties (Solid Line), Established Parties (Those that Have Been in Parliament at Least Two Terms, Short-Dashed), Non-Established New Parties (Long-Dashed), and the Detailed Development of Tenure on a Daily Basis (Thin Saw-Tooth Patterned Line)



(a) Switzerland



(b) Germany



(c) Netherlands

temporal resolution. First, we investigate whether reduced average tenure might be caused by emerging **parties**. Success for such parties is easier in the Dutch and Swiss voting systems where, unlike in the German Bundestag, there is no election threshold. Interestingly, the breakdown presented in figure 6 reveals a clear trend towards **reduced seniority**, even among established parties. The idea that emerging parties played a role is thus not supported by the data.

Second, utilizing the database's high **temporal resolution**, we inspect the **day-by-day** development of tenure (depicted as a slightly thinner saw-tooth patterned lines), and observe three patterns. First, we see that tenure gradually increases over time within parliaments. If nobody leaves, a parliament on average increases its aggregate tenure at a rate of one day per MP for each day that passes. In Germany, with its low between-election turnover (see Figure 1), we indeed see the most steady rise between elections. Second, the occurrence of elections matters. Because of newcomers, tenure drops steeply with every general election. We know from the professionalization literature that some parliaments are more open to newcomers than others. Moreover, dramatic election outcomes might induce stronger shifts by replacing MPs from losing parties with candidates from winning parties. Third, the frequency of elections has an impact: whenever there is an election, tenure drops. This saw-tooth pattern thus reveal three analytical factors to examine when seeking to explain the long-term development of tenure in parliaments: (1) the slope of the climb between elections (early replacement of MPs); (2) the drop in tenure at elections (replacement of MPs at elections); and (3) the frequency of elections. Figure 6 reveals how the relative strength or frequency of these three different factors shapes the long-term trends that can be observed for each country.

In *Switzerland*, the time between elections and the buildup of tenure between elections is stable. However, the drop in tenure at elections is relatively large, particularly between 1955 and 1971. Overall, the average tenure of Swiss MPs dropped from roughly 7.5 years to roughly 5 years in 1970, and has since stabilized. In Germany, the post-World-War II parliament was inexperienced for obvious reasons. It recovered from this shock by around 1970. We can see a steady upward line, resulting from stable membership between elections. We can also see that early (snap) elections are rare. As a result, tenure in the German parliament is quite stable. The current average is around ±7 years, and it has been like this

since roughly 1990. Finally, in the Netherlands, a relatively large number of MPs leave parliament between elections. Also, snap elections are common. Because of this, drops in tenure occur frequently. This seems to have driven a gradual **de**crease in Dutch tenure that currently approaches an average of four years at the start of each legislative term.

All in all, the professionalization literature has emphasized the importance of openness to newcomers, and hence turnover at elections. The PDBD data bring nuance to this idea by showing that other important factors are also at play.

Part 4: How to Contribute to PDBD

The success of the PDBD open-source database is reliant on its use, maintenance, and contributions by the parliamentary research community. We, as part of the PCC project intend to gradually expand this database with, among others electoral information (e.g., election lists positions, candidacy type, votes) and political career data (e.g., political and non-political jobs). We also invite researchers to contribute their own data. Templates for submitting data will be made available. Validation of these data will be done by the team, using a combination of automated and manual checks depending on a mutual agreement between the involved researchers. Information on how to contribute to the PDBD database will be made available at https://parlcc.net/.

Conclusion

In spite of ongoing developments, legislative research still suffers from a lack of detailed basic data on democratic representatives, even in otherwise information-dense contexts like Western European parliaments. The PDBD database can overcome many of the problems related to membership data on *parliaments*, *parties*, and *party groups*. Our naturally occurring identifier the "pers_id" is an important standard that could substantially ease the exchange and merging of information from different sources.

Our brief analysis of gender and seniority shows that there is much to be gained from who-is-who data with higher temporal resolution. We learned that replacement candidates might play a hitherto unexplored role in the increased representation of women. We also illustrated how the tenure profile of parliaments results from an under-theorized interplay between electoral (in)stability,

early departures and turnover.

Legislative studies require reliable cross-country data on who was a member of which *parliament*, *party*, and *party group* on what day. We hope that the PDBD database will become the central collective platform for maintaining and growing this important resource for the parliamentary research community.

Acknowledgements. For this project from the Swiss National Science Foundation (SNF 162427) and the German National Science Foundation (DFG MA-4023/4-1) is gratefully acknowledged. Open Access Funding provided by CSAL.

Conflict of interest. The authors certify that they have NO affiliations with or involvement in any organization or entity that could cause a conflict of interest.

Data Availability Statement. The **data** that support the findings of this study are openly available on Dataverse at https://doi.org/10.7910/DVN/PYGBDO. The **R-scripts** used to generate the graphs presented in the paper are available via https://github.com/TomasZwinkels/PCC_daybyday

Tomas Turner-Zwinkels is an assistant professor in sociology at Tilburg University. His research interests include political human capital, gender equality and the formal modelling of (political) career decisions.

Oliver Huwyler is a PhD candidate in the Department of Political Science at the University of Basel. He conducts research in the area of legislative studies and interest representation.

Elena Frech is post-doctoral researcher at the University of Bamberg. Her research interests include political institutions, parliaments, political behavior, representation and gender.

Philip Manow is professor of Comparative Political Economy at the University of Bremen.

Stefanie Bailer is professor of political science at the University of Basel. Her research interests encompass decision-making at the European and international level, parliamentarians and parliamentary careers.

Niels Goet is a Data Scientist at Inspera AS

Simon Hug is professor in the Department of political science and international relations at the University of Geneva.

REFERENCES

Bäck, Hanna, Marc Debus, and Jochen Müller. 2014. "Who Takes the Parliamentary Floor? The Role of Gender in Speech-Making in the Swedish Riksdag." *Political Research Quarterly* 67: 504–18.

Heinrich Best, and Maurizio Cotta (eds). 2000. *Parliamentary Representatives in Europe, 1848–2000: Legislative Recruitment and Careers in Eleven European Countries*. Oxford: Oxford University Press.

Ceron, Andrea, and Elisa Volpi. 2020. "Breakups Hurt: Party Switching and Perceived Proximity Between Politicians and Their Party." *Party Politics*, forthcoming.

Döring, Holger. 2013. "The Collective Action of Data Collection: A Data Infrastructure on Parties, Elections and Cabinets." *European Union Politics* 14: 161–78.

Eggers, Andrew C, and Arthur Spirling. 2014. "Electoral Security as a Determinant of Legislator Activity, 1832–1918: New Data and Methods for Analyzing British Political Development." *Legislative Studies Quarterly* 39: 593–620.

Frech, Elena, Niels D. Goet, and Simon Hug. 2020. "(forthcoming). The Politics of Legislative Debate in Switzerland." In *The Politics of Legislative Debate Around the World*, edited by Hanna Bäck, Marc Debus, and Jorge M. Fernandes. Oxford: Oxford University Press.

Göbel, Sascha, and Simon Munzert. 2021. "The Comparative Legislators Database." *British Journal of Political Science*1–11.

Gerring, John, Erzen Oncel, Kevin Morrison, and Philip Keefer. 2014. *The Global Leadership Project: A Comprehensive Database of Political Elites*. Online available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2491672.

Heller, William B., and Carol Mershon. 2008. "Dealing in Discipline: Party Switching and Legislative Voting in the Italian Chamber of Deputies, 1988–2000." *American Journal of Political Science* 52: 910–925.

William Heller, and Carol Mershon (eds). 2009. *Legislative Party Switching and the Foundations of Political Parties*. Palgrave Macmillan.

Louwerse, Tom, Simon Otjes, and Cynthia van Vonno. 2018. "The Dutch Parliamentary Behaviour Dataset." *Acta Politica* 53: 149–66.

McAllister, Ian, and Donley T. Studlar. 2002. "Electoral Systems and Women's Representation: A Long-Term Perspective." *Representation* 39: 3–14.

McElroy, Gail, and Kenneth Benoit. 2009. "Party Group Switching in the European Parliament." In *Legislative Party Switching and the Foundations of Political Parties*, ed. William Heller, and Carol Mershon. Palgrave Macmillan.

McElroy, Gail, and Kenneth Benoit. 2010. "Party Policy and Group Affiliation in the European Parliament." *British Journal of Political Science* 40: 377–98.

Mershon, Carol. 2014. "Legislative Party Switching." In *The Oxford Handbook of Legislative Studies*, ed. Shane Martin, Thomas Saalfeld, and Kaare Strøm, 418–35. Oxford University Press.

Mershon, Carol, and Olga Shvetsova. 2013. *Party System Change in Legislatures Worldwide: Moving Outside the Electoral Arena*. Cambridge University Press.

Proksch, Sven-Oliver, and Jonathan B. Slapin. 2012. "Institutional foundations of legislative speech." *American Journal of Political Science* 56: 520–37.

Saalfeld, Thomas. 1997. "Professionalisation of Parliamentary Roles in Germany: An Aggregate-Level Analysis, 1949–94." *The Journal of Legislative Studies* 3: 32–54.

Salmond, Rob. 2006. "Proportional Representation and Female Parliamentarians." *Legislative Studies Quarterly* 31: 175–204.

Sciarini, Pascal, Manuel Fischer Roy Gava, and Frédéric Varone. 2019. "The Influence of Co-Sponsorship on MPs' Agenda-Setting Success." *West European Politics, forthcoming*.

Studlar, Donley T, and Ian McAllister. 2002. "Does a Critical Mass Exist? A Comparative Analysis of Women's Legislative Representation Since 1950." *European Journal of Political Research* 41: 233–53.

Turner-Zwinkels, Tomas. 2020a. "Creating Political Science Data from Secondary Sources: A Classification of Issues and Their Solutions." In *A New Macro-Micro Approach to the Study of Political Careers: Theoretical, Methodological and Empirical Challenges and Solutions*, ed. Tomas Turner-Zwinkels, 47–85. chapter 2, Interuniversity Center For Social Science Theory and Methodology (ICS).

Turner-Zwinkels, Tomas. 2020b. *A New Macro-Micro Approach to the Study of Political Careers: Theoretical, Methodological and Empirical Challenges and Solutions PhD thesis Interuniversity Center For Social Science Theory and Methodology (ICS)*. Groningen.

Turner-Zwinkels, Tomas, and Elena Frech. 2020. "More Women in Parliament, but Why? The Underrated Role of Replacement Candidates." *Working Paper*.

Andrea Volkens, Werner Krause, Pola Lehmann, Theres Matthiess, Nicolas Merz, Sven Regel, and Bernhard Wessels (eds). 2019a. *The Manifesto Data Collection. Manifesto Project (MRG/CMP/MARPOR). Version 2019a.*. Berlin: Wissenschaftszentrum Berlin für Sozialforschung (WZB)

Volpi, Elisa. 2019. "Ideology and Party Switching: A Comparison of 12 West European Countries." *Parliamentary affairs* 72: 1–20.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web site:

Table 2: Summary of Supplementary Materials and Data on Dataverse
Figure 7: Stability of Party Membership within Parliamentary Terms
Figure 8: Stability of Party-Group Membership within Parliamentary Terms
Table 3: Overview of Key Sources and Extraction Techniques Used Per Data Frame
Table 4: The Parliaments-Day-by-Day Database: All Variables
Table 5: Example Projects or How the PDBD Data Could be Used