PCC - Codebook V4.0.0

The PCC project team

December 2, 2024

Contents

List of Tables **3**

1 Project Description **4**
1.1 Aim of this Document . . . 4
1.2 Short Project Summary . . . 4
1.3 Mid-term Road-map of the Project . . . 4

2 Data Structure, Sources & Sampling **4**
2.1 Data Structure . . . 4
2.2 Data Sources . . . 5
2.3 Sample . . . 5

3 How to Read this Codebook **5**
3.1 Labels, Abbreviations, Etc. . . . 5
3.2 Collection Effort Per Variable . . . 5

4 Politician Level Data Frame (POLI) **7**

5 Parliamentary Episode Data Frame (PARE) **10**

6 Resume Entries Data Frame (RESE) **11**
6.0.1 left- and right censoring of dates . . . 15
6.1 The (political) functions coding scheme . . . 16
6.1.1 Principles for Classification of (political) functions . . . 16
6.1.2 Primary Political Function Coding . . . 16
6.1.3 Secondary Political Function Codes . . . 30
6.1.4 Secondary Political Function Coding: Bi-Cameral System (Lower House, Upper House) Applies to: X_LE: . . . 33
6.1.5 Political functions coding: some final specifications . . . 33

7 Parliament Data Frame (PARL) **34**

8 Membership Episode Data Frame (MEME) **35**

9 Party Data Frame (PART) **37**

10 Faction Data Frame (FACT) **38**

11 Committee Data Frame (COMM) **39**

12 Election Data Frame () **42**

13 Electoral List Data Frame (ELLI) **43**

14 Electoral List Entry Data Frame (ELEN) **45**

15 Electoral District Data Frame (ELDI) **47**

16 Organisations Module(ORGS) **49**

17 The Quota Module (QUOT) **53**

18 Appendices **54**
18.1 Alternative IDs: ALID . . . 54
18.2 Regional Abbreviations Used . . . 55
18.2.1 Regional Abbreviations Used for CH . . . 55
18.2.2 Regional Abbreviations Used for DE . . . 55
18.2.3 Regional Abbreviations Used for NL . . . 56

List of Tables

1 Summary of Politician Level Variables (POLI) . . . 7
2 Summary of the Parliamentary Episode Level Variables (PARE) . . . 10
3 Summary of the Resume Entry Variables (RESE) . . . 11
4 Political functions pf_orglevel coding specifications . . . 23
5 Example of Three-tier Coding . . . 23
6 Example of Secondary Coding . . . 24
7 Political functions pf_policy_area coding specifications: policy area codes Note: The letter S in front of the nummerical code specifies newly created codes based on the U.S: SIC classification. . . 28
8 pf_position: Occupational Position Codes . . . 28
9 Elected or Non-Elected Political Functions - **DNC** . . . 30
10 Specification of political function secondary code: legislative steering body codes . . . 31
11 Youth or Regular Party Political Functions . . . 31
12 Shadow or Regular Political Functions - DNC . . . 31
13 Important Party Committees . . . 32
14 Specification of political function secondary code: interest group codes . . . 33
15 Upper or lower house . . . 33
16 Summary of Parliament Level Variables (PARL) . . . 34
17 Summary of Membership Episode Variables (MEME) . . . 35
19 Summary of Faction Variables (COMM) . . . 40
20 Summary of Election Level Variables () . . . 42
21 Summary of the Electoral List Level Variables (ELLI) . . . 43
22 Summary of the Electoral List Entry Variables (ELEN) . . . 45
23 Summary of District Level Variables (ELDI) . . . 47
24 Summary of Organisation Level Variables (ORGS) . . . 49
25 Summary of Quota Variables (QUOT) . . . 54
26 Alternative IDs . . . 55

1 Project Description

1.1 Aim of this Document

The primary purpose of this codebook is to provide detailed information on the data structure and variables of the parliamentary careers in comparison data. This codebook's secondary - much more ambitious, long-term and as such necessarily more controversial - purpose is to establish a new[1] data-standard. As such, we did aim the design a data-structure and coding-standard that can be used to study political careers in any context (country, level, system..). Any input that informs our success in this regard is very welcome!

1.2 Short Project Summary

The research project "Parliamentary Careers in Comparison" (PCC) aims to investigate political careers and activities of parliamentary candidates and parliamentarians in Switzerland, Germany and the Netherlands[2] since the Second World War. Based on an extensive collection of detailed individual-level data, we investigate biographical and behavioural dynamics to obtain a full and dynamic picture of parliamentary careers. The research interests covered by the project cover three broad areas. The first set of questions revolves around career paths on different levels, how they typically develop, and how these patterns can be explained. With the second set of questions we focus on the impact of political institutions on political careers while a third strand highlights the consequences of the different career paths on parliamentary behaviour and future post-parliamentary careers.

The PCC data-set contains detailed information on the political careers of parliamentarians and parliamentary candidates. Examples of variables include socio-demographic information of national candidates and parliamentarians of the three countries since the second world war, the candidate list positions, political functions (if elected) of national parliamentarians. Future versions of this data will also include data on regional parliamentarians and expand the range of included variables. See `parliamentarycareersincomparison.org` for more information.

1.3 Mid-term Road-map of the Project

Data will be collected in several waves over a three year period (2016-2019). The codebook below summarizes the data-collections efforts in the first wave. Future waves will likely include the collection of both *similar data on other politicians* (e.g. candidates for the national parliament, regional parliamentarians) and *different data for the same politicians* (e.g. additional variables in parliamentary behavior). Please drop us an email if you would like to see specific variables included. We can't promise anything but are always happy to consider!

2 Data Structure, Sources & Sampling

2.1 Data Structure

The PCC data-structure is organized into five levels and eleven data-frames. Examples of these data-frames can be found here (.xlsx download). The logic behind this data-structure is to collect static and time-varying variables as much is possible in an 'intuitive' way, sticking close to the structures that researchers are used to and the format that different kinds of data (e.g. biographical profiles, electoral lists) are typically stored in. In practice this implies that each data frame only contains information varying on the same level (e.g. the individual or the electoral list). We use a system of identifiers to allow data-points to be connected up together. Figure 1 summarizes the data-frames and the relations between them.

- Politician level data
  - Stable characteristics - (see page 4) - POLI
  - Parliamentary episodes - (see page 5) - PARE
  - Resume entries - (see page 6) - RESE
  - Membership episodes - (see page 8) - MEME
- Parties (see page 9) - PART
- Parliaments (see page 7)- PARL

[1] See Turner-Zwinkels 2016 for a specification of the requirements of such a new standard.

[2] Most of the Dutch data has already been collected as part of the PhD thesis of PCC project member Tomas Turner-Zwinkels.

- Factions (see page 10)- FACT
- Committees (see page 11)- COMM
- Electoral list data
  - Electoral districts (see page 15) - ELDI
  - Electoral lists (see page 13) - ELLI
  - Electoral list entries (see page 14) - ELEN

2.2 Data Sources

Data was collected by a combination of different **sources**. A main source for the CH and DE parliamentary career data are yearbooks. A main source in the Netherlands was the website of the Dutch parliamentary documentation centre. Other data stem from different statistical offices or are based on newspaper articles or personal websites.

2.3 Sample

Data can be regarded a **complete (population) sample of national parliamentarians since 1946 of all three countries (CH,DE[3],NL)** (wave one data collection).Both parliamentarians in the lower-house (i.e. Bundestag, Nationalrat, Tweede Kamer) and the upper-house (i.e. Staenderat, Eerste Kamer) are included. Future waves will likely include candidates for national elections in Germany and Switzerland (and maybe the Netherlands) since 1946, and of regional parliamentarians of all three countries (wave two). The data will likely also include data on the candidates for elections to regional parliaments (e.g. German Bundeslaender or Swiss Cantons).

3 How to Read this Codebook

3.1 Labels, Abbreviations, Etc.

- Brackets ('[' and ']') in variable names signify variable names. [country]_ for example means that there will be as many of this variable (set) in the data-frame as their are countries in the data.
- 'DNC' means: Do Not Collect, these are specifications that are kept in codebook for reference purposes who values we do not intend to collect (to this level of detail).
- if a variable name contains an (by underscores) indexed element on the lists which is not used 'NA' will be used a filler.

3.2 Collection Effort Per Variable

Not all variables are collected with the same effort. Our effort to collect different variables varies from the highest level ('Collect and follow-up until COMPlete') to the lowest level ('Collect When Easy'). These are the matching abbreviations:

- COMP - Collect and follow-up until complete
- CWA - Collect when available
- CWE - Collect when easy

[3]Up until 1990 only data on West-Germany is included. For obvious reasons.

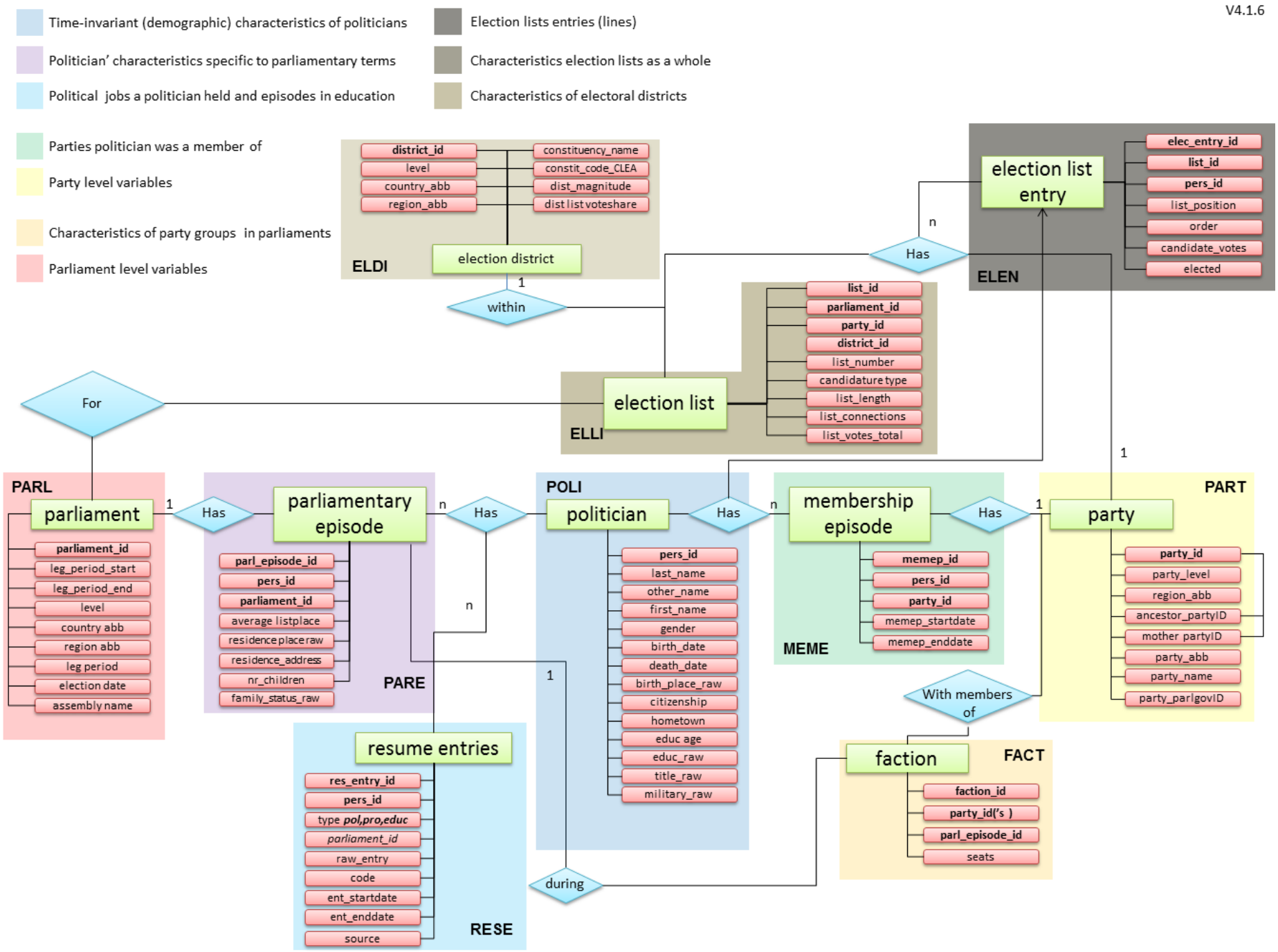

Figure 1: Entity Relationship Diagram of the data

4 Politician Level Data Frame (POLI)

Politician level variables are all static variables on the level of individual politicians. Example data can be found here (.xlsx download).

Table 1: Summary of Politician Level Variables (POLI)

NAME	TYPE	VALUES/EXAMPLE[4]	SHORT DESCRIPTION	EFFORT
`pers_id`[6]	PriID	[country] _[first_last_name] _[first_first_name] _[year_of_birth] CH_Abate _Fabio_1966 DE_Mueller _Johann_1956dec NL_Mark _Rutte_1956dec-1	**parliamentarians individual identification code:** a combination of `country_abb`, the `last_name`, the"first_name" and the `birth_date`	COMP
`id_[country]_parl`[7]	ID	2739 = Gobbi, Norman	**paliamentsdienst CH identification number:** identification number, as used by the CH Paliamentsdienst	CWA
`last_name`	String	Mueller deCourten Mutlu-Blum	**parliamentarian' last name:** contains the last name. With 'von' e.t.c. connected (space removed), double last names hyphenated and special characters replaced.	COMP
`first_name`	String	"Jules-Henri" "Elisabeth" "Johann-Ulrich-Werner" "Heinrich-E."	**parliamentarian' first name:** contains the first name. With double first names hyphenated and special characters replaced.	COMP
`other_name`	String	Hans Muller	**parliamentarian' alternative name or alias:** contains an (array of) - first or last - name(s) or alias(ses). For matching purposes.	CWA
`gender`	Factor	m = male f = female nb = non-binary tm = trans male tf = trans female	**parliamentarian' gender**	COMP
`birth_date`	Date	16apr1897 29may1930	**parliamentarian' birth date**	COMP
`death_date`	Date	16apr1897 29may1930	**date of death of a parliamentarian** empty if still alive and NA when unknown	CWA

[6]To ensure compatibility of data-sets and make usage and merging of our data set easier we will add several id variables. The first one (`pers_id`) will be constructed by us, consists of the country, last name, first name, and birth year, and is unique to parliamentarians across countries and levels. This "naturally occuring" or "information-based" ID format allows consistent IDS to be constructed across a variety of sources while minimizing the need to look-up IDS in the already collected data.

[7]We will add all important id variables (often numbers) that are used by important data sets and/or by the source data set. These additional id variables are mostly unique to a country and/or level and are named by the following principle: first "id", then a country abbreviation (see variable "country_abb") and then a name or names that indicate the usage or source, all added by underscores.

Table 1: Summary of Politician Level Data-frame (Continued)

NAME	TYPE	VALUES/EXAMPLE[5]	SHORT DESCRIPTION	EFFORT
`birth_place_raw`	String	"Berlin" "Zuerich" "Baden-Wuerttemberg"	**place of birth:** place of birth of the parliamentarian	CWA
`citizenship`	String	DE DE; SK CH	**parliamentarians citizenship(s):** Array of - citizenship(s) as country abbreviation(s)	CWA
`citizenship_canton`	String	"ZH"	**Swiss canton of citizenship:** Switzerland-specific variable. Array of - citizenship(s) as canton abbreviation(s). Contains all citizensships over the entire lifetime.	CWA
`hometown_raw`	String	"Fehraltorf"	**Hometown of a person:** Switzerland-specific string variable. Name (and description) of place/city/municipality that is the candidates/parliamentarians home town.	CWA
`educ_age`	Integer	0 - 100	**age at completion of education:** ratio variable indicating how old a politician was when he/she finished her last degree	CWA
`educ_raw`	String	School 'The Bear' 1932	**education:** raw text string field with all available educational information	COMP
`title_raw`	String	Prof. Dr.	**academic title:** Academic title(s), if any.[8]	COMP
`military_raw`	String	Major Appointé	**military rank:** raw string information about the military career	CWA
`twitter_screen_name`	String	Petra_Sitte_MdB KathyR111	**twitter screen name:** raw string information about the twitter screen name of parliamentarians, this screen name can be changed	CWA
`twitter_id`	Integer	17535941 2179010672	**twitter id :** this id refers to the specific account and can not be changed	CWA
`facebook_username`	String	john.doe janedoe123 100009711991629	**facebook username :** this is a unique identifier for users on Facebook, which is sometimes numerical has sometimes been changed by the user. For latest version, query again from wikidata when possible.	CWA

pers_id parliamentarians individual primary identification code: Identification code, individual to each parliamentarian across levels. The code consists of the country abbreviation (as specified in variable `country_abb`), the last name (as specified in variable `last_name`). The first name (as specified in `first_name`) and the year of birth (as specified by the last four digits of variable `birth_date`) connected by underscores. In case this id is not unique to one person, the month of birth (as specified in the middle of the `birth_date` variable string) is appended by underscore. If the month of birth is not known or does still not uniquely identify the person, an additional index -'x' is appended (additionally to the month of birth) by hyphen. If a birth-year is unknown this the part of this identification string will be written as '9999'. In case the first name or the last name are unknown, this the part of this identification string will be written as 'XXX'. As soon as the

[8] In case it is known that the MP has no academic title, this variable is set to "none".

information becomes know this placeholder is replaced with the actual value (name or year). Alternative IDs and how they correspond to the main ID can be found in the appendix.

last_name & first_name For the cleaning of names the following rules apply:

- Names always start with a capital letter
- an umlaut (ö,ä,ü) is written as "oe", "ae", "ue"
- a "ß" will be written as "ss"
- all accents or similar (é, ã, ê, š, ğ, ç, ÿ) are left out, instead just the basic letter is written (e, a, e, s, g, c, y).
- if names contains prepositions like 'Von ' or ' Van der' then the space between these prepositions (name-particle) and the actual lastname are removed and capitals are replaced. So, 'Von Liebig' is spelled as 'vonLiebig' and 'Van Der Maden' is spelled as 'vanderMaden'.
- if the names contain multiple parts (e.g. first_name Jean Luc) then the spaced are replaced with hypens (e.g. Jean-Luc)
- Additional non-name particles such as junior are included after underscores (Carl_jun).

last_name parliamentarians last name: The last name(s) of the parliamentarian, following the cleanup rules specified above.

first_name parliamentarians first name: The first name(s) of the parliamentarian, following the cleanup rules specified above.

other_name parliamentarians alternative names or alias: Contains an (array of) - first or last - name(s) or alias(ses) in case there is or was another first or last name used than mentioned in "first_name" or `last_name`. This might be due to name change, a maiden name, or a commonly used shortage or alias.

gender parliamentarians gender: Gives the gender of the parliamentarian in string format. With 'm' signifying male and 'f' signifying female.

birth_date parliamentarians birth date: Mentions the date of birth of a parliamentarian consisting of the two-digit day, the month (first three letters, small, of the English name of the month) and the four-digit year of birth. If only the month is known the the days are dropped and of only the years are know the day and month letters are dropped.

death_date parliamentarians death date: Mentions the date of death of a parliamentarian consisting of the two-digit day, the month (first three letters, small, of the English name of the month) and the four-digit year of death. Is empty if still alive at the moment of data-collection and NA when unknown.

Effort: collect for those who died 'in office' and limit to MPs.

citizenship parliamentarians citizenship(s): Mentions the parliamentarians - array of - citizenship(s) by mentioning the abbreviation of country(s) he/she is citizen of. The used abbreviations follow the ISO "ALPHA-2" code. (See: `http://www.nationsonline.org/oneworld/country_code_list.htm`)

Citizenship_Canton Swiss canton of citizenship: This variable is Switzerland-specific and contains the abbreviation of the canton(s) in which a parliamentarian holds municipal citizenship(s). If several municipal citizenships exist, the abbreviations are separated by semicolon. Municipal citizenship is inherited from the parents or obtained by way of naturalisation. The same abbreviation is used as in variable "region_abb"

hometown_raw hometown of a person: Switzerland-specific string variable. Name (and description) of place/city/municipality that is the candidates/parliamentarians home town.

educ_age Politician its age when finished highest degree (with Diploma).

educ_raw Is a text (string) field containing all available information (in the original fora and language) on the education of a person.

military_string String variable with all information available on the military rank of the politician (in the original form and language).

twitter_screen_name_string raw string information about the twitter screen name of parliamentarians, this screen name can be changed.

twitter_id_integer this id refers to the specific account and can not be changeds.

5 Parliamentary Episode Data Frame (PARE)

The parliamentary episode data-frame contains information that varies across parliamentary episodes. This data-frame connects the parliament (PARL) data-frame with the politician (POLI) data-frame. It contains the crucial information of what politician was in what parliament and stores additional information that is typically available with this resolution. The data-frame does not only contain elected politicians and the parliament they were member of, but also candidates. Candidates are associated to the parliament they ran for. Several time-varying variables are collected for which we know they vary across episodes but on which we do not have more detailed time-stamped information. For example, the number of children. Naturally, this would also be the data-level where a lot of parliamentary behavior variables end up in.

Table 2: Summary of the Parliamentary Episode Level Variables (PARE)

NAME	TYPE	VALUES/EXAMPLE[9]	SHORT DESCRIPTION	EFFORT
`parl_episode_id`	PriID	CH_Abate _Fabio_1966 __CH_NT-NR_1946	**parliamentary episode ID** combination of pers_id and parliament_id	COMP
`pers_id`	String	CH_Abate _Fabio_1966	See above. A politician is considered a part of a specific parliament as soon as he/she has spend at least one day in the parliament.	COMP
`parliament_id`	String	[country]_[level] -[abbr]_firstyear NL_NT_1946	**parliament identifier:** See the specification below.	COMP
`member_ofthisparliament_atsomepoint`	String	"yes" "no"	**member of this parliament at some point:** dummy indicating, whether someone wa in the current parliament or not	
`residence_place_raw`	String	"Mecklenburg" "Zuerich" "Essen"	**place of residence raw:** all available information on the place a parliamentarian lives in	CWA
`residence_address`	String	"Bismarkstrasse 57, 10627 Berlin"	**residence_address:** complete residential address of the parliamentarian, only full address otherwise "residence_place_raw" is used	CWE
`residence_address_longi`	String	"Bismarkstrasse 57, 10627 Berlin"	longitude of the residence address	CWE
`residence_address_lati`	String	"Bismarkstrasse 57, 10627 Berlin"	latitude of the residence address	CWE
`family_status_raw`	String	'Widowed'	**any (other) family information**, in a raw-text format	CWA
`married`	Binary	1 = 'married' 0 = 'not married'	**parliamentarian its marital status:**, including registered partnerships	CWA
`nr_children`	Integer	0:15	**number of children**	CWA

[9]See text below table for more detailed variable specifications

parl_episode_id **unique identifier of the parliament**. Combination of the ...

birth_place_raw **place of birth:** Mentions the place of birth of the parliamentarian. The "place" that is mentioned may be very specific (e.g. street, city) to very broad (e.g. district or state). Which one is mentioned depends on the information available. In case more than one information/level of detail is available, the information is appended to the existing information by a semi-colon. As for the spelling of the names, the same rules apply as for the last_name_full (see "last_name_full"). In case the place has different names in different languages and these are available, the name in the local language is chosen.

residence_place_raw **place of residence:** Mentions the place a parliamentarian lives in. The "place" that is mentioned may be very specific (e.g. city) to very broad (e.g. district or state). Which one is mentioned depends on the information available. In case more than one information/level of detail is available, the information is appended to the existing information by semicolon. In case we have the full address of residence (i.e. street name, house number and city with postcode), the information is recorded in variable "residence_address". As for the spelling of the place names, the same rules apply as for the last_name_full (see "last_name_full"). In case different place names exist in different languages and are available, the name in the local language is chosen. (see variable "birth_place")

residence_address **residence_address:** Complete residential address of the parliamentarian. Only full addresses are mentioned (otherwise use the variable "residence_place_raw"). The house number is separated from the street name by a blanc space, the city name and code are separated from the street name and house number by a comma. As for the spelling of the street and city names, the same rules apply as for the last_name_full (see "last_name_full"). In case different names exist in different languages and are available, the name in the local language is chosen (see variable "birth_place").

married **parliamentarian was married at entry:** Dummy variable capturing whether a parliamentarian was married (or in a registered partnership). Time-varying. If the status is not known, then the status during the last parliamentary term is used.

family_status_raw **family status raw:** the raw - in own words - description of one' relationship status. If the status is known, but not when then the last parliamentary term is used.

nr_children **number of children:** Numeric variable that captures the number of children a parliamentarian has. Time-varying. If the status is not known, then the status during last parliamentary term is used. In the case of conjoined twins, the number of brains are counted.

6 Resume Entries Data Frame (RESE)

What are resume entries? In a nutshell, resume entries refer to the multitude of experiences, activities, and positions that individuals have held throughout their careers. It answers the question who held what political job when. This data-frame contains a long list of all functions (political- and non-political jobs and side-functions) that MPs held throughout their (political) career. Functions are defined as *all position politicians can hold*. This entails paid and unpaid functions as well as full-time and part-time ones. It can be as big as being a prime-minister and as a small as a short voluntary activity for a local sport-club. Everything that entails an activity that a politician considers worth reporting on their resume is included. Each politician occurs multiple times in this data. A politician that held a total of 20 different (political) functions throughout her career will occupy 20 lines in this data-frame. A complete resume entry contains reference to the role somebody played in an organization, the name of the organization, where in the organization this individual played this role and a start and end-date.

A typical resume entry as used in biographies and CVs contains a description of the activity as well as a start and an end date.

The RESE therefore collects all resume entry variables and variables directly based on or related to these entries.

Table 3: Summary of the Resume Entry Variables (RESE)

NAME	TYPE	VALUES/EXAMPLE[10]	SHORT DESCRIPTION	EFFORT

[10]See text below table for more detailed variable specifications

res_entry_id	String	[country_abb] _[first_last_name] _[first_first_name] _[year_of_birth] __[number] CH_Abate _Fabio_1966__01 DE_Mueller _Johann_1956_dec__06	**parliamentarians' resume entry code:** a combination of "pers_id" and an index for the number of the entry that counts from 01 to ..	COMP
pers_id	String	CH_Abate _Fabio_1966 DE_Mueller _Johann_1956_dec	**parliamentarians' individual identification code:** a combination of country_abb, the last_name, the first_name and the birth_date	COMP
res_entry_type	Integer	pol = Political prof = Professional educ = Educational iden = Occupational Identity oth = Other	**parliamentarians' resume entry type:** Categorical variable that defines the type of resume entry: Entries can refer to political (1), professional (2), or educational (3) sequences.	COMP
res_entry_start	Date	29apr1976 aug1953 2012	**resume entry start date:** Variable that captures when a specific resume entry started.	COMP
res_entry_end	Date	29apr1976 aug1953 2012	**resume entry start date:** Variable that captures when a specific resume entry ended.	COMP
res_entry_at	Date	29apr1976;aug1953 2012	**array of at date** (array of) dates indicating that this postion was held at this point in time.	COMP
res_entry_raw	String	Conseiller communal de 1970 à 1978.	**raw entry:** The resume entry as is taken from the source.	COMP
res_entry_source	String	Parlement.CH CH Yearbooks	**raw entry:** The resume entry as is taken from the source. See appendix XX for an overview of used RESE sources.	COMP
res_entry_index	Integer	01 02	**raw entry:** The position of the entry in the source. e.g. 1 means first mentioned. 2 means second etc	COMP
political_function	String	[PrimaryCode1]_ [PrimaryCode2]- [SecondaryCode2]_ [PrimaryCode3]- [SecondaryCode3]_ [PrimaryCode3]_ [PrimaryCode4]_ [PrimaryCode5]_ RE_PA-MA_T1_300_09 NT_LE_T2-CO_1500_03 MU_EX_T3_900_03	**political function:** Code assigned to the raw entry depending if it is of type "political".	COMP

pf_geolevel	String	[PrimaryCode1] RE NT MU	**geographical level:** Code assigned to level in the multi-level political system (see 6.1.2)	COMP
pf_instdomain	String	[PrimaryCode2]- [SecondaryCode2] [PrimaryCode2] PA-MA LE EX	**institutional domain** Code assigned to the raw entry according to institutional domain, e.g. party or legislative e.t.c. (see 6.1.2)	COMP
pf_orglevel	String	[PrimaryCode3]- [SecondaryCode3] [PrimaryCode3] T1 T2-CO T3	**tier in the organizational hierarchy:** Code assigned to the raw entry according to general location of the position in the organizational hierarchy. (see 6.1.2)	COMP
pf_policy_area	String	[PrimaryCode4] 0300 1500 0900	**related policy area (CAP)** Code assigned to the raw entry according to adjusted version of the Comparative Agenda's Project policy areas (see 6.1.2)	COMP
pf_position	String	[PrimaryCode5]_ 09 03 03	**Type of position** Code assigned to the raw entry according to type of position, e.g. regular member or chair e.t.c. (see 6.1.2)	COMP
parliament_id('s)	String	[country_abb]_[level]_[year] NL_NT_1946 DE_NT_2009 CH_NT_1971 [country_abb]_[level] -[reg_abb]_[year] CH_RE-BS_1997 DE_RE-BW_1956	**parliament identifier:** Unique code for different parliaments, both across countries/levels and over time. Used if political_function refers to a certain parliament(s). Is an array seperated by ' ; ' when there are multiple parliaments	COMP
prof_field	Integer	100:2300	**professional field:** Professional field codes based on the CAP / U.S. Standard Industrial Classification (cf. pf_policy_area of the political functions primary codes) assigned to the raw entry depending if it is of type "professional".	
educ_level_isced2011	Integer	1:8	**educational level:** ISCED level code assigned to the raw entry depending if it is of type "eductational".	COMP
educ_field_isced2013	Integer	00:10	**educational field:** ISECD field code assigned to the raw entry depending if it is of type "eductational".	COMP

occ_identity_raw	string	e.g. lawyer, student, teacher, businessman	**occupational identity:** mentions what individuals identify as their occupation or profession at a certain point in time, in original wording.	CWA
occ_identity_code_DE	integer	1:1000	**occupational identity code:** variable mentions the occupational identity of an individual at a certain time as classified by the German electoral bureau or statistical office	CWA

pers_id: Unique identification code for different politicians. See POLI for details on page 4.

res_entry_id: Identification code for resume entries, individual to each parliamentarian and entry. The code consists of "pers_id" and a number per entry. Numbers below 10 are written with an additional zero (e.g. 05) and connected to "pers_id" by two underscores. The order of numbers is arbitrary but roughly corresponds to the order of appearance in the source.

res_entry_type: Categorisation of resume entries. This variable captures the type to which a certain entry can be attributed to. Entries can refer to either political (pol), professional (prof), educational (educ), an occupational identity[11] (iden), or other (oth) biographical episodes. Political functions (pol) are functions that are directly or indirectly aimed at creating and shaping policy.

res_entry_start: This variable gives the start date of a particular resume entry. These date entries consist of the two-digit day, the month (first three letters, small, of the English name of the month) and the four-digit year of birth. In case of missing information on days or month, shortened versions of the date are used.

res_entry_end: This variable gives the end date of a particular resume entry. These date entries consist of the two-digit day, the month (first three letters, small, of the English name of the month) and the four-digit year of birth. In case of missing information on days or month, shortened versions of the date are used.

res_entry_raw: This variable gives the raw resume entry string. Generally this will be the text as it has been extracted from the original source. However, completeness is more important then sticking to the original source text. For example 'bis 21 oktober 2014' would be replaced with 'Mitglied des bundesrat von 12. Jun 2000 bus 21. oktober' if this concerns an RESE entry which reflects a position in the German Bundestag (NT_LE_T3_NA_01).

political_function_code, pf_geolevel, pf_instdomain, pf_orglevel, pf_policy_area, pf_position Multi-level code that captures various facets of resume entries that are categorized as political under "res_entry_type". A thorough presentation of the entire political functions coding scheme is provided in the subsequent chapter!

parliament_id: Unique identification code for different parliaments. See PARL for details on page 7. It is added whenever ''political_function'' refers to a function connected to national and regional parliaments. Is an array separated by ' ; ' when there are multiple parliaments.

prof_field: This variable captures the professional area of resume entries that were categorised as professional under "res_entry_type". Professional field codes are based on the CAP / U.S. Standard Industrial Classification. The coding scheme corresponds to the policy areas (level 4) of the political functions primary codes. An overview of the codes is provided in the subsequent subchapter.

isco08 This variable takes a value when the entry of concern is of res_entry_type 'prof'. Professional jobs and captures the professional area of resume entries that were categorized as professional ('prof') or occupational identity ('iden') under "res_entry_type". This functions are coded into the International Standard Classification of Occupations (ISCO). This code has four levels of detail or 'digits'. All four digits are used. See https://www.ilo.org/public/english/bureau/stat/isco/isco08/ for details. The main categories of ISCO-08 are:

[11] the job that an MP mentions when asked by the election council to summarize their main occupation so it can be mentioned on the election list

- ISCO-08 major group 1 – Managers
- ISCO-08 major group 2 – Professional
- ISCO-08 major group 3 – Technicians and associate professionals
- ISCO-08 major group 4 – Clerical support workers
- ISCO-08 major group 5 – Service and sales workers
- ISCO-08 major group 6 – Skilled agricultural, forestry and fishery workers
- ISCO-08 major group 7 – Craft and related trades workers
- ISCO-08 major group 8 – Plant and machine operators, and assemblers
- ISCO-08 major group 9 – Elementary occupations
- ISCO-08 major group 10 – Armed forces occupations

educ_level_isced2011: Ordinal variable capturing the level of education of particular resume entries. Based on the ISCED 2011 international coding standard, only first digit is coded. see `http://www.uis.unesco.org/Education/Documents/isced-2011-en.pdf` for the codes.

isced2011 levels

- ISCED level 0 – Early childhood education
- ISCED level 1 – Primary education
- ISCED level 2 – Lower secondary education
- ISCED level 3 – Upper secondary education
- ISCED level 4 – Post-secondary non-tertiary education
- ISCED level 5 – Short-cycle tertiary education
- ISCED level 6 – Bachelor's or equivalent level
- ISCED level 7 – Master's or equivalent level
- ISCED level 8 – Doctoral or equivalent level

educ_field_isced2013: The isced field classification see `http://www.uis.unesco.org/Education/Documents/isced-fields-of-education-training-2013.pdf` only the first digit is coded.

isced2013 fields

- ISCED Field 00 - Generic programmes and qualifications
- ISCED Field 01 - Education
- ISCED Field 02 - Arts and humanities
- ISCED Field 03 - Social sciences, journalism and information
- ISCED Field 04 - Business, administration and law
- ISCED Field 05 - Natural sciences, mathematics and statistics
- ISCED Field 06 - Information and Communication Technologies
- ISCED Field 07 - Engineering, manufacturing and construction
- ISCED Field 08 - Agriculture, forestry, fisheries and veterinary
- ISCED Field 09 - Health and welfare
- ISCED Field 10 - Services

6.0.1 left- and right censoring of dates

In some cases we know for sure that a career episode **at least** lasted until (or started') at a certain moment in time. Such 'censored' dates are indicated in our data with '[[lcen]]' and '[[rcen]]' respectively. If we for example know that a politician was in parliament at least from 19 Jan 1985 onward then we would write this as '19jan1985[[lcen]]'. If we would check on 31th of December 2018 and found out she is still there then res_entry_end will be written as '31dec2018[[rcen]]'

6.1 The (political) functions coding scheme

To investigate political careers, we need to know who had what (political) job when. As such, a core part of our data concerns the resume entry (RESE) data-frame. This data-frame contains a long list of all functions (political- and non-political jobs and side-functions) that the politicians in our data held throughout their (political) career. Functions are defined as *all position politicians can hold*. This entails paid and unpaid functions as well as full-time and part-time ones. It can be as big as being a prime-minister and as a small as a short voluntary activity for a local sport-club. Everything that entails an activity that a politician considers worth reporting on their resume is included. Each politician occurs multiple times in this data. For example, a politician that held a total of 20 different (political) functions throughout her career will occupy 20 lines in this data-frame.

As the focus is on political career the majority of the functions in the RESE data-frame are 'political functions' (defined as ' functions that are directly or indirectly aimed at creating and shaping policy'). This definition is quite all encompassing on purpose; it for example also includes being the CEO of a largere company and part of her job is as well to shape polocy relevant for the company its profits. There are however also several none-political functions in the data. These are marked as 'prof' in the res_entry_type column. There are also several educational episodes (periods of their lives in which MPs where not working but following educational programs instead) in the data. These are marked as 'educ'. The classification scheme below is focused is our own new classification of political functions. Professional jobs are code into the excisting; ISCO-08 coding scheme (see above) and educational entries into the excisting ISCED field and level classifications (see above).

6.1.1 Principles for Classification of (political) functions

The proposed coding scheme for the classification of political functions consists of two parts. The first one is labelled **primary political function coding** while the second one can be referred to as **secondary political function coding**.

- **Primary function coding** entails information that can be coded for all political functions. For example, we should be able to code all political functions with regard to the geographical level (pf_geolevel, e.g. municipal / regional / national) they are located on in the multi-level hierarchy.
- **Secondary function coding** refers to information that is specific to subsets of the primary coding. With the primary coding scheme for instance, we are unable to distinguish whether some-one's party function was in a youth party (YO) or the main party (MA). Secondary codings are therefore introduced whenever an even more detailed distinction is of interest for particular research questions.

6.1.2 Primary Political Function Coding

pf_geolevel: Position of organization in the multi-level hierarchy Coding the position in the multi-level structure depends on the institutional affiliation of the political function (thus often answering the question *who pays you?*). Example: Working as an ambassador to the EU will therefore be coded as national level (NT).

Effort: collect until complete.

- IN - International level
- EU - EU level
- NT - National level
- SR - Superregional level (regional, super) (below national, above regional level)
- RE - Regional level (NL: provinces; DE: Bundesländer; CH: Cantons)
- SD - Superdistrictual level (districutal, super) (below regional, above district level)1
- DI - District level (CH: Ämter, Bezirke; DE: Kreise)
- SM - Supermunicipal level (municipal, super) (below district, above municipal level; CH: Kreise)1
- MU - Municipal level
- BM - Below-Municipal level (e.g. neighbohrhood counsel)

pf_instdomain: Organisational type *Effort: collect until complete.*

- LE - Legislative
- EX - Executive
- JU - Judiciary
- AD - Administration (including dependent agencies)
- SI - Semi-independent agencies
- PA - Party
- IG - Interest Group
- OT - Other

Notes with institutional domain coding:

- if you are a CEO of a large company and it is part of your job to influence policy you would be coded as organisational type: IG - Interest Group
- Semi-independent agencies are public bodies that resemble administrations in their behavior. They have goals defined by the state such as financial regulation (e.g. FINMA), drug licencing (SWISSMEDIC), or nuclear power safety (ENSI). Furthermore, they can also include state-owned research institutions (e.g. AGROSCOPE). They do not pursue financial goals. Publically owned companies (e.g. Swiss Federal Railways, Swisscom, Swiss Post) are therefore NOT semi-independent agencies but IGs.

pf_orglevel: Internal structure of organisational types, up to 3 tiers *Effort: collect when available.*

- T1 Top-tier
- T2 Mid-tier
- T3 Low-tier

How to code the tiers The terms *top-tier, mid-tier, low-tier* would be used as overarching descriptions. For the specific combinations of pf_geolevel and pf_instdomain codes (e.g. B-A EU-level legislative), the name of the actual corresponding tier name would be used e.g. *parliamentary presidency* instead of *top-tier*.

Institutions have highly differing structures that do not only vary across but even within countries. The use of three tiers is necessary to enable comparisons between institutions within and across countries. The tiers offer a meaningful simplification of institutional structure in terms of policy competences. Almost all institutions depict a similar structure: the highest and leading instance holding the broadest/deepest competences (T1), the unit right below the highest instance which holds specialised competences and reports to or is controlled by the highest instance (T2), and the rest of the organizational hierarchy which covers the remaining less powerful and less prestigious functions (T3).

It should be noted that the tiers do not (necessarily) refer to a division of labor though. Support bodies tasked with administrative and implementing functions (e.g. secretariats affiliated with leadership or steering bodies) are coded on the same tier as their principals'. However, bodies to which policy competences are delegated are coded on a separate level.

The table below serves to assist with the classification process. We can see here that almost all institutions under consideration either follow such a three tier structure or the structure can be meaningfully divided into three tiers.

Code	Description	Categories	Examples
IN-...	International level	...	...
EU_LE	EU-level legislative	T1 Parliamentary presidency T2 Committee / delegation / party group T3 Assembly	T1 "Schulz' cabinet" T2 "Committee on Legal Affairs" T3 "MEP"
EU_EX	EU-level executive	T1 Commission presidency T2 College of Commissioners T3 Secretariat-General	T1 "First Vice-President" T2 "Commissioner for Energy Union" T3 "Unit B3 Ethics"
EU_JU	EU-level judiciary	DNC T1 European Court of Justice T1 General Court T1 Specialised court	T1 "Third Chamber" T1 "Registry" T1 "EU Civil Service Tribunal"
EU_AD	EU-level administration	T1 Directorate-General T1 Directorate / department T1 Unit	T1 "DG Communication" T1 "Directorate Resources" T1 "Energy policy coordination unit"
EU_SI	EU-level semi-independent agencies	T1 Decision-making body T2 Department T3 Unit	T1 "ECB's executive board" T2 "Department of air and climate change" T3 "Nutrition unit"
EU_PA	EU-level party	T1 Party leadership body T2 Steering body(commission, delegation, council, 'bureau', non-administrative secretariat)	T1 "Presidency" T2 "PES Council"
EU_IG	EU-level interest group	T1 Leadership body T2 Steering body / department (committee, delegation, council, bureau, secretariat, office)	T1 "Board", "presidential council", "Directorate-General" T2 "General committee", "department", "committee", "working groups", "Brussels office"
EU_OT	Other EU level entities	...	...

NT_LE	National parliament	T1 Parliamentary presidency T2 Committee / delegation / party group T3 Assembly	T1 “Nationalratspräsidentin” T2 “Commissie voor de Rijksuitgaven” T3 “MP”
NT_EX	National government	T1 Government Leadership T2 Ministerial Level T3 Junior Ministerial Level	T1 “Bundeskanzlerin” T2 “Bundesminister des Innern” T3 “Staatssecretaris van Onderwijs”
NT_JU	National judiciary	DNC T1 Highest court T2 Court of Appeal / 2nd highest court T3 -	T1 “Raad van State“; Hoge Raad“, “Bundesfinanzhof“, “Bundesgericht” T2 “Bundesverwaltungsgericht”
NT_AD	National administration	T1 Department T2 Subdepartment T3 Unit	T1 “Secretariaat-generaal“, “Generalsekretariat”, “Staatssekretariat” T2 “Bundesamt (CH)”, Abteilung (DE)“, Directie-Generaal“ T3 Referat“, Stab“, Unterabteilung“, Amt“ Directie“
NT_SI	National semi-independent agencies	T1 Decision-making body T2 Department T3 Unit	T1 “Bestuur”, “Direktion (CH)” T2 “Directie“, “Abteilung“ T3 “Team (NL)”, Referat (DE)“
NT_PA	National party	T1 Party leadership body T2 Steering body (commission, delegation, council, bureau, secretariat)	T1 “Partijbestuur“, Parteileitung“ T2 Partijbureau“, “Generalsekretariat“, Bundesgeschäftsstelle“, Bundesparteigericht“
NT_IG	National interest group	T1 Leadership body T2 Steering body / department (committee, delegation, council, bureau, secretariat,	T1 “Board”, “presidential council”, “Directorate-General” T2 “General committee”, “department”, “committee”, “working groups”, “Brussels office”

NT_OT	Other national entities	...	e.g. intercantonal concordats
SR_	Superregional level	...	CH: 11 Kreispostdirektionen (bis 1997), 3 Kreise der Bundesbahnen (bis 1998), 4 Kreise der Zollverwaltung (bis 2000), 14 Kreise der Alkoholverwaltung NL: 5 gerechtshoven (courts of appeal in five regions)
RE_LE	Regional parliament	T1 Presidency of the regional parliament T2 Committee / delegation / faction T3 Assembly (of the regional parliament)	T1 "Parlamentspraesident", "Kantonsratspraesident", "Grossratspraesidentin", Landtagspraesident" T2 "Ausschuss fuer Laendlichen Raum und Verbraucherschutz", "Presidium", "Fractievoorzittersoverleg" T3 "Provinciale Staten", Landtag", Kantonsrat", Grossrat"
RE_EX	Regional government	T1 Regional Government Leadership T2 Regional Ministerial Level T3 Regional Junior Ministerial Level	T1 "Ministerpraesident", "Comissaris van de Koning", "Regierungspraesident", Staatsratspraesidentin", T2 "Gedeputeerde Staten", "Regierungsrat", "Staatsrat", Innenminister" T3 ...
RE-JU	Regional judiciary	T1 Highest Court T2 Court of Appeal/ 2nd highest court T3 -	T1 "Oberlandesgericht Karlsruhe" (highest ordinary court) "Rechtbank", "Kantonsgericht" T2 "Amtsgericht Achern"
RE-AD	Regional administration	T1 Department T2 Subdepartment T3 Unit	T1 "Direktion" T2 "Kantonsamt (CH)", Abteilung (DE)", Directie-Generaal" T3 Referat", Stab", Unterabteilung", Amt" Directie"
RE-SI	Regional semi-independent agencies	T1 Decision-making body T2 Department T3 Unit	T1 "Bestuur", "Direktion (CH)" T2 "Directie", "Abteilung" T3 "Team (NL)", Referat (DE)"

RE-PA	Regional party	T1 Party leadership body T2 Steering body (commission, delegation, council, bureau, secretariat)	T1 “Partijbestuur“, Landesvorstand“, Landespraesidium“, Kantonalparteileitung“ T2 Partijbureau“, Landesgeschäftsstelle“, Landesparteigericht“, “Parteisekretariat”
RE_IG	Regional interest group	T1 Leadership body T2 Steering body / department (committee, delegation, council, bureau, secretariat,	T1 “Board”, “presidential council”, “Directorate-General” T2 “General committee”, “department”, “committee”, “working groups”, “Bremen office”
RE_OT	Other regional entities	...	e.g. private-public partnerships with political functions
SD_	Superdistrictual level	...	NL: 19 rechtbanken (separated into judicial arrondisements), waterschapen CH: Zivilkreisgerichte
DI_LE	District parliament	T1 Presidency of the district parliament T2 Committee / delegation / faction T3 Assembly (of the district parliament)	T1 “Kreistagspraesident”, “Landrat” T2 “Ausschuss fuer Migration” T3 “Kreistagsabgeordnete”
DI-EX	District government	T1 District government leadership T2 District government membership T3 District government steering body	T1 “Landrat” “Landrat”, “Amtsstatthalter”, “Regierungsstatthalter“, Bürgerworthalter“ T2 “Kreisvorstand” T3 ...
DI-JU	District judiciary	dnc T1 District Level Court	T1 “Amtsgericht Achern”, “Kreisgericht“, Amtsgericht“, Bezirksgericht“, Kantongerecht“
DI-AD	District administration	T1 Department T2 Subdepartment T3 Unit	T1 “Kreisamt (CH)”, “Dezernat (DE)”, “Abteilung (DE)”, T2 “Referat”, “Stab” T3 “Unterabteilung”

DI-SI	District semi-independent agencies	T1 Decision-making body T2 Department T3 Unit	...
DI-PA	District level party	T1 Party leadership body T2 Steering body (commission, delegation, council, bureau, secretariat)	T1 “Partijbestuur“, Präsidium“, Kamerkringbestuur“, “Geschäftsleitung”, Bezirksvorstand“, Kreispraesidium“, Partijbureau“, Kreisgeschäftsstelle“, T2 Kamerkring“, Kamercentrale“, Provinciaal contact“
DI-IG	District level interest group	T1 Leadership body T2 Steering body / department (committee, delegation, council, bureau, secretariat,	T1 “Board”, “presidential council”, “Directorate-General” T2 “General committee”, “department”, “committee”, “working groups”, “Bremen office”
DI-OT	Other district level entities	...	... e.g. private-public partnerships with political functions
SM-	Supermunicipal level	...	CH: “Gemeindeverband”
MU-LE	Municipal parliament	T1 Presidency of the municipal parliament T2 Committee / delegation / faction T3 Assembly	T1 “Parlamentspraesident” T2 “Ausschuss” T3 Gemeindeparlament“, “Einwohnerrat”, “Stadtparlament”, “Gemeenteraad”
MU-EX	Municipal government	T1 Municipal government leadership T2 Municipal government membership T3 Municipal government steering body	T1 Gemeindepräsident“, Stadtverordnetenvorsteher“, Ratsvorsitzender“, Präsident der Bürgerschaft“, Bürgervorsteher“, (Ober-)Bürgermeister“, Burgemeester“ T2 Gemeinderat“, Stadtrat“, College van burgemeester en wethouders“ T3 “Welstand”, “SER”, “Einbürgerungskommission”
MU-JU	Municipal judiciary	DNC T1 (see District Judiciary)	...

MU-AD	Municipal administration	T1 Department T2 Subdepartment T3 Unit	T1 “Ressorts”, “Amt”, “Dezernat” T2 “Bereich” T3 “Sachbereich”
MU-SI	Municipal semi-independent agencies	T1 Decision-making body T2 Department T3 Unit	...
MU-PA	Municipal party	T1 Party leadership body T2 Steering body (commission, delegation, council, bureau, secretariat)	T1 “Partijbestuur“, Vorstand des Ortsverbandes“, “Geschäftsleitung“, Präsidium“, “Geschäftsleitung“ T2 Partijbureau“, Parteibüro“
MU-IG	Municipal interest group	T1 Leadership body T2 Steering body / department (committee, delegation, council, bureau, secretariat,	T1 “Board” “presidential council”, “Directorate-General” T2 “General committee”, “department”, “committee”, “working groups”
MU-OT	Other municipal entities	...	...

Table 4: Political functions pf_orglevel coding specifications

While this three-tier code condenses information, the more detailed information will not necessarily be lost. On the one hand, it is still part of the string variable that is used to code the tier; on the other hand specific secondary codes can be introduced. The secondary code (see 2 Secondary Coding) can be used to introduce more specificity whenever it is required. Take for example national level parliaments and say we are particularly interested in positions in national parliamentary bodies. The primary coding does not provide detailed information on that (see table 5)

Code	Description	Categories
NT-LE	National parliament	T1 - Parliamentary presidency T2 - Committee / party group / delegation T3 - Assembly

Table 5: Example of Three-tier Coding

The information on committees, delegations, and party groups is condensed into one single category (mid-tier). If the interest arises to study membership in these different bodies separately, the secondary code would capture that (See table 6)

Code	Description	Categories	Secondary code options for NT-LE-02
NT-LE	National parliament	T1 - Parliamentary presidency T2 - Committee / delegation / party group T3 - Assembly	CO: Committee FA: Faction (party group) DL: Parliamentary delegation

Table 6: Example of Secondary Coding

Beyond the coding of just NT-LE-02 for national parliamentary bodies, we can thus in the same step code this additional information.

pf_policy_area: 21 policy areas (taken from the Comparative Agendas Project and extended by areas of economic activities taken from the U.S. Standard Industrial Classification). Politicians are typically expected to hold a certain policy expertise. When applicable we code the policy areas expertise a function is related to. For example, a function for an climate change related interest group was coded as '1900 - environment' and if someone was a minister of education this would be marked as '0600 - education'. Functions can be related to multiple policy areas. Being the dean of an academic hospital for example would be marked as entailing expertise on both '0300 - health' and '0600 - education'). Information for policy area may not always be available or applicable (e.g. municipal councilors dealing with a wide range of policy issues). Whenever none of the subcategories (sub-areas) apply, the main topic (policy area) is to be coded.

-

Effort: collect when available.

Code	Policy Area	Sub-Area	Notes
(0)100	**Domestic Macroeconomic Issues**		
(0)200	**Civil Rights, Minority Issues, and Civil Liberties**		
(0)201		Ethnic Minority and Racial Group Discrimination	
(0)202		Gender Discrimination	Given particular research interests in gender issues, the aspect of sexual orientation discrimination has been dropped from the original CAP code.
(0)300	**Health**		All health related activities and entities are coded here. This includes aspects related to health facilities, insurance, pharmaceutical products, tobacco, health research, and the health care system in general.
(0)400	**Agriculture**		
(S0)401		Crop and livestock	Based on the (U.S.) Standard Industrial Classification SIC 0100 and 0200. This category covers aspects related to the production poultry , livestock, dairy, fruit, nuts, and crop plants.
(S0)402		Forestry	Based on the (U.S.) Standard Industrial Classification SIC 0800. If forestry is is related to Forest Protection, then it is coded as 700 Environment.

(0)403		Agricultural Primary Processing	This category is not based on any CAP or SIC code. It covers aspects related to primary processing of poultry , livestock, dairy, fruit, nuts, and crop plants. These are the steps of food processing directly after production. Wine and wineries are also coded here. This category is also given precedence over S1505 Manufacturing of Food and Kindred Products whenever a clear connection to agriculture is predominant in later food processing steps.
(0)404		Agricultural Trade, Retail, Services, and Supply	This category is not based on any CAP or SIC code. It covers all activities related to wholesale and retail trade of agricultural products, services and supply provided for / or by agriculturally focussed entities.
(0)405		General Agricultural Activities	This category is not based on any CAP or SIC code. It covers all activities related to agriculture that are not in anyway covered by the other agriculture categories. This includes for instance agricultural interest groups, farmers' newspapers and mountain farmers' relief organizations, agricultural disaster insurance, agricultural research institutions, animal welfare, pest control etc.
(0)406		Fisheries, Fishing, and Hunting	Based on CAP code 408.
(0)500	**Labour and Employment**		E.g. Functions in pension funds
(0)600	**Education**		
(0)700	**Environment**		
(0)800	**Energy**		
(0)900	**Immigration and Refugee Issues**		
1000	**Transportation**		Everything environment-related is coded under 700 Environment. Transportation that serves primarily a touristic purpose should be coded under S1515.
1200	**Law, Crime, and Family Issues**		
1300	**Social Welfare**		
1400	**Community Development and Housing Issues**		
(S)1401		Real Estate	Partially based on the (U.S.) Standard Industrial Classification SIC 7000.
1500	**Economic Activities and Domestic Commerce**		This category encompasses a wide array of economic activities which do not fall under the scope of the other 20 policy areas. When economic activities come with interactions with public officials, they are coded according to their economic area. Note that activities covered by other policy areas such as farming (400 Agriculture) or the production of pharmaceutical goods (300 Health) are coded in the corresponding policy areas.

(S)1501		Mining and Quarrying of Metalls and Nonmetallic Minerals, except Fuels	Based on the (U.S.) Standard Industrial Classification SIC 1000 and 1400. Energy resources (oil, gas, coal etc.) are coded under 800 Energy.
(S)1502		Construction, Engineering, Architecture, and Surveying	Based on the (U.S.) Standard Industrial Classification SIC 1500-1799 and 8710
(S)1503		Secondary Processing of Food and Kindred Products	Partially based on the (U.S.) Standard Industrial Classification SIC 2000. This category comprises secondary processing of food products. Entities active in this area (often) do not execute primary processing OR produce mainly for end customers. Hence, agricultural policies are mostly NOT the main concern of these actors active in this area. Non-grape based alcoholic beverages are coded here.
(S)1504		Manufacturing of Leather, Textile Products, and Apparel	Based on the (U.S.) Standard Industrial Classification SIC 2200, 2300, 3100.
(S)1505		Manufacturing of Lumber and Wood Products, incl. Furniture and Paper	Based on the (U.S.) Standard Industrial Classification SIC 2400, 2600, 3500. Activities related to unprocessed timber are coded under S402 Forestry.
(S)1506		Manufacturing of Chemicals and Allied Products	Based on the (U.S.) Standard Industrial Classification SIC 2800. Pharmaceutical companies are coded under 300 Health.
(S)1507		Manufacturing of Rubber and Miscellaneous Plastics Products	Based on the (U.S.) Standard Industrial Classification SIC 3000.
(S)1508		Manufacturing of Stone, Clay, Glass, and Concrete Products	Based on the (U.S.) Standard Industrial Classification SIC 3200.
(S)1509		Manufacturing of Primary Metal Products	Based on the (U.S.) Standard Industrial Classification SIC 3300.

(S)1510		Manufacturing of Machinery, Computer Equipment, Electronic and other Electrical Equipment, Measuring, Analysing, and Controlling Instruments, Photographic, Medical and Optical Goods; Watches and Clocks.	Based on the (U.S.) Standard Industrial Classification SIC 3400-3600, 3800. Equipment related to transportation (motor vehicles, aircrafts, railroad equipment, ships, bicycles , motorcycles etc.) are coded under 1000 Transportation. Manufacturing of equipment mainly for agricultural purposes / agricultural customers is coded under 404 Agricultural Trade, Retail Services, and Supply.
(S)1511		Wholesale Trade	Based on the (U.S.) Standard Industrial Classification SIC 5000-5199. Wholesale directly tied to agricultural organisations is coded under 404 Agricultural Trade, Retail Services, and Supply.
(S)1512		Retail Trade	Based on the (U.S.) Standard Industrial Classification SIC 5200-5999. This category includes activities related to General Merchandise Stores, Food Stores, Building Materials, Hardware, Garden Supply, and Mobile Home Dealers, Apparel and Accessory Stores, and Home Furniture, Furnishings, and Equipment Stores. Automotive Dealers and Gasoline Service Stations are coded under 1000 transportation. Retail directly tied to agricultural organisations is coded under 404 Agricultural Trade, Retail Services, and Supply.
(S)1513		Banks, Financial Institutions, and Non-Health / Non-Accident Insurances	Based on the (U.S.) Standard Industrial Classification SIC 6000-6200 and 6400. Also includes 8720 (Accounting, Auditing, and Bookkeeping Services) and 8740 (Management and Public Relations Services). Health and accident related insurances (6320) are coded under 300 Health.
(S)1514		Tourism, Hotels, and Restaurants	Based on the (U.S.) Standard Industrial Classification SIC 7000 and 5800
(S)1515		Other Services	Based on the (U.S.) Standard Industrial Classification SIC 7000-8999. This category includes a wide array of services including Personal and Business Services, Amusement and Recreation Services (if S1515 [Tourism, Hotels, and Restaurants] and 2300 [Cultural Policy Issues] are **not** applicable), and Social Services (e.g. child day care).
1516		Interbranch and Cross-Sectoral Activities	Economic activities that span multiple industries, sectors, and/or services where no clear emphasis on a particular area of business activity emerges. Examples include holding companies and general business associations.
1517		Legal Services	Legal practitioners (e.g. lawyers and notaries), not classified to any other industry, primarily engaged in providing legal and paralegal services (based on SIC 54119).
1600	**Defense**		

1700	**Space, Science, Technology, and Communications**		
(S)1701		Research, Development, and Testing	Based on the (U.S.) Standard Industrial Classification SIC 8730.
1702		Newspaper, Publishing, and Broadcast Industry Regulation	Dissemination of information related to any other policy area (e.g. publishing a farmers' newspaper) is coded in the corresponding policy area. Based on CAP code 1707.
1703		Computer Industry and Computer Security	Based on CAP code 1709.
1800	**Foreign Trade**		
1900	**International Affairs and Foreign Aid**		
2000	**Government Operations**		
2100	**Public Lands, Water Management, and Territorial Issues**		
2300	**Cultural Policy Issues**		Also includes sports, heritage, and history
2400	**General Ideological Issues**		Only applicable to interest groups. Captures general ideological stances (e.g. liberalism, conservatism, socialism) of organisations that do not compete for public office

Table 7: Political functions pf_policy_area coding specifications: policy area codes Note: The letter S in front of the nummerical code specifies newly created codes based on the U.S: SIC classification.

pf_position: occupational position within the hierarchical level (categorical scale!) *Effort: collect when available.*

The codes capture occupational position in a standardized form. Every piece of in-formation is assigned a numerical value (e.g. the president of an organisational unit is coded as leadership [01]).

Code	Category	Examples
01	Leadership	"Präsident", "Vorsitzende", "Teamleiterin", "Chair", 'Co-président", "Co-Direktor", "Gesellschafter"
02	Deputy leadership	"Vice-chair", "stv. Amtsleiter"
03	Ordinary worker / member	"Member of the council", "Mitarbeiter", "Erstatzmitglied"
04	Leading support staff[12] to leadership	"Head of the party bureau"
05	Support staff to leadership	"Personal assistant to the secretary-general"
06	Leading support staff	"Head assistant for MEP"
07	Support staff	"Assistant for MEP", "secretary"
08	Replacement member / worker	"Ersatzmitglied", "Stellvertretendes Mitglied"
09	Active (position not specified)	"Amt für Volksschulbildung 1974-9"

Table 8: pf_position: Occupational Position Codes

Distinguishing pf_tier and pf_position: A Short Note on the Difference Between Leaders and Members

The distinction between the internal structure of organisations and the occupational position can be a cause for confusion. Depending on whether a position is considered over the entirety of the organisation or merely the tier (T1, T2, or T3) at hand, the pf_position code may vary. However, only the second option is correct. Occupational positions are always only coded in relation to their tier.

To illuminate this issue more, let us consider an example: how do we code ministers? It is quite clear that ministers are located on tier T2 (i.e. on the ministerial level; _T2_). They are neither government leadership (tier 3; _T1_) nor located on the junior ministerial level (_T3_). The more difficult question is then, however, how to code their occupational position (pf_position). In relation to the head of government (_T1_NA_01), one might **incorrectly** assume that ministers are ordinary workers / members. They are not though. Rather, they are considered the leadership of their own tier (i.e. tier 2). Hence ministers are always coded as (_T2_01).

Figure 2 provides an overview of several political functions in the executive and what code they are assigned. The key element to this overview is that the distinction between leadership and other functions is always made at the tier at hand, not across tiers.

In analogy, the same applies to the legislative. Positions in parliament are always coded only with reference to the tier at hand. This might seem obvious for some positions and counterintuitive for others: The parliamentary president is naturally coded as the leader (01) of the highest level in parliament (T1) i.e. as _T1_NA_01. Similarly, members of committees, parliamentary party groups, and delegations are coded as members of their respective bodies (_T2_NA_03).

On every tier, the highest position is **always** coded as leadership. Confusion might arise when the distinction between leadership and ordinary members is concerned on the assembly level (T3). Members of parliament are always coded as leaders because they are the highest instance on their level. Leadership, in that sense, this does not only refer to exercising power but also relates to independence in making one's decisions.

Figure 3 provides an overview of several political functions in the legislative and what code they are assigned.

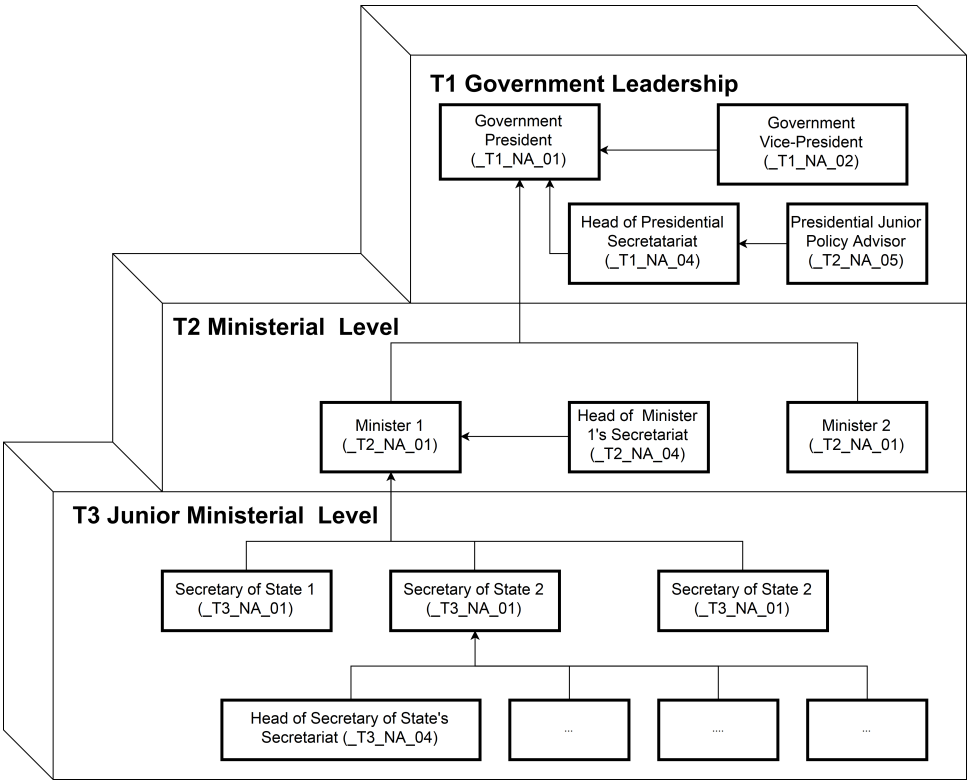

Figure 2: Overview of political functions in the executive and how they are coded on pf_tier and pf_position.

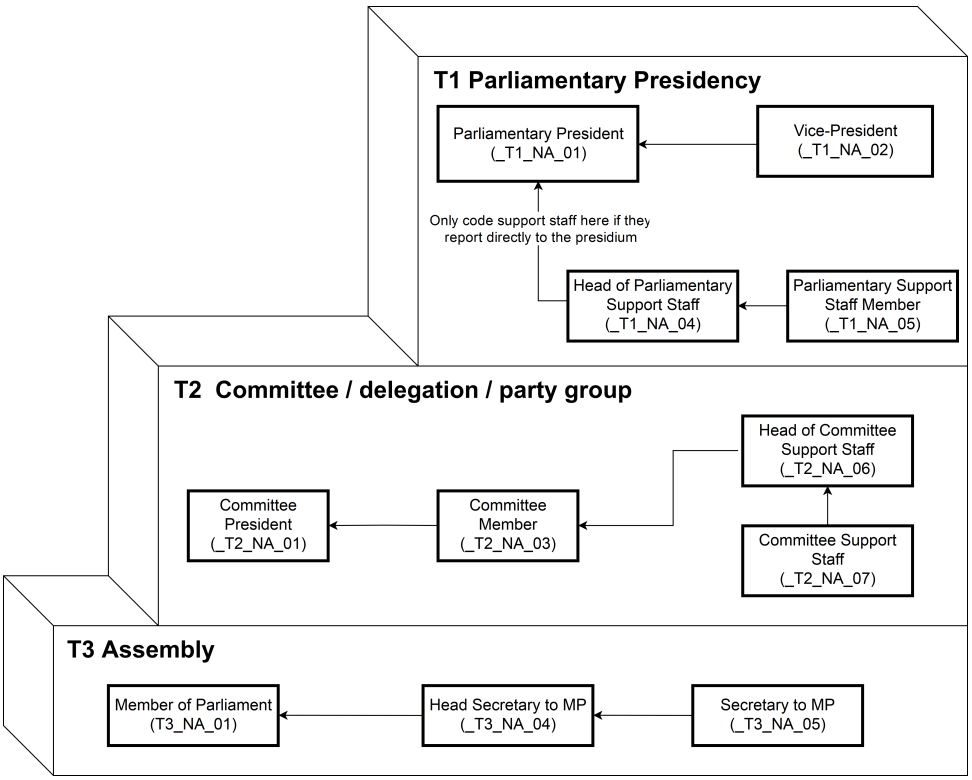

Figure 3: Overview of political functions in the legislative and how they are coded on pf_tier and pf_position.

Added
Nov 7, 2019 10:56 AM • You 524

6.1.3 Secondary Political Function Codes

Secondary coding's are research project (e.g. research question) specific 'even more detailed' distinction in the coding of political functions. We are happy to facilitate you with the coding of for additional secondary codes. Drop us a message at info@parliamentarycareersincomparison.org The secondary codes in the next paragraphs are already available.

Secondary Political Function Coding: Elected or Non-Elected Applies to: 'X'_LE_ 'X'_EX_: Whenever the organisational type of a political function is specified as either legislative (pf_instdomain = LE) or executive (pf_instdomain = EX) or judical (JU), then it will be specified if this functions was **elected** or **non-elected**.

Code	Category	Examples
PE	Probably[13] elected	"Minister", "MP"
NE	Non-Elected	"Assistant to the Minister", "Parliamentary Committee Secretary"

Table 9: Elected or Non-Elected Political Functions - **DNC**

Secondary Political Function Coding: Legislative steering body specification Detail on the second tier of all legislative (X_LE) positions is important. Hence a further distinction between the three types of steering bodies is made as specified below.

Applies to: 'X'_LE_T2-: Whenever the organisational type of a political function is specified as a legislative body and the tier as T2 (pf_instdomain = LE, and pf_orglevel = T2), then more detailed information on the type of legislative body is collected:

Code	Category	Explanation / notes	Example
CO	Committee	Legislative committee	"Committee on Legal Affairs", "Kommission für Verkehr und Fernmeldewesen des Nationalrates"

DL	Delegation	Legislative delegation	"Delegation bei der parlamentarischen Versammlung der NATO", "Deutsche Delegation in der OSZE"
FA	Faction party group	Legislative faction, please note that board members of factions get pfposition 02.	"Grüne Fraktion", "Groenlinks"
FA-T1	Faction committee	Sub-committee of a faction	Example goes here
FA-T2	Faction committee	Sub-committee of a faction	Example goes here
FA-T3	Faction committee	Sub-committee of a faction	Example goes here
AG	All-Party Parliamentary Groups	All-Party Parliamentary Groups (APPGs) are informal topical cross-party groups that have no official status within Parliament.	"Parlamentarische Gruppe für Altersfragen", "Parlamentarische Gruppe Binnenschifffahrt"

Table 10: Specification of political function secondary code: legislative steering body codes

Secondary Political Function Coding: Advisory 'extra parliamentary' expert committee Applies to: 'X?/NT?'_AD_T2-: Whenever the organisational type of a political function is specified as a administrative body (pf_instdomain = AD), then we have the option to specify the membership of special expert committees. E.g. 'commitee installed by the Minister' in the Netherlands. [[this is not in the coding space yet!]]

Secondary Political Function Coding: Youth versus Regular party Applies to: 'X'_PA_: Whenever the organisational type of a political function is specified as a party (pf_instdomain = PA), then more detailed information on the type of party is collected:

Code	Category	Examples
MA	Main party	"FDP"
YO	Youth party	"JUSO"

Table 11: Youth or Regular Party Political Functions

Secondary Political Function Coding: Regular versus Shadow positions - DNC (Table 12) Applies to: 'X'_[LE/EX]_: Whenever the organisational type of a political function is either legislative or executive, then it will be specified if this function is the regular or the shadow function.

Code	Category	Examples
01	Regular position	
02	Shadow position	"Official Opposition Shadow Cabinet"

Table 12: Shadow or Regular Political Functions - DNC

Secondary Political Function Coding: Important party committees (Table 13) Applies to: 'X'_PA_T2: Whenever the organisational type of a political function is specified as a party (pf_instdomain = PA) steering body / committee (pf_orglevel = T2), then more detailed in-formation on the type of steering body is collected. Please note that steering bodies can consist of one or more individuals.

Code	Category	Examples
10	Substantive committee per policy area. **See pf_policy_area!** (6.1.2)	
20	Election specific committees	
_21	(_Election) programme committee	"Wahlprogrammkommission"
_22	_Election campaign team	"Wahlkampfkommission"
_23	_Candidate selection committee	
_24	_Referendum / popular initiative campaign team (CH)	
_25	_Other Election Committee	
30	Other organisational (not policy-area specific) committees	
_31	_Ideological faction	"Seeheimer Kreis"
_32	_Regional committee	"Forum Ostdeutschland der Sozialdemokra
_33	_Training/development/recruitment	
_34	_Party magazine	
_35	_Other organisational commitee	
40	External party-controlled organisations	Note: Need to be legally independent
_41	_General scientific bureau	
_42	_Fundraising body	

Table 13: Important Party Committees

Secondary Political Function Coding: Interest groups codes Closely based on the typology of the `http://interarena.dk/` project and the coding applied by Mach, Varone, Eichenberger, and Gava (see Gava et al (2016). "Interest Groups in Parliament: Exploring MPs' interest affiliations (2000-2011)". Swiss Political Science Review) Applies to: 'X'_IG_: Whenever the organisational type of a political function is specified as an interest group (pf_instdomain = IG), then more detailed information on the type of interest group has to be collected. Information is stored in the variable ig_type in the ORGS data frame.

Code	Category	Explanation / notes	Example
10	**Unions**	Associations of employees negotiating work-related terms and conditions	"Blue-collar union", "Federatie Nederlandse Vakbeweging"
20	**Business org.**		
21	Business / trade association	Associations of firms, and associations of business associations	"Verband der Chemischen Industrie e.V. (VCI)", "Arbeitgeberverband der Schweizer Maschinenindustrie"
26	Private Firm / company	Firms by majority in possession of private owners. Firms are only code-able as interest groups if the position within the firm falls mainly into the area of public affairs.	"Board of oversight at Shell Global"
27	Public Firm / company	Firms by majority in possession of public owners (state, regions, municipalities etc.).	"Board of oversight at SBB AG"
30	**Institutional organisation**	Organisation of public authorities or institutions that are primarily publically funded	"swiss universities", Netz der Berner Spitäler und Kliniken"
40	**Occupational organisation**	Organisation of employees not negotiating terms and conditions	"Schweizerischer Bankpersonalverband", "Doctors' associations ", "Teachers' association" ETC
50	**Identity groups**	Organisation where supporters have a selective interest in group goals (not work related)	Patients, Elderly, Students, Friendship groups, Racial or ethnic, Gender "ProSenectute", "Procap"

60	**Hobby/leisure groups**	Functions in organisation for people with a common sport/leisure interest including arts and culture. This can only be coded as political if they touch on matters of public affairs.	"United Hockey Union"
70	**Religious group**	Organisation representing churches and religious communities	"Schweizerischer Evangelischer Kirchenbund"
80	**Public interest group**	Organisations aimed at the promotion of issues of general public concern (e.g., environ-mental protection, human rights, and consumer rights).	"WWF", "ICRC"
90	**Other interest group**	...	...

Table 14: Specification of political function secondary code: interest group codes

6.1.4 Secondary Political Function Coding: Bi-Cameral System (Lower House, Upper House) Applies to: X_LE:

With the secondary codes "LH" meaning lower house or "UH" meaning upper house, we distinguish the chambers in a bi-cameral system. With this secondary code it is possible to code representatives of the specific chamber. If there is no bi-cameral system on a certain level, it will be interpreted as a higher level and therefore coded on LE-level. We will only code on LE-level when there is no bi-cameral system. Committees which have members in both chambers will also be coded on the general LE-level.

Code	Category	Examples
UH	Upper house	"Eerste Kamer"
LH	Lower house	"Tweede Kamer"

Table 15: Upper or lower house

6.1.5 Political functions coding: some final specifications

1. Writing numbers: Sometimes, there are more than 10 categories. To keep the hierarchy intact, numbers should always be written as 1 to 9 as 01 to 09 (at least for levels 4).
2. Delimiters: A code consisting of numbers and letters could be difficult to read. Especially if differences in length appear due to more information (policy area). Therefore, we introduce delimiters in the code. A possible delimiter might be "_". While not strictly necessary, delimiters would facilitate reading the codes and keeping an overview. Ex. The code for the Chair of the EP's standing committee on Internal Market and Consumer Protection B_A_02_01_01 (where "B" denotes the EU level, "A" the legislative, "02" Committee / delegation / faction, the first "01" the policy area, and second "01" means leadership)
3. Missingness: If information is missing on policy area, then this will be indicated by either "NA" when a policy area specification is not applicable to the current positions (for example, The code for a municipal councillor might be: I_B_1_NA_03 because this position does not apply to a specific policy area) or "NC" when it is **n**ot **c**oded, which means that I did not yet get around the coding this entry.

7 Parliament Data Frame (PARL)

The parliament data frame (PARL) includes variables that contain relevant information related to the parliament primarily as an institution. It aims to cover both stable and time-variant characteristics of national and regional parliaments. The characteristics of parliaments change over time, differ between countries and differ between levels within countries. A "parliament" thus always entails a combination of a legislative term (e.g. `leg_period`) and either a country (`country_abb`) or a within country region (`region_abb`).

Table 16: Summary of Parliament Level Variables (PARL)

NAME	TYPE	VALUES/EXAMPLE[14]	SHORT DESCRIPTION	EFFORT
`parliament_id`	String	[country_abb]_[level] -[type/reg_abb]_[year] NL_NT-TK_1946 DE_NT-BT_2009 CH_NT-NR_1971 CH_RE-BS_1997 DE_RE-BW_1956	**parliament identifier:** Unique code for different parliaments, both across countries/levels and over time.	COMP
`leg_period_start`	Date	29apr1976 aug1953 2012	**legislative period start date:** Variable that captures when a specific legislative period started.	COMP
`leg_period_end`	Date	29apr1976 aug1953 2012	**legislative period end date:** Variable that captures when a specific legislative period ended.	COMP
`level`	String	NT RE	**Location in the multi-level structure:** The organisational level at which the parliament is situated	COMP
`country_abb`	String	CH = Switzerland DE = Germany NL = Netherlands	**country abbreviation of the parliament**	COMP
`region_abb`	String	BB = Brandenburg ZH = Zürich	**abbreviation of the region:** abbreviation of the name of the region (NL), federal state (DE) or canton (CH) of the parliament	COMP
`leg_period`	Integer	43 13	**number of the legislative period**	CWA[16]
`assembly_name`	String	Kantonsrat Grosser Rat Buergerversammlung Landesparlament	**name of assembly:** as locally used	COMP
`assembly_abb`	String	Bundestag(BT) Tweede Kamer(TK Kantonsrat (KR) Grosser Rat (GR) Landesparlament (LP)	**two letter abbreviation of the assembly**	COMP
`coalition_parties`	String	NL_VVD_NT; NL_PvdA_NT	Array of party_id's indicating which part(ie /y) was / where a part of the governing coalition during this parliamentary term	
`previous_parliament`	ID	NL_NT-TK_1946	Parliament_id of the previous parliament	

`parliament_id`: Unique identification code for different parliaments, both across countries/level and over time. The id consists of the country abbreviation (variable `country_abb`), the organizational level of the parliament

[16] For regional parliaments that have existed since the middle ages, counting legislative periods might not be feasible.

(see `level`), the type (DE: BundesTag(BT) & BundesRat(BR); NL: Tweede Kamer(TK) & Eerste Kamer (EK); CH: National Rat (NR) & StandeRat (SR), the abbreviated region (variable `reg_abb`), and the start year of the legislative period (see variable "leg_period_start") appended by underscore.

`leg_period_start:` Contains the official start date, in the same format as described in variable "birth_date" (two digit day, month (first three letters), and four digit year). The chosen date is the official date mentioned by the parliament as the first day of the respective legislative term, if available. Otherwise the day after the election is used.

`leg_period_end:` Contains the official end date, in the same format as described in variable "birth_date" (two digit day, month (first three letters), and four digit year). The chosen date is the official date mentioned by the parliament as the last day of the respective legislative term (i.e. the last day the parliament was in session), if available, otherwise the day before the next election is used.

`country_abb:` Abbreviated name of the country the parliament (parliamentarian) belongs to (works for). The used abbreviations follow the ISO "ALPHA-2 code. (See: http://www.nationsonline.org /oneworld/country_code_list.htm)

`region_abb:` Abbreviation of the name of the region (NL), federal state (DE), or canton (CH) of the parliament using two captial letters.

`assembly_name:` String variabls that captures the assembly name as commonly used. For the writing of the assembly name, refer to the rules explained regarding the variable `last_name`.

8 Membership Episode Data Frame (MEME)

The membership episode data frame (MEME) includes variables that relate to instances of membership in parties (both national and regional) of an individual. The characteristics of member episodes change over time and differ between individuals. A "member episode" entails a combination of the party in question (i.e. `party_id`), an id specific to the individual politician (`pers_id`), and the id of the party (`part_id`). An example of this data-frame can be downloaded at: `[URL]`.

Table 17: Summary of Membership Episode Variables (MEME)

NAME	TYPE	VALUES/EXAMPLE[17]	SHORT DESCRIPTION	EFFORT
`memep_id`	String	CH_Abate_Fabio_1966 __CH_RL_NA__2	**membership episode identification code:** a combination of `pers_id`, `party_id` and an integer to count episodes.	COMP
`pers_id`	String	CH_Abate _Fabio_1966 DE_Mueller _Johann_1956_dec	**parliamentarians' individual identification code:** a combination of `country_abb`, the `last_name`, the `first_name` and `birth_date`	COMP
`party_id`	String	DE_CSU_RE-BY	**unique party identifier:** a combination of `country_abb`, `party_abb`, level, `reg_abb`	COMP
`memep_startdate`	Date	29apr1976 aug1953 2012	**membership episode start date:** Variable that captures when a specific membership episode started.	COMP
`memep_enddate`	Date	29apr1976 aug1953 2012	**membership episode end date:** Variable that captures when a specific membership episode ended.	COMP

Table 17: Summary of Membership Episode Variables (MEME) (Continued)

NAME	TYPE	VALUES/EXAMPLE[18]	SHORT DESCRIPTION	EFFORT
memep_type_raw	String	honorary member	**membership episode type (raw):** a string variable that describes the kind of party membership the individual in question has.[19]	CWA
meme_entry_raw	String	ARP (Anti-Revolutionaire Partij), tot juli 1970	**membership episode entry (raw):** a string variable that describes the membership in a party.[20]	CWA
meme_entry_source	String	Dutch PDC data May 2019 extraction	**membership episode entry source:** a string variable that indicates the source of the membership episode.[21]	CWA
meme_date_altered	number	0 or 1	**indicator if altered:** a number variable that indicates if the date in the meme_entry was altered manually (= 1) or not (= 0).	CWA

memep_id: Identification code that is specific to this data frame. It combines an individual' (pers_id) with a specific party (party_id). When an MPs has multiple episodes in the same party (e.g. she breaks with party 'X' but later becomes a member again then an index (_01, _02 e.t.c.) is used to signify subsequent episodes.

pers_id personal identifier: Unique identification code for different politicians. See POLI for details on page 4.

party_id unique party identifier: party identifier consisting of the country abbreviation (see 7, variable country_abb), the party abbreviation (see party_abb), the organizational level and the abbreviated region (see 7, variables level, and reg_abb) appended by underscore.

memep_startdate membership episode start date: Contains the start date of the membership episode, in the same format as described in variable birth_date (two-digit day, month (first three letters), and four-digit year). The chosen date is either: i) the date on which the MP in question became a member of the party; ii) the official date mentioned by the parliament as the first day of the respective legislative term on basis of which a party membership episode could be constructed because an MP occurred on this party its election list (see the ELLI data-frame for more details).

memep_enddate: membership episode end date Contains the end date of the membership episode, in the same format as described in variable birth_date (two digit day, month (first three letters), and four digit year). The chosen date is either: i) the date on which the MP left the party or the party membership of the MP was revoked,ii) the day the MP joined another party of faction, iii) the official date mentioned by the parliament as the last day of the respective legislative term following the same logic as just mentioned for the start date.

meme_entry_raw membership episode entry(raw): Contains all raw information about party membership episodes, e.g. party name, faction name, start and end date of the episode if this information was taken from a qualitative source.

meme_entry_source membership episode entry source: Indicates the source from where the information about the party membership episode comes from.

[19] If no information, defaults to "NA".
[20] If no information, defaults to "NA".
[21] If no information, defaults to "NA".

`meme_date_altered` Indicates if the memep_startdate and/or memep_enddate were manually changed (and thus do not correspond to the content of meme_entry_raw). The number 1 indicates a manual change, while the number 0 indicates no change in the date.

9 Party Data Frame (PART)

The party data frame (PART) includes variables that relate to the parties included in the dataset. The characteristics of parties change over time, between countries, and across regions. A "party" entails a combination of the country (`country_abb`), the party (`party_abb`), the level at which the party operates (`level`), and the region (`reg_abb`). An example of this data-frame can be downloaded at: `[URL]`.

m.18— m.06— m.22— m.30— m.08—!

`party_id` **unique party identifier:** party identifier consisting of the country abbreviation (see 7, variable country_abb), the party abbreviation (see `party_abb`), the organizational level and the abbreviated region (7, variables `level`, and `reg_abb`) appended by underscore.

`party_level` **organizational/federal level a party operates on.**

`region_abb` **region abb abbreviation region:** abbreviation of the name of the region (NL), federal state (DE), or canton (CH) of the parliament using two capital letters.

`ancestor_party_id` **ancestor party id:** shows the party id of other party(s) if a party has came about from a merger or after a split from another party (if applicable). Can be empty.

`mother_party_id` **mother party id:** gives the id of (the) mother party (if applicable). Signifies relations between parties in the multilevel structure (e.g. but not limited to lower level membership implying higher level membership). Can be empty.

`party_abb` **abbreviated party names:** party names as used by ParlGov, consisting of a party abbreviation appended to a three digit country abbreviation by underscore.

`party_name`**: party names** Full party name in original language: the same spelling rules as set out for the variable "last name" apply. Furthermore the rules set out by ParlGov apply. See: http://www.parlgov.org/documentation/codebook.

`party_parlgov_id` **party id:** party id as used by ParlGov (see `http://www.parlgov.org/documentation/codebook`).

`old_party_abb` **old party abb:** a(n array of) party potential abbreviation(s) that the party has had in the past, separated by semi-colons.

`old_party_name` **old party name:** a(n array of) potential party name(s) that the party has had in the past, separated by semi-colons.

10 Faction Data Frame (FACT)

The faction data frame (FACT) includes variables that relate to groups of parties in assemblies (i.e. 'factions', or 'parliamentary party groups'). The characteristics of factions change over time, across parliaments, and within parliaments. We define a 'faction' as collection of individuals in parliament that collectively represent either a party or a collection of parties. Not that this thus entails both the group made up entirely by members of the same party, like is for example the case in the Netherlands, as party groups like is typicall in Germany and Switzerland. If factions loose a seat, for example one politicians breaks with them and joins another party or starts her own faction, then a new line is added to the faction data-frame with the new faction layout and the date of the break as the respective start and end dates of the two episodes. Because characteristics of party groups sometimes differ from the characteristics of their subparties (e.g. number of seats, gender quotas), in later cases both the parties collectively and and their party specific subgroups will occur in the data. The DE_NT_2009__ DE_CDU_NT–DE_CSU_NT group will as such occur three times in the data. Once as *"DE_NT_2009__ DE_CDU_NT–DE_CSU_NT"*, but also as *"DE_NT_2009__DE_CDU_NT"* and as *"DE_NT_2009__DE_CSU_NT"* **(We should discuss this as it adds a lot of redundant information: Wouldn't it make more sense to connect QUOT to PART and add start and end dates for quotas in QUOT? Then, there would be no redundant entries in FACT. For seats per party within the same faction, faction composition already covers this.)**.

b

NAME	TYPE	VALUES/EXAMPLE[25]	SHORT DESCRIPTION	EFFORT
	`faction_id`	String		

[25] See text below table for more detailed variable specifications

	NAME	TYPE	VALUES/EXAMPLE[26]
	`faction_name`	String	
	`party_id('s)`	String	
	`parliament_id`	String	
	`seats`	Integer	
	`faction_composition`	Integer	
	`faction_start`	Date	
Date		29apr1976 aug1953 2012	**faction end date:** Variable tha

`faction_id` **faction identifier:** combination of the name of the faction (`faction_name`) and `parliament_id`.

`faction_name` **name of the faction:** mentions the full name of the faction (as given by the relevant parliament/in the original language). As for spelling of the names, see rules at `last_name_raw`.

`party_id('s)` **unique identifiers for faction parties:** party identifiers for all parties included in the faction; each party identifier consists of the country abbreviation (see 7, variable `country_abb`), the party abbreviation (see `party_abb`), the organizational level and the abbreviated region (7, variables `level`, and `reg_abb`) appended by underscore.

parliament_id: Unique identification code for different parliaments. See PARL for details on page 7.

`seats`**:** the number of seats that the faction has in the respective eppisode in parliament.

`fact_composition` **composition of the faction:** the seats held by parties that constitute the faction, in the same order as the parties contained in `faction_id`.

faction_start: This variable gives the start date of a faction. These date entries consist of the two-digit day, the month (first three letters, small, of the English name of the month) and the four-digit year. In case of missing information on days or month, shortened versions of the date are used. In cases when the sources do not provide episode data, periods are inferred based on snapshots of faction compositions in conjunction with parliamentary episodes.

faction_end: This variable gives the end date of a faction. These date entries consist of the two-digit day, the month (first three letters, small, of the English name of the month) and the four-digit year. In case of missing information on days or month, shortened versions of the date are used. In cases when the sources do not provide episode data, periods are inferred based on snapshots of faction compositions in conjunction with parliamentary episodes.

11 Committee Data Frame (COMM)

The committee data frame (COMM) includes variables that relate to non-partisan specific groups in legislative assemblies (i.e. 'parliamentary committees', 'parliamentary delegations', 'all-party parliamentary groups', and their respective subbodies). These bodies can either have official standing in parliament such as committees

and delegations or constitute informal cross-party groups such as all-party parliamentary groups. In either case, the characteristics of these legislative bodies change over time, across parliaments, and within parliaments.

Table 19: Summary of Faction Variables (COMM)

NAME	TYPE	VALUES/EXAMPLE[27]	SHORT DESCRIPTION	EFFORT
comm_id	PriID	CH_NT-NR_1999__CO|FK–SC|EDA/WBF CH_NT-NR_2003__CH_NT-SR_2003__DL|Del-FL CH_NT-NR_2011__CH_NT-SR_2011__AG|Wohn - und Grundeigentum	**committee identifier**: combination of the parliament_id(s), the committee maintype, the abbreviated maintype name, as well as the subtype and the abbreviated subtype name if applicable. parliament_id and maintype are separated by double underscores'__'. Maintype/subtype and abbreviated maintype/subtype name are separated by a vertical bar '|'. In case a subtype is specified, it is separated from the abbreviated maintype name by double hypens '–'.	COMP
parliament_id	ID	NL_NT-TK_1946 DE_NT-BT_2009 CH_NT-NR_1971 CH_RE-BS_1997 DE_RE-BW_1956	**parliament identifier:** Unique code for different parliaments, both across countries/levels and over time.	COMP
comm_name	String	Delegation bei der parlamentarischen Versammlung der OSZE Geschäftsprüfungs-kommission Parlamentarische Gruppe Kultur	**committee name**: name of the committee, delegation, all-party parliamentary group, or subcommittee.	COMP
comm_maintype	String	CO DL FA AG	**type of committee**: Body maintype is either parliamentary committee (CO), parliamentary delegation (DL), faction (FA) or all-party parliamentary group (AG). Main type uses two-letter abbreviations.	COMP
comm_subtype	String	SC SD SA	**subtype of committee**: Body subtype refers to either parliamentary subcommittee (SC), parliamentary subdelegation (SD), or all-party parliamentary subgroup (SA). Subtype uses two-letter abbreviations.	CWA
comm_maintype_abb	String	FK FinDel Drogenpolitik	**main committee abbreviation**: Abbreviation of main body as used by officials. If none is available, a shortened version of comm_name is used instead.	COMP
comm_subtype_abb	String	EDA/WBF AgrRisikorep D	**subcommittee abbreviation**: Abbreviation of subbody as used by officials. If none is available, a shortened version of comm_name is used instead.	CWA

Table 19: Summary of Committee Variables (COMM) (Continued)

NAME	TYPE	VALUES/EXAMPLE[28]	SHORT DESCRIPTION	EFFORT
seats	Integer	• 45 • 27;28;29	**committee seats**: the number of seats that all MPs included in the respective body have combined during the respective parliament.	CWA
comm_start	Date	29apr1976 aug1953 2012	**committee start date:** Variable that captures when a specific committee, delegation, or all-party parliamentary group came into being or changed with regard to any other collected COMM variable.	COMP
comm_end	Date	29apr1976 aug1953 2012	**committee end date:** Variable that captures when a specific committee, delegation, or all-party parliamentary group ended or changed with regard to any other collected COMM variable.	COMP

comm_id: Combination of the parliament_id(s), the committee maintype, the abbreviated maintype name, as well as the subtype and the abbreviated subtype name if applicable:
[parliament_id(s)]__[comm_maintype]|[comm_maintype_abb]–[comm_subtype]|[comm_subtype_abb]
Multiple parliament_id as well as the last parliament_id and the maintype are separated by double underscores'__'. Maintype/subtype and abbreviated maintype/subtype name are separated by a vertical bar '|'. In case a subtype is specified, it is separated from the abbreviated maintype name by double hypens '–'

parliament_id: Unique identification code for different parliaments. See PARL for details on page 7.

comm_name: Official long version of the name of the committee, delegation, all-party parliamentary group, or subcommittee. Exception: In case of faction, "Leadership body" or "Working Groups".

comm_maintype: Main type refers to either parliamentary committee (CO), parliamentary delegation (DL), faction (FA) or all-party parliamentary group (AG). Main type uses two-letter abbreviations.

comm_subtype: Subtype refers to either parliamentary subcommittee (SC), parliamentary subdelegation (SD), or all-party parliamentary subgroup (SA). Subtype uses two-letter abbreviations.

comm_maintype_abb: Abbreviation of main type as used by officials. If none is available, a shortened version of comm_name is used instead.

comm_subtype_abb: Abbreviation of subtype as used by officials. If none is available, a shortened version of comm_name is used instead.

seats: The number of seats that all MPs included in the respective body have combined during the respective parliament. If the number of seats fluctuates during the same term, the different seat numbers are separated by ' ; '.

comm_start: Variable that captures when a specific committee, delegation, or all-party parliamentary group came into being or changed with regard to any other collected COMM variable. These date entries consist of the two-digit day, the month (first three letters, small, of the English name of the month) and the four-digit year. In case of missing information on days or month, shortened versions of the date are used. In cases when the sources do not provide episode data, periods are inferred based on snapshots of faction compositions in conjunction with parliamentary episodes.

comm_end: Variable that captures when a specific committee, delegation, or all-party parliamentary group ended or changed with regard to any other collected COMM variable. These date entries consist of the two-digit day, the month (first three letters, small, of the English name of the month) and the four-digit year. In case of missing information on days or month, shortened versions of the date are used. In cases when the sources do not provide episode data, periods are inferred based on snapshots of faction compositions in conjunction with parliamentary episodes.

12 Election Data Frame ()

The election data frame () includes information on elections: for example when did the election take place? Which parliament was elected? The data frame captures each election (i.e. also the second round of a election, by-elections and replacement-elections) on the level of the district it is executed on (i.e. on the level of the election lists).

Table 20: Summary of Election Level Variables ()

NAME	TYPE	VALUES/EXAMPLE[29]	SHORT DESCRIPTION	EFFORT
`election_id`	PriID	[district_id]_[election_date] DE_NT-BT_2017__ Hessen __24sep2017	**unique election identifier:** Unique code used to identify an election, consisting of the district_id and the election_date	COMP
`parliament_id`	String	[country_abb]_[level] -[type/reg_abb]_[year] NL_NT-TK_1946 DE_NT-BT_2009 CH_NT-NR_1971 CH_RE-BS_1997 DE_RE-BW_1956	**parliament identifier:** Unique code for the parliament the election related to/elects.	COMP
district_id	ID	[parliament_id] __[constituency _name] CH_NT_1971 __Zuerich	**unique district identifier:** identification code of the district, as used in the other data frames, identifying the district the electoral list belongs to.	COMP
`election_date`	Date	18oct1987	**date of election**	COMP
`election_type`	String	regular early	Variable indicating if the election was a normal or an early/snap election due to special circumstance (like a crisis and constitutional change)	CWA
`election_mode`	String	popular by_other_parliament town_square		CWA
`parliament_id_` `of_electing_parliament`	ID	DE_RE-BW_1956		COMP
district_votes _total_source	Numeric	• 1250	**total number of valid votes:** cast within a given district and period	CWA

`election_id`**:** Unique identification code for different parliaments, both across countries/level and over time. The id consists of the

`election_date`**:** Contains the official date(s) the respective parliament was elected, in the same format as described in variable "birth_date" (two digit day, month (first three letters), and four digit year).

district_votes_total_[source] total number of valid votes: a list got within a given district and period according to a given data source (which is mentioned in the end of the variable name).

13 Electoral List Data Frame (ELLI)

The electoral list data frame (ELLI) includes information on the individual electoral lists, which might originate from one or several parties. All variables in this data frame are on the level of the electoral lists and vary per parliamentary term: TV_P (start/end 'parliament_id'). Each list is identified by a combination of the parliament id (parliament_id), the district id (district_id), and the name of the list(list_name). An example of this data-frame can be downloaded at: `http://parliamentarycareersincomparison.org/wp-content/uploads/2017/03/PCC-example-data_v311.xlsx`

Table 21: Summary of the Electoral List Level Variables (ELLI)

NAME	TYPE	VALUES/EXAMPLE	SHORT DESCRIPTION	EFFORT
list_id	PriID	[parliament_id]__[district_id] __[list_name] DE_NT_2009__ CH_NT_1971 __Zuerich __Landesliste-Sozialdemokratische-Partei-Deutschlands	**unique list identifier:** consisting of the parliament id, the list name and the district id	COMP
list_name	String	• Sozialdemokratische-und-Gewerkschaftliche-Liste • NA_ZH_1	**name of the electoral list:** mentions the full name of the electoral list (as given by the relevant electoral office)[31]. For CH only: In case the MP did not run on a list, we use the format `NA_[kanton_abb]_index`, where index starts at 1 and increments by 1 with every unnamed list for the Kanton in the order in which they are listed in the *Bundesblatt*	COMP
parliament_id	ID	[country_abb]_[level] -[type/reg_abb]_[year]	**parliament identifier:** Unique code for different parliaments. See parliaments date-frame on page 7	COMP
party_id	ID	[country_abb]_[party_abb] _[level]-[reg_abb] DE_CSU_RE-BY	**unique party identifier (see above)**	COMP
district_id	ID	[parliament_id] __[constituency _name] CH_NT_1971 __Zuerich	**unique district identifier:** identification code of the district, as used in the other data frames, identifying the district the electoral list belongs to.	COMP
list_level	String	• DE_BY[32] – list on the German level of Bavaria • CH_GR – list on the level of the Swiss Canton Graubuenden	**level a list is on:** the variable specifies the level (national or subnational) a list is on and gives the identity of the level (e.g. Germany federal, or German State of Bavaria)	COMP

[31] In case a full list name (string does not exist, we use the list number (list_number)

[32] In some cases (e.g. the Netherlands) a list of electioral district abbreviations will be used instead of regional abbreviations

list_number	Numeric	• 1 • 13	**number of electoral list:** identifying number of consecutive number of the list as mentioned in the original data source (usually the national bureau for elections). Can be an array (e.g. 1;2;3) if the same election lists (same politicians in same order) was used across different electoral districts (this for example often happens in the Netherlands).	CWA
list_party_id	String (of IDs)	[country_abb]_[party_abb]_ [level]-[reg_abb] DE_CSU_RE-BY	**list party id(s):** id of the party(s) that constructed the list. Can be identical with the variable "party_id". Several party ids can exist in this variable, if the list is a collaboration of several parties	COMP
list_length	Numeric	• 1:99	**total number of spots on the electoral list**	CWA
list_connections	String (of IDs)	CH_NA_Sozialdemokratische-und-Gewerkschaftliche-Liste_36	**Id(s) of list(s) an electoral list is connected with**	CWA
list_votes_total_source	Numeric	• 1250	**total number of valid votes a list got**	CWA
list_type	String	• party = list of political party • youth = youth party list • party_female/ other_female = female-only party/ other list	**type of list: the variable characterizes the electoral list: who composed it and/or what characterizes the list?**	CWA

list_id unique list identifier: electoral list identifier consisting of the parliament id (see variable "parliament_id" in 7, consisting of variables [country_abb]_[level]_[reg_abb]_ [leg_period]), the list name (see "list_name"), the identitiy of the list level (see "list_level") connected by underscores.

list_name name of the electoral list: mentions the full name of the electoral list (as given by the relevant electoral office). As for spelling of the names, see rules at "last_name_raw". Blanc spaces between the words are exchanged by hyphens. When a list_name is unknown NA_[reg_abb]_[integer] is used.

parliament_id Unique identification code for different parliaments. See PARL for details on page 7.

party_id unique party identifier: party identifier. See PART for details on page 9

list_level level the list is on: the variable consists of the country abbreviation (two letters) and an abbreviation of the sub-national unit (region/state/or sub- and inter-regional units in the case of the NL, again two letter abbreviations) appended to each other by underscore. The variable specifies the level (national or subnational) a list is on (a national party list would have a country abbreviation but no sub-national/regional abbreviation) and gives the identity of the level (e.g. DE_XX = German federal level list, or DE_BY = state based list in the German state of Bavaria).[33]

list_number number of electoral list: identifying number of consecutive number of the list as mentioned in the original data source (usually the national bureau for elections).

list_party_id list party id(s): id of the party(s) that constructed the list. Can be identical with the variable "party_id". Several party ids can exist in this variable, if the list is a collaboration of several parties. on the construction of party ids, see variable "party_id" in the PART data-frame.

list_length total number of spots on the electoral list

list_connections list_connections: this variable is NL - specific and mentions the list_id(s) of other lists that the respective list is connected with, which means that these lists will get the votes that the respective list can not use (e.g. because it did not make the quorum).

list_votes_total_source total number of valid votes a list got according to a specific source.

list_type the variable mentions the list-type. The values of this variable provide information about who composed the list (a political party or some other group) and who is typically on the list (e.g. only females). In case the information is not known, the variable takes the value "NA".

Possible codes can be:

- party = list of political party
- youth = youth party list
- party_female/ other_female = female-only party/ other list
- party_male/ other_male = male-only party/other list
- multi_party/ multi_other = multi party/ other groups list
- other = list of non-party groupings
- NA

14 Electoral List Entry Data Frame (ELEN)

The electoral list entry data frame (ELEN) includes information on the candidacy of individual politicians (on lists). All variables in this data frame are on the level of the individual politician (see `pers_id`) and vary per election (see `parliament_id`) and election district see `list_id` . Each election list entry is identified by a combination of the electoral list id (`list_id`), and the ID identifying the person (`pers_id`). An example of this data-frame can be downloaded HERE

Table 22: Summary of the Electoral List Entry Variables (ELEN)

NAME	TYPE	VALUES/EXAMPLE	SHORT DESCRIPTION	EFFORT
elec_entry_id	PriID	[list_id]__[pers_id] DE_NT_2009 __DE_NT_2009 __Saarland __Landesliste-Sozialdemokratische-Partei-Deutschlands __DE_Funk _Alexander_1974	**unique electoral list entry identifier:** consisting of the the electoral list id (`list_id`), and the personal id (`pers_id`)[34]	COMP

[33] Note that the variable "list_level" differs from the variable `level`, in that it does not only indicate the organizational level (e.g. national or regional), but also the identity of the specific level.

[34] In case the same person appears more than once on the same electoral list, a consecutive number is appended by underscore (e.g.: DE_NT_2009 __DE_NT_2009 __Saarland __Landesliste-Sozialdemokratische-Partei-Deutschlands __DE_Funk _Alexander_1974_2)

list_id	ID	[parliament_id]__[disctrict_id] __[list_name] DE_NT_2009__DE_NT_2009 __ Saarland __Landesliste- Sozialdemokratische- Partei-Deutschlands	**unique list identifier:** consisting of the parliament id, the list name and the list level	COMP
`pers_id`[35]	ID	CH_Abate _Fabio_1966 DE_Mueller _Johann_1956_dec	**parliamentarians individual identification code:** a combination of `country_abb`, the `last_name`, the"first_name" and the `birth_date`	COMP
listplace	Numeric	1:99	**listplace the person ran on**	COMP
elected_ [source]	Binary	0=not elected 1=elected	**elected_[source]:** binary variable that specifies a list position or other type of candidature of a candidate was successful (elected) (according to a specific source (mentioned in [external data source]))	CWA
candidature_type	String	L = list D = district LD = list and district	**type of candidature of a person:** only relevant for DE	CWA
seat_type	Numeric	1 direct mandate 2 list without succession 3 succession via list 11 byelection 12 Volkskammer GDR 13 list	**type of seat/mandate in detail:** only relevant for DE	CWA
district_id	ID	see ELDI	Primary ID in ELDI, also included here because candidates might run from subdistricts but be part of a larger primary district. Also included to make DE data easier to understand. Redundant.	CWA
candidate_votes	Numeric	1250	**total number of votes a candidate got in the first (possibly the only) round of elections (for the national parliament)**	CWA
candidate_votes2	Numeric	1250	**total number of votes a candidate got in the second round of elections (if applicable**	CWA

elec_entry_id unique electoral list entry identifier: consisting of the the electoral list id (`list_id`), consisting of [parliament_id]_[disctrict_id]_[list_name], and the personal id (`pers_id`), which is a combination of `country_abb`, the `last_name`, the"first_name" and the `birth_date`, connected by underscores.

[35]To ensure compatibility of datasets and make usage and merging of our dataset easier we will add several id variables. The first one (`pers_id`) will be constructed by us, consists of the country, last name, first name, and birth year (and possibly a running number), and is unique to parliamentarians across countries and levels. This 'information based' ID format allows consistent IDS to be costructed across a variety of sources while minimizing the need to look-up IDS in the already collected data.

list_id unique list identifier: electoral list identifier consisting of the parliament id (see variable "parliament_id" in 7, consisting of variables [country_abb]_[level]_[reg_abb]_ [leg_period]), the list name (see "list_name"), the identitiy of the list level (see "list_level") connected by underscores.

pers_id Unique identification code for different politicians. See POLI for details on page 4.

listplace Variable is a positive number that indicates the position on the electoral list a person ran on and/or was elected on.

candidature_type Variable is mainly important for German candidate selection and mentions the type of candidature (list, district, or both) of a candidate.

seat_type Variable is mainly important for (elected) parliamentarians from mixed systems like Germany. It specifies the seat type or mandate of a parliamentarian. The parliamentarian can either have a direct mandate (based on a district mandate, 1), can be part of the parliament due to a successfull list candidacy without succession (2), he/she could have succeeded another parliamentarian via the list (3), could have entered the parliament due to a byelection (11), could have been part of the GDR Volkskammer (12), or can be part of the parliament due to a successfull list candidacy (13).

candidate_votes Numeric variable mentioning the total number of (valid) votes a candidate/list position got in the first round (which may be the only round) of (national parliament) elections.

candidate_votes2 Numeric variable mentioning the total number of (valid) votes a candidate/list position got in the second round of (national parliament) elections.

elected_[source]: Binary variable that specifies whether, according to a specific source (mentioned in [external data source]), the federal statistical bureau for example, a list position or other type of candidature of a candidate was successful (elected) or not.

15 Electoral District Data Frame (ELDI)

The electoral district data frame (ELDI) includes information on the organizational units (districts) the elections are executed in. For example, the name of a given district or size of it. Variables in this data frame are on the level of the districts and (are allowed to) vary per election (see 'parliament_id'). Each electoral district (constituency) is identified by a combination of the parliament (parliament_id) and the constituency name (constituency _name). An example of this data-frame can be downloaded HERE

Table 23: Summary of District Level Variables (ELDI)

NAME	TYPE	VALUES/EXAMPLE	SHORT DESCRIPTION	EFFORT
district_id	PriID	[parliament_id] __[constituency _name] CH_NT_-NR_1971 __Zuerich	**unique district identifier:** identification code as used in the other data frames	COMP
parliament_id	ID	[country_abb]_[level]_[year] NL_NT_1946 DE_NT_2009 CH_NT_1971 [country_abb]_[level] -[reg_abb]_[year] CH_RE-BS_1997 DE_RE-BW_1956	**parliament identifier:** Unique code for different parliaments, both across countries/levels and over time.	COMP

country_abb	String	• CH = Switzerland • DE = Germany • NL = Netherlands	**country abbreviation of the parliament**	COMP
region_abb	String	• BB = Brandenburg • ZH = Zürich	**abbreviation of the region:** abbreviation of the name of the region (NL), federal state (DE) or canton (CH) of the parliament/district	COMP
constituency_name	String	• Zuerich • Stuttgart	**name of the constituency**	COMP
constituency_code_CLEA	Numeric	• 001–900 • 901–999	**identifying code of the constituency:** as assigned by CLEA	CWA
constituency_name_CLEA	Numeric	• 001–900 • 901–999	**name of the constituency:** as assigned by CLEA	CWA
dist_magnitude	Numeric	• -992 = Uncontested election (i.e. a single candidate contested the election) • -994 = Suspended election	**district magnitude:** measured by the number of seats allocated in a given district	CWA

district_id unique district identifier unique identification code for electoral districts consisting of the parliament id (variable "parliament_id"), and the name of the constituency (variable "constituency_name") appended by underscore.

parliament_id Unique identification code for different parliaments. See PARL for details on page 7.

country_abb country abbreviation of the parliament: abbreviated name of the country the parliament/district belongs to. The used abbreviations follow the ISO "ALPHA-2 code. (See: http://www.nationsonline.org /oneworld/country_code_list.htm)

region_abb abbreviation region: abbreviation of the name of the region (NL), federal state (DE), or canton (CH) of the parliament using two captial letters. For CH Based on urlhttp://swiss-government-politics.all-about-switzerland.info/swiss-federal-states-cantons.html

district_aliases district_aliases (array of) potential aliases the district it also known under. E.g. 'kies kring numbers' for NL.

constituency_name name of the constituency: mentions the full name of the electoral constituency (as given by the relevant electoral office/in the original language). As for spelling of the names, see rules at "last_name_raw". Called 'CST' in CLEA. Dots are removed. Remaining spaces are replaced by hypens ('-')

constituency_code_CLEA identifying code of the constituency: As assigned by the Constituency Elections Archive (CLEA) (see http://www.electiondataarchive.org/about.html). Note: this code is unique only to constituency and election (hence, the same district can have different codes at different elections). CLEA codes exist only for some national elections.

dist_magnitude district magnitude: measured by the number of seats allocated in a given district. The number of seats is expressed in a positive number, the following other codes are possible:

- -992 = Uncontested election (i.e. a single candidate contested the election)
- -994 = Suspended election

(See also urlhttp://www.electiondataarchive.org/about.html)

16 Organisations Module(ORGS)

The organisations data frame (ORGS) captures information on the level of organisations that parliamentarians are members of at some stage of their career. In most instances, these organisations are interest groups as data on parties (PART), factions (FACT), and parliaments (PARL) are collected in other data frames. Responsible for the module and the data contained is: OH.

Some of the data presented here are drawn from secondary sources. For more extensive discussions of these sources, readers are therefore referred to the codebooks and studies for which these data have been collected and coded for in the first place.

Table 24: Summary of Organisation Level Variables (ORGS)

NAME	TYPE	VALUES/EXAMPLE	SHORT DESCRIPTION	EFFORT
org_id	ID	[country_abb]_[12 digit code] CH_H8KkcewN6fGv	**unique organisation identifier:** identification code to link organisations mentioned in RESE	COMP
org_name	String	[name$_1$]/[$name_2$]/[$name_n$], [place of headquarters] Cranex AG, Opfikon Crédit Suisse / Schweizerische Kreditanstalt (SKA) fenaco Genossenschaft / Fédération de coopératives agricoles, Bern	**name of organisation:** Name and synonyms (previously used names, names in other languages) used for an organisation separated by slashes as well as the place of the organisation's headquarters. Minimum requirement: One name.	COMP
country_abb	String	• CH = Switzerland • DE = Germany • NL = Netherlands	**country abbreviation of the organisation's headquarters location**	COMP
legal_form	String	[legal_form$_1$]; $legal_form_2$; [legal_form$_n$] AG; Gen. KollG; EG	**legal form:** Legal form of the organisation as given in the source. Multiple legal forms are separated by semicolons.	CWA

polactive_start_date	Date	29apr1976 aug1953 2012	**political activity start date:** Variable that captures when an organisation was known to be politically active for the first time	CWA
polactive_end_date	Date	29apr1976 aug1953 2012	**political activity end date:** Variable that captures when an organisation was known to be politically active for the last time	CWA
polactive_end_date	Date	29apr1976 aug1953 2012	**political activity end date:** Variable that captures when an organisation was known to be politically active for the last time	CWA
address_raw	String	Musterstr. 3, Basel Postfach, Beispielstadt	**raw address:** Address of the organisation	CWA
GAVA_assoc_id	Numeric	1 329 4605	**Gava association id:** Id assigned to organisations by Gava et al. in the DATABASE ON INTEREST GROUPS IN THE SWISS PARLIAMENT	CWA
GAVA_assoc_name	String	Pro Natura Basel Sportstiftung Thurgau Lutecia SA Genève	**Gava association name:** Name of organisations as used by Gava et al. in the DATABASE ON INTEREST GROUPS IN THE SWISS PARLIAMENT	CWA
GAVA_assoc _aggregate_name	String	Pro Natura Swiss Life Industrie du bois Suisse	**Gava aggregate association name:** Umbrella names of organisations as used by Gava et al. in the DATABASE ON INTEREST GROUPS IN THE SWISS PARLIAMENT	CWA
GAVA_assoc _legal_form	Numeric	11 30 50	**Gava legal form:** Legal form of organisations as used by Gava et al. in the DATABASE ON INTEREST GROUPS IN THE SWISS PARLIAMENT	CWA
GAVA_assoc _general	String	Association Commission Societe	**Gava general form of association:** Form of organisations as used by Gava et al. in the DATABASE ON INTEREST GROUPS IN THE SWISS PARLIAMENT	CWA
GAVA_assoc_ig _type_name	String	Business groups Identity groups Unions	**Gava interest group category:** Categories of interest groups as used by Gava et al. in the DATABASE ON INTEREST GROUPS IN THE SWISS PARLIAMENT. The categories are adapted from the INTERARENA project.	CWA

GAVA_assoc_ig_code	Numeric	26 63 89	**Gava interest group category code:** Category codes of interest groups as used by Gava et al. in the DATABASE ON INTEREST GROUPS IN THE SWISS PARLIAMENT. The category codes are adapted from the INTERARENA project.	CWA
GAVA_assoc_ig_code	String	Private firm/company Arts Local development	**Gava interest group category code name:** Category code names of interest groups as used by Gava et al. in the DATABASE ON INTEREST GROUPS IN THE SWISS PARLIAMENT. The category codes are adapted from the INTERARENA project. This variable describes the numeric values of GAVA_assoc_ig_code.	CWA
GAVA_level	Numeric	1 2 4	**Gava level of organisation:** Level codes ranging from local to international as used by Gava et al. in the DATABASE ON INTEREST GROUPS IN THE SWISS PARLIAMENT.	CWA
GAVA_category	String	Local Regional/Cantonal International	**Gava category of level:** Level codes ranging from local to international as used by Gava et al. in the DATABASE ON INTEREST GROUPS IN THE SWISS PARLIAMENT. This variable describes the numeric values of GAVA_level.	CWA
GAVA_topic	Numeric	6 10 12	**Gava topic:** Policy area in which an organisation is active in as used by Gava et al. in the DATABASE ON INTEREST GROUPS IN THE SWISS PARLIAMENT. Based on the Comparative Agenda Project (CAP) coding scheme.	CWA

GAVA_topic _business	Numeric	1500 1505 1564	**Gava topic business:** For organisations that fall into the category of General Business (1500), Gava et al. provide more detailed information on the area of business in their DATABASE ON INTEREST GROUPS IN THE SWISS PARLIAMENT.	CWA
NCCR_nqid	Numeric	46 827 2608	**NCCR interest group id:** Id of interest groups for which data have been collected in the context of the project NCCR IP 8: MEDIATIZATION OF POLITICAL INTEREST GROUPS by Jarren et al..	CWA
NCCR_A4 Organisation	String	Alpen-Initiative Castagna, Zürich FachFrauen Umwelt	**NCCR interest group name:** Name of interest groups for which data have been collected in the context of the project NCCR IP 8: MEDIATIZATION OF POLITICAL INTEREST GROUPS by Jarren et al..	CWA
GK_interest _group	String	Verein Al Forno Semaine du Gout Swissgas AG	**Giger & Klüver IG name:** Name of interest group for which data have been collected by Giger and Klüver for their project on SWISS MPS AND THEIR INTEREST GROUP AFFILIATION.	CWA
GK_canton _ig_raw	String	LU Halle (BRD) ZH	**Giger & Klüver canton of IG location:** Raw (uncleaned) string of the canton in which an interest group is headquartered. These data have been collected by Giger and Klüver for their project on SWISS MPS AND THEIR INTEREST GROUP AFFILIATION.	CWA
GK_type	String	culture ngo prof	**Giger & Klüver IG type:** Type of interest group for which data have been collected by Giger and Klüver for their project on SWISS MPS AND THEIR INTEREST GROUP AFFILIATION.	CWA

GK_policy_code	String	7 13 28	**Giger & Klüver IG policy code:** Policy area of interest group for which data have been collected by Giger and Klüver for their project on SWISS MPS AND THEIR INTEREST GROUP AFFILIATION. Based on the Comparative Agenda Project (CAP) coding scheme.	CWA

Note on variables: The detailed description of variables, their ranges or categories are only provided for original variables presented in this codebook. For variables from secondary sources that are mentioned in the above table, readers are referred to the original codebooks and studies. The variables in question all start with a capitalised indicator: GAVA, NCCR, and GK.

org_id unique organisation identifier: unique identification code for organisations consisting of the country abbreviation (variable "country_abb"), and a 12 digit combination of letters, numbers, and special characters. The 12 digits code is created in Excel with the VBA hash function BASE64SHA1.

org_name organisation name Name(s) of the organisation as used over time in different languages. The string captures exclusively names that refer to exactly the same organisation (Organisations that split or merged are not added to the same organisation). Alternative names are separated by slashes. If available, the place where an organisation is headquartered is given at the end separated by a comma: $[\text{name}_1]/[name_2]/[name_n], [placeofheadquarters]$.

country_abb country abbreviation of the parliament: abbreviated name of the country the parliament/district belongs to. The used abbreviations follow the ISO "ALPHA-2 code. (See: http://www.nationsonline.org /oneworld/country_code_list.htm)

legal_form legal form: the legal form of organisations as indicated in the data source. The Register of Interest Ties prepared by the Swiss Parliamentary Services for instance distinguishes between 15 different types of legal forms. The current list of abbreviations and their meaning can be looked up online: `https://www.parlament.ch/centers/documents/de/interessen-nr.pdf`

polactive_start_date First known political activity: Captures the time when an organisation was first politically active e.g. had first ties to a parliamentarian. The date consists of the two-digit day, the month (first three letters, small, of the English name of the month) and the four-digit year. If only the month is known the the days are dropped and of only the years are know the day and month letters are dropped.

polactive_end_date Last known political activity: Captures the time when an organisation was last politically active e.g. had last ties to a parliamentarian. The date consists of the two-digit day, the month (first three letters, small, of the English name of the month) and the four-digit year. If only the month is known the the days are dropped and of only the years are know the day and month letters are dropped.

address_raw Raw address: Contains the raw address of the organisation's headquarters. If multiple addresses are known for the organisation, the time period for which the address is valid is added in brackets after the address e.g.[[2015-2017]]. Multiple addresses are separated by the pipe symbol "|".

17 The Quota Module (QUOT)

This is a PCC data module on (gender) quota information (at the level of the national parties, QUOT. The quota data is appended to and compatible with the FACT data frame, however the data is strictly speaking not part of the PCC data. Responsible for the module and the data contained are: EF and TTZ.

Table 25: Summary of Quota Variables (QUOT)

NAME	TYPE	VALUES/EXAMPLE[36]	SHORT DESCRIPTION	EFFORT
`faction_id`	String	DE_NT_2009-DE_CDU_NT	**faction identifier**: combination of the parliament_id, and the party id (see 10).	COMP
`quota_bin`	Binary	0=no gender quota 1=gender quota	**quota dummy**: indicates whether the party has any ('soft' or 'hard') gender quota or not	CWA
`quota_percentage`	Numeric	0:100	**quota value**: mentions the percentage (of candidates or posts) that the quota specifies to be female at least	CWA
`quota_zipper`	Binary	0=no zipper quota 1=zipper quota	**zipper quota dummy**: indicates whether the party reserves alternating seats to different sexes	CWA
`quota_soft`	Binary	0='hard' quota 1='soft' quota	**soft gender quota**: shows if a gender quota is 'soft'/informal/non binding in wording or not	CWA
`quota_execution`	Binary	0=quota not implemented 1=quota implemented	**quota implementation**: indicates whether the party actually implemented the quota or not	CWA

`faction_id` **faction identifier:** combination of the `parliament_id`. and the name of the party (`faction_id`) (see 10). Identifies the party–parliament dyad for which the quota of the national party applies. In case the quota was introduced during a parliamentary term, the quota information is recorded for the next parliamentary term/election.

`quota_bin` **quota dummy:** Dummy variable indicating if a party/fraction in or for a certain parliament has any gender quota ('soft' or 'hard') (1) or not (0). Gender quotas are defined as any ratio/share of females to males (bigger than zero) with regard to electoral candidates, list postions, or elected MPs (party offices and delegates for party conventions are not considered) mentioned in an official party document (party statute, party convention decision, minutes of party convention) as officially decided by the party.

`quota_females` **quota value:** Numeric variable that mentions the highest percentage of females (candidates or positions) that the quota prescribes explicitely. When the quota text prescribes "at least X percent", X is the recorded value. The variable is zero if the party has no gender quota (see variable `quota_bin`). When the wording 'equal' is used this is coded as a 50% quota.

`quota_zipper` **zipper quota dummy:** Binary variable that indicates if (1) the respective party has a gender quota that reserves alternating positions on electoral lists for male and female candidates (zipper system). A zipper system can also leave certain positions open to both sexes (e.g. SPD in Germany, prescribes the electoral list positions to be alternating in sexes but leavs every fivth position open for either sex).

`quota_soft` **soft gender quota:** Dummy variable that shows if a gender quota is 'soft'/non binding in wording (1) or not (0). For 'soft' quotas the party statutes say "ideally there should be..." or "there should be..." a certain share of females (or a similar expression).

`quota_execution` **quota implementation:** The dummy variable indicates whether, according to our information, the gender quota was implemented by the party in the respective legislative election/period (1) or not (0).

18 Appendices

18.1 Alternative IDs: ALID

ALID Different data sources offer different challenges and possibilities for the creation of IDs. To create an ID based on country, last name, first name, and year of birth might not be always be feasible. At the same time,

IDs of existing data sets (e.g. the Manow IDs) might need to be transposed to the PCC IDs by some means. If the creation of the PCC ID is not possible in one step, the use of auxiliary IDs makes sense.
The following table offers an overview of alternative IDs and the alternative values they use. For all alternative IDs, a table of correspondence will then be created to easily allow for them to be transposed to the PCC IDs. Please note that these IDs are not a data-frame in the classic sense, they will just be generated on basis of the information available.

Table 26: Alternative IDs

Alternative Value(s)	ID Structure	Examples
Region	Country_LastName_FirstName_Region	CH_Allenspach_Heinz_ZH, DE_Doerflinger_Thomas_BW
Party	Country_LastName_FirstName_Party	CH_Allenspach_Heinz_FDP—PRD, DE_Doerflinger_Thomas_CDU, NL_Elias_Ton_VVD
Year of Parliament Entry	Country_LastName_FirstName_EntryYear	CH_Allenspach_Heinz_E1979, DE_Doerflinger_Thomas_E1998, NL_Elias_Ton_E2008

18.2 Regional Abbreviations Used

18.2.1 Regional Abbreviations Used for CH

- AG = Aargau
- AI = Appenzell Innerrhoden
- AR = Appenzell Ausserrhoden
- BS = Basel-Stadt
- BL = Basel-Landschaft
- BE = Bern
- FR = Fribourg
- GE = Geneva
- GL = Glarus
- GR = Graubünden
- JU = Jura
- LU = Lucerne
- NE = Neuchâtel
- NW = Nidwalden
- OW = Obwalden
- SH = Schaffhausen
- SZ = Schwyz
- SO = Solothurn
- SG = St. Gallen
- TG = Thurgau
- TI = Ticino
- UR = Uri
- VS = Valais
- VD = Vaud
- ZG = Zug
- ZH = Zürich

18.2.2 Regional Abbreviations Used for DE

- BW = Baden-Württemberg
- BY = Bavaria
- BE = Berlin
- BB = Brandenburg
- HB = Bremen
- HH = Hamburg
- HE = Hesse
- NI = Lower Saxony
- MV = Mecklenburg-Vorpommern
- NW = North Rhine-Westphalia
- RP = Rhineland-Palatinate
- SL = Saarland
- SN = Saxony
- ST = Saxony-Anhalt
- SH = Schleswig-Holstein
- TH = Thuringia

18.2.3 Regional Abbreviations Used for NL

- DR = Drenthe
- FL = Flevoland
- FR = Friesland
- GD = Gelderland
- GR = Groningen
- LB = Limburg
- NB = Noord-Brabant
- NH = Noord-Holland
- OV = Overijssel
- UT = Utrecht
- ZH = Zuid-Holland
- ZL = Zeeland
- BO = Openbare lichamen Bonaire, Sint Eustatius en Saba