

PolPaL (Political Party Links) Merging Data with PolPaL v1.0

Paul Bederke & Lea Kaftan

As of October 2026

Table of Contents

1. Introduction 3

2. Download of the linkage keys 3

3. Merge data with PolPaL 4

3.1 Assessing the PolPaL linkage key tables 4

3.1.1 Groupings 4

3.1.2 Multiple Links 5

3.2 Merging the PolPaL linkage key tables 5

3.3 Controlling 7

4. Summary 7

5. References 8

6. Version history 8

1. Introduction

The main goal of PolPaL is to improve the merging of two or more datasets on political party data for researchers. In general, the process of merging datasets is time-consuming and error-prone. PolPaL provides linkage keys for a multitude of social science datasets which contain data on political parties. With the help of these linkage keys, the process of linking political party data is drastically sped up, and the number of errors is reduced. Scholars benefit from this because instead of building their own crosswalk tables, they merely need to merge existing linkage keys, make minor changes to suit their interests, and control the final list of linked political parties.

This document explains how to download linkage keys from the PolPaL website, merge two datasets using the PolPaL linkage keys, and control the final merged dataset.[1] Importantly, the linkage keys provided by PolPaL are not access-restricted; everyone can download and use those data.

2. Download of the linkage keys

The PolPaL website provides a page to download linkage keys for external datasets. This page can be found in the navigation bar at the top of the website called *Download*. On the Download page, the linkage keys can be downloaded from the PolPaL Linkages of External Data section. The user is required to select the (1) *data project*, (2) *dataset*, and (3) *data version* of the external dataset. Afterwards, the user can filter the selected dataset to get a subset of the linkage keys (see Table 1).

Table 1. Filter options for PolPaL linkages of external data.

Filter	**Explanation**
Short Format	Exclude variables containing the names of the relevant variables
Include catch-all entities	Include PolPaL IDs of PolPaL special entities (residual, independent, others, unknown)
Year Range	Select a specific time period
Region	Select all countries of a region
Countries	Select specific countries
Variable	Select specific variables that contain party IDs in the external dataset

[1]We also recommend reading the PolPaL codebook to understand our coding decisions, especially of PolPaL parties and PolPaL groupings.

3. Merge data with PolPaL

After downloading the PolPaL linkage key table for the relevant external datasets, users can begin merging these datasets. Because of the variety and complexity of external datasets that contain political party data, the process of merging multiple datasets is the responsibility of the individual researcher.

3.1 Assessing the PolPaL linkage key tables

We recommend that users first assess

- whether the linkage key tables for the data they want to use contain groupings (variable *entityType*).
- whether any entity in the external data has been linked to several entities within PolPaL (variable *multipleLinks*).

If either of these two cases applies, researchers should investigate whether the decisions each external dataset made regarding entity coding overlap or conflict with one another. If they overlap, researchers can simply proceed with merging the two datasets.

3.1.1 Groupings

One common conflict in the coding decisions for external datasets concerns the handling of political party groupings. For example, the German CDU and CSU are coded as two single entities in a variety of datasets. However, some datasets code the union between the CDU and CSU, namely CDU/CSU. If two datasets differ in how they code the CDU and CSU cases, merging the two datasets is not straightforward. It is now up to the researchers to decide which decision – keeping CDU and CSU separate or combining them as CDU/CSU – is more suitable for their research. As a concrete data example, the 1999-2024 Chapel Hill Expert Survey (CHES) trend file (Rovny et al., 2025) contains expert judgments on the CDU and CSU as separate parties. However, the Manifesto Project (Lehmann et al., 2025) provides data for the CDU/CSU.

With PolPaL, researchers can now easily detect this mismatch. The PolPaL linkage key table for the Manifesto Project flags the appearing CDU/CSU entries as a grouping. PolPaL provides a downloadable data table containing information on the PolPaL grouping members. For this

example, the table contains information that the CDU and the CSU are members of the CDU/CSU starting from 1949 onwards. This allows researchers to quickly assess the members of a grouping. The PolPaL linkage key table for the CHES trend file contains the CDU and CSU as separate entities. The researcher can now decide to (1) merge CDU/CSU with both CDU and CSU, (2) merge CDU/CSU with a (weighted) average of CDU and CSU, or (3) not merge CDU/CSU to CDU and CSU – depending on the researcher's primary interest.

3.1.2 Multiple Links

Another conflict arising from divergent coding decisions is the presence of multiple links from a single external data observation to several PolPaL entities. Datasets combine multiple political parties together as one observation. However, these might not align with the definition of PolPaL groupings (see *PolPaL Codebook*). For example, the World Values Survey (WVS) cumulates three far-right German parties into one response category (*DEU: REP, NPD or DVU*). As this cumulation is the design choice of the survey authors, this is not considered a PolPaL grouping. Instead, the observation is linked to the three separate PolPaL parties, The Republicans (ID 720), The Homeland (ID 1723), and German People's Union (ID 695). This results in a **one-to-many match (!)**, meaning the observation in the WVS gets duplicated for each linked PolPaL entity if merged with another dataset.

The researcher needs to decide whether (1) to aggregate the data from another external dataset for the separate PolPaL parties, (2) to merge data from another external dataset of only one of the separate PolPaL parties, or (3) not to merge data for this observation. There might be other options depending on the researchers' primary interest.

3.2 Merging the PolPaL linkage key tables

After assessing the PolPaL linkage key tables, the researcher can begin merging them. Here the researcher must decide whether (1) to keep the temporal dimension of the set linkages or (2) drop the temporal dimension. We recommend keeping the temporal dimension, as the set links between external dataset observations and PolPaL entities are more accurate.

For example, the ParlGov dataset on elections (Döring & Manow, 2024) codes the German PDS and The Left as one party, namely *PDS*|*The Left*. In PolPaL, those ParlGov observations up to (and including) the election in 2005 are linked to the PolPaL entity of the *PDS* (ID 86), observations for later elections are linked to the PolPaL entity of *The Left* (ID 1545). The temporal

dimension allows the election specific linking of the ParlGov election data to other external datasets by using PolPaL. Should the temporal dimension be dropped, the only information remaining is that the ParlGov single observation for *PDS|The Left* is linked to the PolPaL entities of the *PDS* (ID 86) and *The Left* (ID 1545) (see Table 2).

Table 2. Comparison of linkages between ParlGov (Döring & Manow, 2024) and PolPaL for the observation 'PDS|The Left'.

Without Temporal Dimension		With Temporal Dimension	
ParlGov (Elections)	PolPaL Entities	ParlGov (Elections)	PolPaL Entities
PDS | The Left	The Left (1545) Party for Democratic Socialism (86)	PDS | The Left (1990-12-02)	Party for Democratic Socialism (86)
		PDS | The Left (1994-06-12)	Party for Democratic Socialism (86)
		PDS | The Left (1994-10-16)	Party for Democratic Socialism (86)
		PDS | The Left (1998-09-27)	Party for Democratic Socialism (86)
		PDS | The Left (1999-06-13)	Party for Democratic Socialism (86)
		PDS | The Left (2002-09-22)	Party for Democratic Socialism (86)
		PDS | The Left (2004-06-13)	Party for Democratic Socialism (86)
		PDS | The Left (2005-09-18)	Party for Democratic Socialism (86)
		PDS | The Left (2009-06-07)	The Left (1545)
		PDS | The Left (2009-09-27)	The Left (1545)
		PDS | The Left (2013-09-22)	The Left (1545)
		PDS | The Left (2014-05-25)	The Left (1545)
		PDS | The Left (2017-09-24)	The Left (1545)
		PDS | The Left (2019-05-26)	The Left (1545)
		PDS | The Left (2021-09-26)	The Left (1545)

The researcher merges the PolPaL linkage key tables for two or more external datasets based on two dimensions:

1. The PolPaL ID, identifying the matching PolPaL entities for the external dataset observations.
2. The temporal dimension, the point in time of the external dataset observation.

While the match on the PolPaL ID is relatively straightforward, the match on the temporal dimension might require some additional adjustments.

For example, merging the PolPaL linkage key table for the ParlGov Elections dataset with that for the CHES trend file requires a researcher's decision. Should an observation in the CHES trend file be merged with the ParlGov observation for the previous or next election? This kind of decision depends on the researcher's primary interest!

3.3 Controlling

After the merging of two or more PolPaL linkage key tables, the researcher should control the resulting table. This should ensure that the researcher is aware of the matched observations and also identify mistakes in the linking and/or merging process.

The researcher should compare the names of the entities in the external datasets with those of the linked PolPaL entity. This reveals any mistakes made in the linking of the external dataset observations to PolPaL entities. While political party names often differ across datasets, substantial deviations should be controlled for to ensure that the same political party is actually being referred to.

The researcher should also review non-matches. While non-matches can occur due to only partial overlap among the covered political parties across two or more external datasets, the researcher should be aware of the reasons for those non-matches.

4. Summary

The merging of two or more datasets on political party data is a time-consuming and error-prone process. With PolPaL, this process is drastically sped up, and the number of errors is reduced. Scholars do not need to build their own crosswalk tables; they merely need to merge existing linkage keys. However, we want to emphasize that merging the PolPaL linkage key tables requires researcher-made decisions. There is no solution that fits all potential research interests and political party data. Researchers have to carefully undertake the merging process and make appropriate decisions dependent on their needs.

5. References

Döring, Holger. & Manow, Philip. 2024. ``ParlGov 2024 release [Dataset]''. Harvard Dataverse. https://doi.org/10.7910/dvn/2vz5zc

Lehmann, Pola, Simon Franzmann, Denise Al-Gaddooa, Tobias Burst, Christoph Ivanusch, Sven Regel, Felicia Riethmüller, Andrea Volkens, Bernhard Weßels, Lisa Zehnter. 2025. "The Manifesto Data Collection. Manifesto Project (MRG/CMP/MARPOR). Version 2025a." Berlin: Wissenschaftszentrum Berlin für Sozialforschung (WZB). Göttingen: Institut für Demokratieforschung (IfDem). https://doi.org/10.25522/manifesto.mpds.2025a

Rovny, Jan, Jonathan Polk, Ryan Bakker, Liesbet Hooghe, Seth Jolly, Gary Marks, Marco Steenbergen, and Milada Anna Vachudova. 2025. "The 2024 Chapel Hill Expert Survey on political party positioning in Europe: Twenty-five years of party positional data." Electoral Studies 97 (October). https://doi.org/10.1016/j.electstud.2025.102981

6. Version history

- Proof reading; correcting typos and adjusting the layout of the document