SOCIUM · SFB 1342
WorkingPapers No. 32

Holger Döring

Advancing data infrastructures on democratic representation: An open-source version of ParlGov

socium
Forschungszentrum
Ungleichheit und Sozialpolitik

Globale
Entwicklungsdynamiken
von Sozialpolitik
SFB 1342

Holger Döring
Advancing data infrastructures on democratic representation: An open-source version of ParlGov
SOCIUM SFB 1342 WorkingPapers, 32
Bremen: SOCIUM, SFB 1342, 2025

Universität Bremen

SOCIUM Forschungszentrum Ungleichheit und Sozialpolitik /
Research Center on Inequality and Social Policy
SFB 1342 Globale Entwicklungsdynamiken von Sozialpolitik /
CRC 1342 Global Dynamics of Social Policy

Postadresse / Postaddress:
Postfach 33 04 40, D - 28334 Bremen

Websites:
https://www.socium.uni-bremen.de
https://www.socialpolicydynamics.de

[DOI https://doi.org/10.26092/elib/4362]

[ISSN (Print) 2629-5733]
[ISSN (Online) 2629-5741]

Gefördert durch die Deutsche Forschungsgemeinschaft (DFG)
Projektnummer 374666841 – SFB 1342

DFG Deutsche Forschungsgemeinschaft

Holger Döring

Advancing data infrastructures on democratic representation: An open-source version of ParlGov

SOCIUM · SFB 1342
No. 32

Holger Döring (doering@uni-bremen.de), Affiliated Researcher, SOCIUM Research Center on Inequality and Social Policy, University of Bremen

Acknowledgement:
I thank the two reviewers and Thomas Mustillo for their comments on the paper.

Abstract

This research note presents an open-source version of the ParlGov website and discusses the state of cross-national data on parties, elections, and cabinets. The ParlGov project is a prominent political science data infrastructure with information for EU and OECD democracies since 1900. The data has been publicly available since 2010. ParlGov is an example of a new type of data infrastructure in political science that uses modern software tools, collaborative data collection, and open data principles. In this research note, I introduce a new open-source software of ParlGov that demonstrates how modern software practices enhance the accessibility, transparency, and reliability of political science data. In addition, I assess the contribution of the ParlGov project to advances in comparative political data collection by comparing it with four key datasets on elections and cabinets (EJPR-PDY, PPEG, V-Party, and PAGED). I emphasize the importance and benefits of open science principles, arguing that their full potential has yet to be realized in political science. By adhering to open science principles, ParlGov provides a reference point for future data projects aiming to increase the availability and interoperability of research data in political science.

Zusammenfassung

Dieser Artikel stellt die Open-Source-Version der ParlGov-Website vor und diskutiert vergleichende Daten zu Parteien, Wahlen und Regierungen in der Politikwissenschaft. Das ParlGov-Projekt ist eine wichtige Dateninfrastruktur, die detaillierte Informationen über Demokratien ab 1900 (EU und OECD) bietet und seit 2010 öffentlich zugänglich ist. ParlGov ist ein Beispiel für eine neue Art von Dateninfrastruktur in der Politikwissenschaft, die moderne Software-Tools, kollaborative Datenerhebung und Open-Data-Prinzipien nutzt. In diesem Artikel präsentiere ich eine neue Open-Source-Software von ParlGov, die veranschaulicht, wie moderne Methoden der Softwareentwicklung die Zugänglichkeit, Transparenz und Reliabilität politikwissenschaftlicher Daten verbessern können. Zudem diskutiere ich den Beitrag des ParlGov-Projekts zur Datenerhebung in der Vergleichenden Politikwissenschaft, indem ich es mit vier zentralen Datensätzen zu Wahlen und Regierungen vergleiche (EJPR-PDY, PPEG, V-Party und PAGED). Ich betone die Bedeutung von Open-Science-Prinzipien und argumentiere, dass ihr Potenzial in der Politikwissenschaft noch nicht ausgeschöpft ist. Durch die Anwendung dieser Prinzipien bietet ParlGov einen wichtigen Referenzpunkt für zukünftige Datenprojekte, um die Zugänglichkeit und Interoperabilität von Forschungsdaten in der Politikwissenschaft zu erhöhen.

Contents

1 Introduction

Research data has become more prominent in political science. One notable project is ParlGov, a data infrastructure on political representation. ParlGov covers national election results and cabinet compositions for all EU and most OECD democracies since 1900. An initial version and an agenda for the ParlGov project were presented in Döring (2013), and a detailed coding scheme as well as an extension of the scope of the data were presented in Döring (2016). Here, I introduce a new open-source version of the ParlGov website and assess ParlGov's contribution to political science research by comparing it to other prominent data sources. Throughout this research note, I discuss the role of open science principles in political science data practices and evaluate their current role in shaping research practices.

The open-source software of the new ParlGov site presented in this note follows open science principles. The open science movement has provided a framework and promoted data-sharing, transparency, and replicability. The new ParlGov site offers several advantages over the legacy version used by the project for almost two decades. The software's source code is publicly available under an open-source license and can be built upon. The new implementation uses modern software development approaches such as testing, continuous integration, and a reproducible environment, which enable collaborative software development. Here, I highlight why this approach is valuable for open infrastructures and discuss its advantages.

Over the past two decades, political science research has benefited from technological advancements such as free software, online tools, and public data repositories, making data collection and quantitative analysis more accessible. Comparative political data on democratic representation is now available in structured data formats and digitally accessible, information that was scattered and sometimes still paper-based a decade ago. There is significant progress and substantial potential to further improve political science data. To assess this development, I compare established cross-national data projects with ParlGov and discuss their commitment to open data principles.

The note is structured into three parts. First, I briefly present data types and infrastructures in political science and discuss their relation to open science principles. Second, I present a short history of the ParlGov project and introduce the new open-source version of the ParlGov website. Third, I compare ParlGov's contribution with four other prominent sources and assess current challenges to further improve comparative political data. I conclude by assessing the role of open science in political science data collection and the contributions the paradigm provides for the social sciences.

2 Political science data today

Over the last two decades, we have seen a significant increase in political science research data. Technological innovations such as advanced, often free, research software, online collaboration tools, public data repositories, and more general availability of IT resources have made collecting, structuring, and presenting data significantly more manageable. There is now a broad use of different data types in social science research. Larsen (2024) gives an overview of about six hundred political science datasets, and many of these sources have been created or updated over the last decade. Advances in data-sharing practices and data infrastructures have enabled a wider distribution of research data. The open science principles have provided a framework and motivation for such open data sharing.

There are several examples of new data collection practices and new types of information sources in political science research. Online collaboration tools have enabled large-scale data collection in expert surveys, bringing together political scientists across the globe (e.g. Coppedge et al., 2020; Jolly et al., 2022). Various forms of text data, such as parliamentary speeches, parliamentary questions, bills, and laws, have become prominent large-scale sources in comparative legislative research (cf. Sebők et al., 2023). So-

cial media data has been used to study political communication (e.g. Silva & Proksch, 2022). These three approaches are examples of the increasingly diverse data sources used in political science research today. New forms of data based on images, audio, video, or sensor data are increasingly accessible for social science research and may provide the next wave of research data sources.

The open science framework has highlighted the importance of (cf. UNESCO, 2021):

1. Scientific publications (open access)
2. Open research data
3. Open educational resources
4. Open-source software
5. Open hardware

We have been able to share data more widely through the evolution of norms for data sharing and improvements in research data infrastructure. The open science paradigm has emphasized sound replication standards and highlighted that data sharing and reproducible workflows are essential prerequisites for open social science research (Christensen et al., 2019). Platforms for online collaboration and data sharing have enabled researchers to cooperate more widely (e.g., Dataverse, GitHub, Open Science Framework, Zenodo), and open data licenses (e.g., Creative Commons) have allowed permissive data sharing. It is increasingly the norm to share all elements of a research project (data, documentation, scripts, etc.) in an institutional data repository with an open license, and journals increasingly require an archiving of these research elements upon publication of an article (cf. Basile et al., 2023). Open science promotes transparency by fostering the accessibility of research outputs, and framing publications as a part of research dissemination.

A few attempts have succeeded in establishing open-source software as an element of political science research. Two prominent examples are Quanteda (Benoit et al., 2018), a software for quantitative text analysis, and the QCA package (Thiem & Duşa, 2013), a software to conduct Qualitative Comparative Analysis (QCA). Both packages are widely used within the respective fields of text analysis and qualitative research. These packages rely on the open programming language R and recent data science practices (R Core Team, 2024; Wickham et al., 2023). Open research software to analyze and manage data has become prominent within academic research, and researchers rely significantly more on open-source software. However, political science has only produced a few software infrastructures that contribute to creating open research and has only started using collaborative coding approaches. Overall, adding to open-source research software is still a niche in political research that is underacknowledged and rarely cited (Arel-Bundock & McCrain, 2023).

Political scientists have embraced research data over the last decade, and many datasets are now free to access and contribute to knowledge generation. The status of open-source software is different in political science. It is increasingly used in research, but there are only a few long-term projects and contributions to open-source software development. Overall, the role of open science principles is mixed in political science research. There are some vocal proponents, but it seems that the movement is more vital in other scientific fields.

3 ParlGov: An evolving data infrastructure

3.1 The ParlGov project approach

ParlGov is a data infrastructure for political science that presents data on parties, elections, and cabinets in all EU and most OECD democracies since 1900. The first data version was published in 2010. The project data has been widely used and cited over a thousand times, according to Google Scholar. The concept and an initial version were presented in Döring (2013, p. 166) and suggested four elements:

- A database to structure the data tables
- A website to present the information
- Collaborative data collection
- Scripts to calculate and link additional data

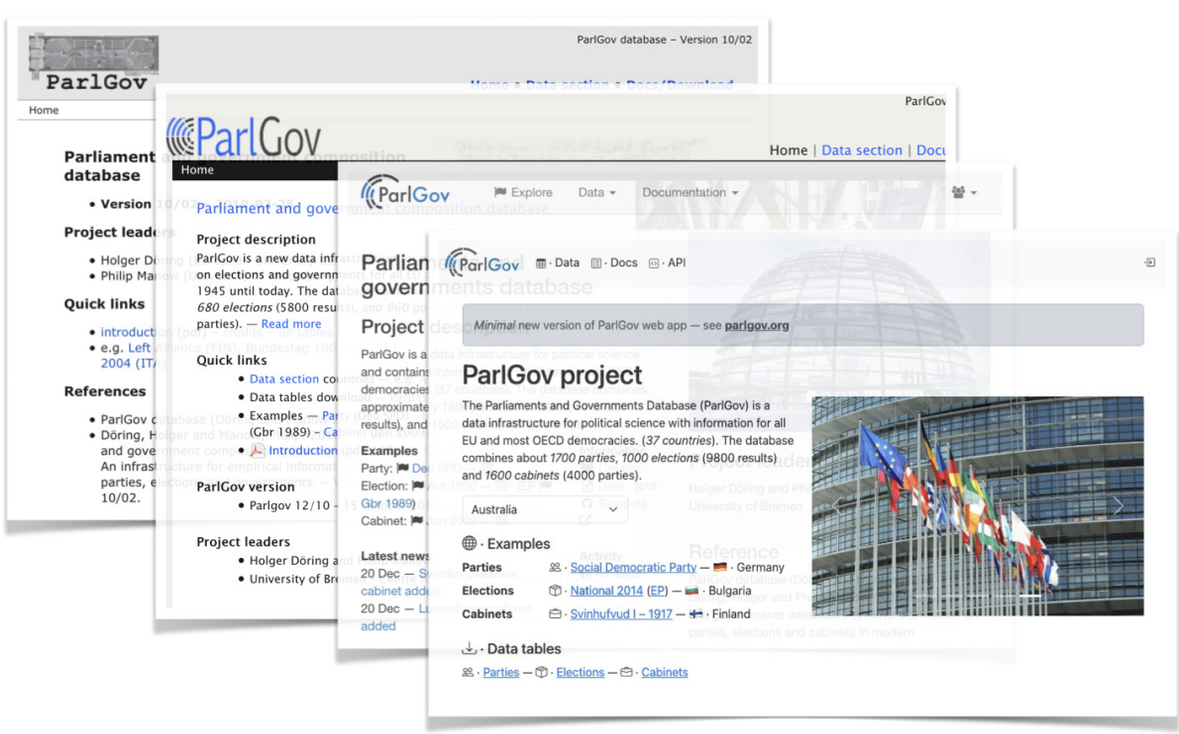
A full version and a detailed specification of the coding rules were published later (Döring, 2016). One of the advantages of ParlGov has been the systematic structuring of information in a database to map the relations between different entities and their presentation on a website. Databases are essential for recording and querying information with a schema that structures data non-redundantly. Their use for political research has been encouraged, but it is still a little-used approach to structure information in political science (cf. Mustillo & Springer, 2015; Weidmann, 2023). In the ParlGov database, for example, parties, elections, election results, cabinets, and cabinet parties are stored in five data tables, and the relations of these tables are defined in the database schema with unique identifiers for parties, elections, and cabinets. Further information is included in other database tables, among them all pages from the codebook.

ParlGov uses explicit criteria to include political parties, mainly based on a threshold of 1% in national parliamentary or EP elections. These parties are systematically linked to their predecessors and successors in the database. In addition, party name changes are systematically recorded. For the parties, this information is presented on a website in addition to each party's election results and cabinet memberships. The various links to the data points are more easily explored with a website, and a database structures the information coherently.

A website presents information recorded in a database in a more accessible format and allows for editing database entries. For ParlGov, three website versions were publicly accessible between 2010 and 2021 (see Figure 1). Due to technical issues, the dynamic site was replaced with a static site in 2021. However, the software's source code for the website used to collect the data in ParlGov was never publicly available. The new open-source version introduced in this note makes the software publicly available, promoting transparency and collaboration.

Over the years, ParlGov has benefited from close scrutiny of its data and valuable feedback. Nevertheless, only a small group of maintainers and data editors provided the extension and update of the data. The project has never succeeded in gaining significant research funding. Hence, the ParlGov project has been less successful in

Figure 1. Screenshots of the dynamic ParlGov website (2007, 2010, 2015, 2024)



establishing collaborative data collection. Finally, the project has provided scripts and examples to analyze the data in a public code repository since 2017 (ParlGov authors, 2024).

3.2 A new implementation of ParlGov

Open-source software is a crucial element of open science principles. Political scientists have mainly remained users of open-source software and have provided only a few contributions, as I have discussed above. So far, the ParlGov project has also not included an open-source implementation of its website that others could build upon. Here, I present such an implementation of the ParlGov website that follows open science principles. The software code follows best practices in software development and is released under an open license. The major open-source software used to develop the new ParlGov site is summarized in Appendix B. The main data tables of the original ParlGov database were migrated for the initial version of the software presented here.

The new version of the ParlGov software provides several improvements over the legacy version used by the project since the late 2000s. The new software is provided in a fully reproducible environment using containers and can be run with minimal configuration (cf. Moreau et al., 2023). It uses software testing to improve the code's robustness and to document implementation details. Software testing is an engineering technique that enhances the overall quality of the software and is important for collaborative software development. It promotes coding standards and best practices, encouraging developers to write cleaner, more maintainable code. Reproducible environments and software testing are important for high-quality open science data infrastructures. Details about these approaches may be beyond the expertise of most social scientists today but are important when cooperating with IT specialists and data experts.

The new implementation has improved the data quality of the ParlGov database. The database schema is now explicitly documented (see Appendix C). Technical data validation checks ensure the consistency of added information. Standard checks ensure that a variable is of a particular type (e.g., number, string, boolean). In fact, new information can only be added if it passes the standard check, and these checks have been used for reimporting the data into the new version. Additional checks provide technical solutions to verify that the data follows the ParlGov coding criteria. For example, one data check determines the inclusion criteria of a party. A party may be included in ParlGov because it won more than 1% vote share in an election, is a cabinet party that formed during a legislative term, or won twice one seat with less than 1% vote share (see ParlGov codebook). The data check determines the criteria and ensures each party meets a ParlGov inclusion rule. Other checks validate, for example, that the sum of the vote shares in an election does not exceed 100%. The data checks implemented in the new version of ParlGov are based on the project's long-term experience and ensure that previous data entry errors are not repeated.

Finally, the new version of ParlGov provides a simple, standards-based API (Application Programming Interface). An API is a structured way to access data that is widely used for integrating and exchanging information. For example, social scientists have used APIs to access social media data in structured formats (cf. Bauer et al., 2024). In the initial version of the new ParlGov site, the API can be used to read data. The approach also allows data to be updated with API requests. For example, the ParlGov API could be extended to automatically add results of recent elections based on a different source that also provides an API. Additionally, the API enables the documentation of ParlGov's database structure using the OpenAPI specification (see spec.openapis.org).

The new version of the ParlGov software offers more fine-grained documentation of several project details. The open-source code, software tests, and a reproducible environment allow others to run and extend the software. These tools may require particular IT expertise beyond quantitative researchers' skill sets. However, this approach, high-quality open science software, provides significant value to large-scale data collaboration in political research and adds to comprehensive data sharing.

4 Cross-national political data: Progress in collaborative data integration

4.1 Comparing five datasets

To understand ParlGov's contribution to political research, it is useful to evaluate it alongside other core datasets, focusing on their coverage, structure, and adherence to open science principles. By now, several data sources present information on parties, elections, and cabinets in Europe and beyond. Table 1 summarizes four main datasets and ParlGov. The table gives an overview of the datasets' country coverage and assesses the principles used by these sources. To compare them, I evaluate whether the datasets include a codebook and visualization, are structured with a database, use institutional collaboration, and archive in an institutional repository. These dimensions are relevant to the open science principles of open research data, open educational resources, and open-source software.

The European Journal of Political Research Political Data Yearbook (EJPR-PDY) has been a prominent source since the 1990s, relying on country experts to cover recent political events (Gomez et al., 2023). The data was first published in articles and later supplemented by data sheets and a website to present the information. The EJPR-PDY provides neither a codebook nor a database structuring of its information. The PPEG – Political Parties, Presidents, Elections and Government (Krause et al., 2024) covers elections and cabinets in 73 countries worldwide. The coding criteria are specified in a codebook, and the data is structured similarly to a database, but a database is not provided. The Varieties of Party Identity and Organization (V–Party) dataset includes election results since 1900 and party positions as well as cabinet compositions since 1970 worldwide (Lindberg et al., 2022). The Party Government in Europe Database (PAGED) gives detailed information on government dynamics in 28 European countries since 1945 (Hellström et al., 2024). Several of these datasets use an interactive visualization of the data on a dynamic website. However, additional scripts and database structuring are rarely included. Furthermore, several of the sources are not archived in an institutional data repository and may not be accessible in the future when a project and its website don't exist anymore.

Compared to the situation a decade ago when ParlGov was initially published, there are now multiple data sources on parties, elections, and cabinets that cover various regions of the world. A decade ago, political scientists relied on different sources to compile individual datasets, but now, this information is provided in structured digital formats. Hence, describing and analyzing political representation with sound data sources is significantly easier today. Much core information, such as election results and cabinet compositions, overlaps between these sources. However, there are also significant differences between the datasets. They vary in the number of countries and years they cover as well as in the depth of information they provide. There are differences in the

Table 1. Major datasets on parties, elections, and cabinets

Dataset	Initial version	Countries	Codebook	Visualization	Database (Schema)	Scripts	Public repository	Institutional collaboration
ParlGov	2010	37	yes	yes	yes	yes	yes	no
EJPR-PDY	1992	37	no	yes	no	no	no	yes
PPEG	2022	73	yes	no	no	no	yes	no
V-Party	2020	169	yes	yes	no	no	no	yes
PAGED	2012	28	yes	yes	no	No	no	yes

Notes: 'Initial version': first public release of the data; 'Codebook': documentation of major coding decisions and inclusion criteria; 'Database': data structuring in a relational database with unique identifiers; 'Visualization': presentation of data in tables or graphs on interactive web pages, 'Scripts': inclusion of scripts to calculate parameters and data usage examples; 'Public repository': dataset versions are archived in an institutional repository with FAIR data sharing principles (e.g., Harvard Dataverse). 'Institutional collaboration': the project cooperates with other researchers and institutions beyond the core group. See Appendix A for details.

institutionalization of the projects, with some projects based on long-term institutional cooperations (EJPR-PDY and PAGED) and others based on a small core group (ParlGov and PPEG). Finally, the degree to which the five projects provide open-source software and open educational resources differs significantly.

Where does ParlGov stand out in this new landscape of data sources? At the data level, it includes some details not covered in the other sources. ParlGov links parties with their predecessor and successor parties, and tracks name changes. It provides an accessible integration of national and European Parliament elections and records members of electoral alliances. Some of the fine-grained coding rules offer detailed information about the effects of electoral systems (e.g., "no-seats" inclusion rule for election results, see codebook). Technically, it is still the only source that systematically structures the information in a database and includes the database in its version releases. Furthermore, some additional information about observations in ParlGov is structured as key/value entries. Additionally, the project provides applications, examples, and scripts in an open code repository of data snippets. Finally, the new open-source implementation of the website that I introduced above is a unique feature compared to other data sources. However, the ParlGov project is institutionally less embedded and was less successful in finding institutional funding than most other sources. It has remained the result of a small core group and a loose net of individual contributors.

4.2 Towards integrating comparative political data

What is missing, and what might future comparative political data on parties, elections, and cabinets look like? The core information that the five sources cover has significant overlap. Basic information such as election results or cabinet compositions is included in all sources, and this information is added to new versions independently. The various independent coding efforts allow us to validate, harmonize, and improve these sources by comparing and revising them systematically. More coordinated efforts to collect this basic information would benefit the discipline. Each project has particular strengths, such as the scope of the country/year coverage, the established institutional network, or the technical approach used. What is missing are systematic efforts to combine this data and the expertise in joint long-term institutional cooperation. Needed is a collaboration similar to the V-Dem network that has led to large-scale collaboration among democratization researchers, including the expertise of country experts, political methodologists, and data experts. Missing is an institutionalized cooperation that brings together domain knowledge, country expertise, data, and technological skills in open online collaboration on the dynamics of democratic representation.

Data on parties, elections, and cabinets should also be better linked. Ideally, each unit (e.g., a party, a politician, a place) is recorded with a unique identifier and linked to other sources. Data creators should make sure that a data set can be linked to other sources by relying on standards (e.g., ISO country codes), community-based identifiers (e.g., Manifesto Project party IDs), or generic identifiers (e.g., Wikidata QID). As I have argued repeatedly in this note, using a database facilitates the coherent usage of identifiers across data types. However, modern approaches based on knowledge graphs may offer a better alternative to structure entities used in comparative political data (c.f. Kejriwal et al., 2021). The main goal is to systematically organize and link data sources on democratic representation, thereby simplifying the integration of existing sources for specific research questions.

A decade ago, Schedler (2012) offered a critical assessment of large-scale cross-national political data collection, pointing out its decentralized and non-institutionalized nature. While some progress has been made, as shown in my comparison of core datasets, these improvements fall short of the scale Schedler had envisioned. Overall, the benefits of open science principles can be more effectively leveraged by political science data projects. In my opinion, greater efforts should be directed toward contributing to open educational resources and open-source software. Providing open educational resources, such as examples

of data usage, replication materials, and tutorials, would enhance the accessibility of the data. Contributions to open-source software, such as libraries and analytical examples, would further reduce barriers to access. Online platforms that support collaborative work, such as GitHub, provide an ideal environment for fostering such cooperation, in alignment with open science principles.

5 Conclusion

Over the last two decades, technological advancements have made it easier for researchers to gather, analyze, and share data on political representation, such as information about parties, elections, and cabinets. The shift from scattered, often inaccessible data sources to structured, digital formats has enhanced the research quality and fostered a collaborative environment where information is shared more widely and transparently. In this research note, I have shown that scholars of democratic representation have successfully contributed to this development through new types of datasets. Nevertheless, the full potential of the open science framework to improve research has yet to be used in political science and its data collection practices. The research community would benefit from more collaboration, better integration of sources, and by relying on collaborative data infrastructures.

This note presented ParlGov's new open-source implementation as an example of an open data infrastructure. By adhering to open science principles, the new ParlGov version offers a more reliable and transparent platform for collecting comparative political data on parties, elections, and cabinets. Modern software development practices, including reproducible environments and testing, have further improved the database's quality and accessibility in the new version. The provision as open-source software allows others to use, extend, and learn from the experience gathered by the ParlGov project over almost two decades.

Looking ahead, I argue that the future of political science data lies in more coordinated and collaborative efforts. Political scientists have made progress by developing key comparative political data in recent years. By combining technical expertise, domain knowledge, and country-specific insights, large-scale collaborations could further advance the study of democratic representation, as the V-Dem project has demonstrated. Key data sources should be better integrated and linked across sources. There should be more effort to establish long-term collaboration. Professional associations, journal editors, and funding agencies should further strengthen their efforts to establish and reward collaborative data collection practices. Open science principles, including open research data, open-source software, and open educational resources, provide important guidelines. This approach promises to further enhance transparency, reproducibility, and the overall quality of political science research, thereby making research results more accessible to a wider audience. I hope the ParlGov project and the new software implementation presented in this article have contributed to this development by demonstrating the feasibility and benefits of open science approaches.

References

Arel-Bundock, V., & McCrain, J. (2023). Software citations in political science. *PS: Political Science & Politics, 56*(3), 398–401. https://doi.org/10.1017/s1049096523000239

Basile, L., Blair, A., & Buckley, F. (2023). Research transparency and openness. *European Political Science, 22*(2), 177–181. https://doi.org/10.1057/s41304-023-00424-x

Bauer, P. C., Landesvatter, C., & Behrens, L. (Eds.). (2024). *APIs for social scientists: A collaborative review.* Retrieved from https://paulcbauer.github.io/apis_for_social_scientists_a_review

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software, 3*(30), 774. https://doi.org/10.21105/joss.00774

Christensen, G., Freese, J., & Miguel, E. (2019). *Transparent and reproducible social science research: How to do open science.* Oakland, CA: University of California Press.

Coppedge, M., Gerring, J., Glynn, A., Knutsen, C. H., Lindberg, S. I., Pemstein, D., ... Wang, Y. (2020). *Varieties of Democracy: Measuring two centuries of political change.* Cambridge, UK: Cambridge University Press.

Döring, H. (2013). The collective action of data collection: A data infrastructure on parties, elections and cabinets. *European Union Politics, 14*(1), 161–178. https://doi.org/10.1177/1465116512461189

Döring, H. (2016). Mapping established democracies: Integrated data on parties, elections and cabinets. *Electoral Studies, 44*, 535–543. https://doi.org/10.1016/j.electstud.2016.07.002

Gomez, R., Mustillo, T., & Vercesi, M. (2023). Political developments and data in 2022: Introducing the 2023 Political Data Yearbook. *European Journal of Political Research Political Data Yearbook, 62*(1), 3–15. https://doi.org/10.1111/2047-8852.12431

Hellström, J., Bergman, T., Lindahl, J., & Sychowiec, M. (2024). *Party Government in Europe Database (PAGED): Basic dataset, Version 2024.06* [Data set]. Retrieved from https://repdem.org

Jolly, S., Bakker, R., Hooghe, L., Marks, G., Polk, J., Rovny, J., Vachudova, M. A. (2022). Chapel Hill Expert Survey trend file, 1999–2019. *Electoral Studies, 75*, 102420. https://doi.org/10.1016/j.electstud.2021.102420

Kejriwal, M., Knoblock, C. A., & Szekely, P. (2021). *Knowledge graphs: Fundamentals, techniques, and applications.* Cambridge, MA: The MIT Press.

Krause, W., Stelzle, R., & WZB Berlin. (2024). *PPEG – Political Parties, Presidents, Elections and Governments, Version 2024v1* [Data set]. Harvard Dataverse. https://doi.org/10.7910/dvn/k5Ajaw

Larsen, E. G. (2024). *PolData: A dataset with political datasets* [Data set]. Retrieved from https://github.com/erik-gahner/PolData

Lindberg, S. I., Düpont, N., Higashijima, M., Kavasoglu, Y. B., Marquardt, K. L., Bernhard, M., ... Seim, B. (2022). *Varieties of party identity and organisation (V–Party) dataset V2* [Data set]. Varieties of Democracy (V–Dem) Project. https://doi.org/10.23696/vpartydsv2

Moreau, D., Wiebels, K., & Boettiger, C. (2023). Containers for computational reproducibility. *Nature Reviews Methods Primers, 3*(1), 1–16. https://doi.org/10.1038/s43586-023-00236-9

Mustillo, T., & Springer, J. A. (2015). Seeing beyond the datasets: The unexploited structure of electoral and institutional data. *International Journal of Social Research Methodology, 18*(4), 349–365. https://doi.org/10.1080/13645579.2014.895288

ParlGov authors. (2024). *ParlGov snippets: Examples for dataset generation and analysis with ParlGov.* Zenodo. https://doi.org/10.5281/zenodo.14282725

R Core Team. (2024). *R: A language and environment for statistical computing.* Vienna, Austria: Foundation for Statistical Computing. Retrieved from https://r-project.org

Schedler, A. (2012). The measurer's dilemma: Coordination failures in cross-national political data collection. *Comparative Political Studies, 45*(2), 237–266. https://doi.org/10.1177/0010414011421308

Sebők, M., Proksch, S.-O., Rauh, C., Visnovitz, P., Balázs, G., & Schwalbach, J. (2023). Comparative European legislative research in the age of large-scale computational text analysis: A review article. *International Political Science Review.* https://doi.org/10.1177/01925121231199904

Silva, B. C., & Proksch, S.-O. (2022). Politicians unleashed? Political communication on Twitter and in parliament in Western Europe. *Political Science Research and Methods, 10*(4), 776–792. https://doi.org/10.1017/psrm.2021.36

Thiem, A., & Duşa, A. (2013). QCA: A package for Qualitative Comparative Analysis. *The R Journal, 5*(1), 87–97. https://doi.org/10.32614/rj-2013-009

UNESCO. (2021). *UNESCO recommendation on open science*. UNESCO. https://doi.org/10.54677/MNMH8546

Weidmann, N. B. (2023). *Data management for social scientists: From files to databases*. Cambridge, UK: Cambridge University Press.

Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). *R for data science: Import, tidy, transform, visualize, and model data* (2nd ed.). Sebastopol, CA: O'Reilly Media.

Appendix

A · Datasets notes (Table 1)

ParlGov — https://parlgov.org

- initial version — see news for 2010 version, first Dataverse release in 2016
- database — SQLite database file includes all project data and documentation
- visualization — data presentation on a website from 2010–2021 and since 2024
- scripts — https://github.com/hdigital/parlgov-snippets
- public repository — https://dataverse.harvard.edu/dataverse/parlgov

EJPR-PDY — https://politicaldatayearbook.com

- initial version — Mackie, Thomas T. 1992. "General Elections in Western Nations during 1990." European Journal of Political Research 21(3): 317–32. https://doi.org/10.1111/j.1475-6765.1992.tb00301.x
- no database — data is provided in complex country spreadsheets
- visualization — graphs in an interactive web application (custom build)
- institutional collaboration — a network of country experts provide country reports

PPEG — https://ppeg.wzb.eu

- initial version — first publication in 2022, unpublished versions since 1999 (see PPEG codebook)
- no database — three different tables are provided for parliamentary elections, presidential elections, and cabinets, but data is not structured with a database schema
- public repository — https://dataverse.harvard.edu/dataverse/ppeg

V-Party — https://www.v-dem.net/data/v-party-dataset

- initial version — Lührmann, Anna, Nils Düpont, Masaaki Higashijima, Yaman Berker Kavasoglu, Kyle L. Marquardt, Michael Bernhard, Holger Döring, et al. 2020. "Varieties of Party Identity and Organisation (V–Party) Dataset V1." doi:10.23696/vpartydsv1
- visualization — graphs in an interactive web application with R Shiny
- institutional collaboration — a network of country experts and expert survey

PAGED — https://repdem.org

- initial version — Andersson, Staffan, and Svante Ersson. 2012. "The European Representative Democracy Data Archive."
- visualization — graphs in an interactive web application with R Shiny
- institutional collaboration — a network of country experts and edited volumes
  - Bergman, Torbjörn, Hanna Back, and Johan Hellström, eds. 2021. Coalition Governance in Western Europe. Oxford, New York: Oxford University Press.
  - Bergman, Torbjörn, Gabriella Ilonszki, and Wolfgang C. Müller, eds. 2020. Coalition Governance in Central Eastern Europe. Oxford, New York: Oxford University Press. https://dx.doi.org/10.1093/oso/9780198844372.001.0001

B · Software

See software versions specified in project files pyproject.toml, requirements.txt, and requirements-dev.txt.

Programming languages and frameworks

» Python: A high-level programming language used in web development, data analysis, and machine learning.
» R: A programming language for statistical computing, data analysis, and graphical representation.
» Tidyverse: A collection of R packages for data science, making data manipulation, visualization, and modeling easier through a consistent, cohesive syntax.

Web development

» Django: A Python web framework for fast, secure, scalable web development.
» Django REST Framework: A toolkit for building Web APIs with Django.
» Bootstrap: A frontend framework for building responsive websites using HTML, CSS, and JavaScript.

Software development

» MkDocs: A static site generator for project documentation in Markdown.
» Pytest: A Python testing framework.
» Ruff: A Python linter and code formatting tool to improve code quality and enforce style guidelines.
» Git: A distributed version control system to track source code changes and collaborate on projects.
» GitHub: A web-based platform for version control, issue tracking, and collaboration.
» Docker: A platform to package applications and their dependencies into containers for consistent deployment.

Global Dynamics of Social Policy
CRC 1342
socium
Research Center on Inequality and Social Policy

C · ParlGov database schema