8. Data and data sets

Rahel Freiburghaus

INTRODUCTION

How to keep track of the tons of publicly shared data sets?[1] How to figure out which open data sets fit the purpose of my own research best?[2] And how to identify which data are valid empirically – and also well received in the field that I am contributing to?

It is actually a good sign that comparativists feel at times even a bit overwhelmed by the sheer off-the-shelf availability of publicly shared data sets that can be utilized for their research endeavours. First, the overwhelming availability of off-the-shelf data sets powerfully demonstrates that the 'open government data' (OGD) movement has left its traces. Championing accessibility and transparency, the OGD offers unprecedented advantages for (political science) researchers. Armed with novel techniques of computational social sciences, OGD, often provided in real time, can be leveraged in extracting insights, patterns, and trends from large and complex data sets. For example, web-scraping tools can automatically navigate websites, collect relevant information, and store it in a structured format. Application programming interfaces offer a particularly structured and reliable way to fetch OGD from one country that can then be merged, using common identifiers such as country codes, to combine data from diverse OGD sources for a comparative study. Text mining and natural language processing techniques help to analyse textual data, such as government reports, policy documents, and legislative texts across a wide range of countries (see Chapter 7).

Second, researchers' overwhelmingness in selecting appropriate data sets also shows that the community is making real progress in fighting the 'replication crisis' affecting the social sciences in the 2010s (e.g., Korbmacher et al., 2023). While in 2014, when the DA-RT symposium was published in *PS: Political Science & Politics*, only about 12 per cent of quantitative research articles pointed to the data and code affiliated with the researcher, the share increased to 31 per cent in 2022 (Rainey et al., 2024). This shows a growing awareness within the community that acknowledges the importance of sharing and providing open access to data, often explicitly mandated by journal publication policies. Only replication materials allow researchers to replicate and verify studies, contributing significantly to the credibility of (political science) research.

For while the increasing off-the-shelf availability of an ever-expanding number of political science data sets is clearly desirable, it also raises new and pressing questions, including those posed at the beginning. (Graduate) students may require additional guidance from their supervisors to navigate data selection processes. Even more, senior researchers may encounter difficulties, such as assessing the data quality and reliability. Notably, the opportunity to download large amounts of well-documented data anytime and anywhere has enabled even well-established data sets to undergo critical review, facing sudden scrutiny in terms of, for example, underlying measurement models or data aggregation steps (see, e.g., Knutsen et al., 2024 or Little & Meng, 2024 for a recent controversy about the advantages and disadvantages of measuring the scope of global democratic backsliding via expert coding in V-Dem data).

However, taking stock of the existing open data sets to allow for informed data selection choices in one's research is no trivial task. The 'data set landscape' is evolving extremely quickly. Any attempt to provide an overview, say, in a handbook chapter, runs the risk of becoming outdated even before the contribution is out in print. A handbook chapter advising in favour of the usage of one particular data set for one specific research goal is hardly helpful if the suggested data set is no longer available, or no longer updated, at the time of reading.

Still, many lectures may recall office-hour debates with (graduate) students where the 'data (selection) question' suddenly arose, and one might have wished for the possibility to assign such an overview in one's own class to help them find a way. Other colleagues may reflect on their own empirical research endeavours, realizing that reaching agreement in the selection of valid measurements and, consequently, suitable data is indeed inevitable.

Acknowledging both the challenges and widespread desire to offer such an overview of the 'data set landscape', this chapter is, to the best of the author's knowledge, the first that dares to attempt to take stock of the plethora of publicly available political science data sets.[3] The chapter intends to deliver on two fronts: comprehensively familiarizing readers with the full breadth of data sets, while also offering in-depth guidance in selecting data suitable for one's own research purposes. Ideally, the overview can help avoid 'path-dependent' rather than goal-orientated data selection processes – to combat 'atomized' usage of data sets a certain department or research group have always been utilizing, so to speak. To achieve this, the chapter first introduces the PolData repository; an open-source tool designed to keep track of the dynamically evolving landscape of political science data sets. Subsequently, and guided by systematic sampling criteria, the chapter delves into a collaboratively curated selection of data sets that are particularly tailored to the comparative study of political institutions. The chapter concludes with a call to action, encouraging readers to leverage available data resources and embark on their own research endeavours.

POLDATA: A TOOL TO KEEP TRACK OF POLITICAL SCIENCE DATA SETS

General Principles in Using PolData

'PolData' is an open-source GitHub repository that provides 'a collection of political data sets' (Gahner Larsen, 2024).[4] In early 2024, it included some 600 political data sets. The PolData GitHub repository is assembled and administered by Erik Gahner Larsen, a Danish data scientist trained in political science. The repository's goal is straightforward: it compiles a vast number of publicly available political science data sets in a spreadsheet. These data sets are categorized across a broad array of specific topics, such as cabinets, citizens, constitutions, parties, political institutions, political elites and politicians, democracy, economics, elections, international relations, media, public policy, as well as political speeches and debates. In the first run, these specific categories serve to narrow down the exhaustive list of political data sets to a sample that thematically aligns with one's research goals.

Once the PolData spreadsheet has been filtered according to the respective categories, the sample of thematically suitable data sets can be explored further. To do so, the PolData GitHub repository provides 'detailed information on the topics, coverage, and availability of the respective data sets' (Gahner Larsen, 2024), along with more technical or metadata information such as the date for the last revision and updates. These comprehensive variables assist the researcher in navigating inevitable trade-offs and, ultimately, in making essential data selection choices, including the following:[5]

- *Data relevance*: Does the scope and focus of the data set align with my research questions? And does the data set cover the variables and parameters relevant to my study?
- *Sampling strategy and coverage*: Are there biases or limitations in the sampling strategy? Is the spatial coverage of the data set suitable for my research context? Is the data cross-sectional or longitudinal? If longitudinal, what time period does the data cover, and does it match the timeframe of my study?
- *Data quality and data source reputation*: What is the quality of the data? Are there known issues with accuracy, completeness, or consistency? Has the data been subject to rigour (peer-review) validation? And how has the data set been received in the community as measured through conventional metrics (with all their strengths and weaknesses), such as the number of citations?
- *Data format and structure*: In what format is the data available, and can it be easily integrated into my analysis tools (e.g., software)? Is the data well structured, or does it require substantial pre-processing and cleaning?
- *Consistency with previous studies in the field*: Has the data set been used in previous studies in the field, specifically by leading authorities? How does my own data analysis approach, build on, or differ from existing work that utilized the same data?
- *Scalability and 'matchability'*: Can the data set accommodate potential future expansions or additional variables (e.g., through matching)? Is the data set scalable to broader research questions?
- *Documentation and metadata*: How well is the data documented? Are there detailed metadata and codebooks available? Does the documentation provide sufficient information for understanding the data structure and variables?

Hence, the PolData GitHub repository enables the researcher not only to narrow down political science data sets into thematic categories, but also to make well-informed, inevitable data selection choices. PolData is a collaborative effort, with scholars and users worldwide invited to provide updates, hints, and suggestions for novel and emergent political data sets to its owner, Erik Gahner Larsen. This approach acknowledges the ever-expanding and dynamically evolving 'data set landscape' and ensures that the repository stays relevant well beyond the present day. Hence, researchers are encouraged to consistently draw on the most recent version of the PolData GitHub repository.

Using PolData for the Comparative Study of Political Institutions

In addition to the general and aforementioned benefits that the PolData GitHub repository brings to political science research, it offers a wealth of opportunities specifically for the comparative study of political institutions. I shall highlight the following main assets that comparativists are most likely to appreciate.

Firstly, the PolData GitHub repository incorporates a category named 'political institutions', allowing users to subsample the repository based on this specific criterion (or search string). In this intentionally generic and generalist category, approximately 25 data sets are included (as of early 2024). Examples of data sets categorized under 'political institutions' include, e.g., the 'Political and Economic Database', also named 'Democracy and Dictatorship

Dataset' (Cheibub et al., 2024), the 'Comparative Political Data Set' (Armingeon et al., 2024), and the 'Database of Political Institutions' (Inter American Development Bank, 2024). These data sets provide numerous variables related to broad topics such as regime types and political systems, (the quality of) democracy, government, as well as potentially relevant controls (e.g., demographics and economy).

Secondly, the PolData GitHub repository allows comparativists to dive deeply into the specific political institutions included in this volume such as political parties and party system institutionalization, (digital) media and (digital) media systems, political executives (cabinets), parliaments, or constitutions – to name just a few.

TAKING STOCK: A SHORT INTRODUCTION OF RELEVANT DATA SETS

Selection Approach: Which Data Sets Merit Closer Introduction?

With an abundance of data sets providing a wealth of variables on political institutions, both in a general sense and with a focus on specific aspects, the decision on which data sets to introduce in greater depth poses a challenge. Which data sets tailored to the comparative study of political institutions merit putting in the spotlight, and why?

A straightforward sampling approach may be to check metrics like the number of citations a given data set has received so far (probably combined with the 'impact factor' of the outlets in which they have been cited). Citations provide a quantitative measure of the visibility of and impact of a data set within the academic community, and well-cited data sets may have an established reputation for reliability and relevance. However, there are also considerable downsides in using the number of citations to determine which data sets 'deserve' a closer introduction. Highly cited data sets may be biased toward popular topics, potentially neglecting valuable but less mainstream data sets. Also, citation numbers may be subject to manipulation or inflation (e.g., by self-citations), leading to inaccurate assessments of the data set's impact. Even more problematic is the publication lag: the citation count may not reflect the current impact of a data set as there could be a publication lag between the creation of the data set and subsequent research publications. Hence, the focus on citation numbers might lead researchers to overlook newer or emerging data sets that have not had sufficient time to accumulate citations.

To avoid these fallacies, the approach taken at hand to identify the data sets that receive more in-depth introduction relies on a deliberately forward-looking peer assessment.[6] A well-grounded selection of leading authorities around the globe have been consulted and asked which data sets they, based on their own expertise, deemed to be the most relevant in the current and future comparative study of political institutions.

Brief Data Set Profiles

In the following sections, a total of 13 data sets are introduced, selected based on the aforementioned expert-guided approach. Each data set is presented in a dedicated table, aiming to offer a concise profile that captures key features at a glance (Tables 8.1–8.13). The tables are designed to facilitate the quick identification and assessment of the major characteristics of each data set, enabling easy comparisons across the various data set options. Each brief data set profile, showcased in its respective table, encompasses the seven dimensions of comparison:

- *Data type*: What kind of data does the data set provide? What is the aggregation level and/or the unit of analysis (e.g., country-year)?
- *Method of data collection*: How was the data collected? Is the data original, gathered through e.g., expert coding? Or is the data compiled from secondary sources such as official data sources (e.g., national statistical offices) and established secondary data sets?
- *Number of variables*: How many variables or indicators does the data set include?
- *Country coverage*: How many countries does the data set cover? What is the geographical scope of the data set?
- *Temporal coverage*: What is the time span covered by the data set?
- *Potential focus of analysis*: What political institutions does the data set cover, and which political institutions discussed in this volume is it particularly suitable for?
- *Particularly notable features*: What are the unique selling points of the data set? What notable features make it particularly compelling and handy for own research endeavours? Also, what other data sets can it be combined or merged with?

Needless to say, with these seven dimensions, it is impractical to provide a full account of the large and complex data sets presented in each table. However, the intention of this section is not to exhaustively introduce each data set in all its facets. Instead, these brief data set profiles shall assist (graduate) students and researchers in making informed data selection choices. They aim to offer a better understanding of what these data sets entail and, once interest is sparked, invite readers to explore each data set more extensively on their own.

Table 8.1 *Varieties of Democracy*

Data type	Multidimensional and disaggregated; country-year
Method of data collection	Expert coding (e.g., of project managers, research assistants, country coordinators, and 4000 country experts)
Country coverage	All countries in the world
Time coverage	1789–2023[a]
Potential focus of analysis	● Regime types ● Elections and electoral systems ● Party systems, party system institutionalization, and party politics ● Governments and political executives ● Parliaments ● Regional and local political institutions ● Direct democracy ● Judiciary
Potential focus of analysis	● Media and media systems ● Civil society (including social movements)
Particularly notable features	● Includes the world's most comprehensive and detailed democracy ratings ● Reflects the complexity of the concept of democracy via five high-level principles of democracy (electoral, liberal, participatory, deliberative, egalitarian) ● Features a 'Historical V-Dem' data section with merged time series V-Dem data extending all the way from 1789 to the present

Note: [a] Coding starts from when a country first enjoyed at least some degree of functional and/or formal sovereignty. The information refers to the data set version, 'V-Dem (v14)', published in March 2024.
Source: Coppedge et al. (2024).

Table 8.2 *Power Diffusion and Democracy*

Data type	Country-year
Method of data collection	Data infrastructure that compiles information from official data sources (e.g., from national statistical offices) with own calculations of well-established indices
Country coverage	61 countries
Time coverage	1990–2022
Potential focus of analysis	● Political-institutional configurations ● Types of democracy ● Forms of government ● Electoral systems and elections ● Party systems, party system institutionalization, and party politics ● Governments and political executives ● Parliaments ● Bicameralism ● Direct democracy ● Federalism and federal systems ● Judiciary
Particularly notable features	● Encompasses a broad country sample beyond OECD member states ● Introduces novel country scores on four distinct dimensions, categorizing different types of democracy (proportional power diffusion, decentral power diffusion, presidential power diffusion, direct power diffusion; see Bernauer & Vatter, 2019) ● Includes replicated ‘Lijphart-styled scores’ that address critiques of Lijphart’s (2012) seminal typology while adhering to his fundamental approach ● Offers comprehensive replication code for analysing the effects of various types of democracy on multiple dimensions, including public policy, corruption, economic inequality, and satisfaction with democracy

Note: The information refers to the updated data set version published in January 2024.
Source: Bernauer and Vatter (2024).

Table 8.3 *Comparative Political Data Set*

Data type	Country-year
Method of data collection	Data infrastructure that compiles information from official data sources (e.g., from national statistical offices) with own calculations of well-established indices
Country coverage	36 OECD and/or EU member states
Time coverage	1960–2022[a]
Potential focus of analysis	● Political-institutional configurations ● Electoral systems and elections ● Party systems, party system institutionalization, and party politics ● Governments and political executives ● Parliaments
Particularly notable features	● Features a supplement with detailed government composition data (e.g., party composition, reshuffles, duration, reason for termination, and the type of government) ● Allows to assign programmatic scores to parties, governments, or parliaments based on ‘The Manifesto Project Database’ and ‘Chapel Hill Expert Survey’ ● Provides a wealth of data related to the economy (e.g., openness of the economy, macroeconomic data, labour force data, and public expenditures)

Note: [a] Political data is only collected for the democratic periods. The information refers to the data set version published in August 2024.
Source: Armingeon et al. (2024).

Table 8.4 Quality of Government

Data type	Country-year
Method of data collection	Data infrastructure that compiles information from official data sources (e.g., from national statistical offices) with original data sets (expert coding)
Country coverage	193 UN member states
Time coverage	1946–2023
Potential focus of analysis	● Regime types ● Political-institutional configurations ● Types of democracy ● Elections and electoral systems ● Party systems, party system institutionalization, and party politics
Potential focus of analysis	● Interest groups ● Social movements and civil society ● (Digital) media and (digital) media systems ● Political executives ● Bureaucracy and public administration ● Regional political institutions ● Judiciary and judicial behaviour
Particularly notable features	● Offers a data finder, variable search tools, and some visualization tools to help you find and understand the QoG data faster ● Provides regionally specified data sets, including 'QoG OECD' and 'EU Regional data', offering a focused perspective on OECD member states and the EU ● Presents the 'European Quality of Government Index', containing information on subnational governance in Europe derived from a comprehensive pan-European survey on citizen perceptions and experiences with public services ● Compiles the 'Environmental Indicators Dataset', major indicators measuring environmental performance of countries over time (e.g., presence and strictness of environmental policies, level of pressure on the environment, and public opinion on environmental matters) ● Incorporates a wealth of data on public policy (e.g., education, energy and infrastructure, environment, gender equality, health, labour market, media, migration, and economy), offering an extensive resource for studying the effects of, or correlations with, political institutions

Note: The information refers to the data set version 'The QoC Standard Dataset 2024 (v2024)', published in January 2024.
Source: Teorell et al. (2024).

Table 8.5 *Our World in Data*

Data type	Country-year
Method of data collection	Data infrastructure that compiles information from official data sources (e.g., from national statistical offices) with well-established data sets
Country coverage	All countries of the world
Time coverage	1789–2024
Potential focus of analysis	● Regime types ● Political-institutional configurations ● Types of democracy ● Elections and electoral systems
Particularly notable features	● Compiles the most state-of-the-art data sets to assess democracy and the quality of democracy (e.g., V-Dem/Episodes of Regime Transformation, Lexical Index, Polity, Freedom House, Bertelsmann Transformation Index, and Economist Intelligence Unit) ● Offers a plethora of interactive visualization tools, facilitating exploration of data online, creation of ready-publishable charts, and the ability to filter tailored data sets based on criteria such as geographical regions and time frames

Note: The information refers to the data set version 'Democracy', published in January 2024.
Source: Herre et al. (2024).

Table 8.6 *Veto Points Data Set*

Data type	Country-year
Method of data collection	Expert coding, own calculations of well-established indices
Country coverage	49 countries
Time coverage	1940/1990–2018[a]
Potential focus of analysis	● Political-institutional configurations ● Political executives ● Legislatures and legislative politics
Particularly notable features	● Informs about the existence of veto points in the political executive, legislative chambers, judiciary, electorate, and territory (e.g., federalism) through both quantitative and qualitative data ● Offers detailed fact sheets for every country in the sample, providing information on formal rules for core political institutions (e.g., rule of investiture, rule dissolution, and rule of decision-making), describing the political institutions' role in the legislative process, and documenting institutional reforms ● May be linked to some of the main data sets on political parties such as the Manifesto Project, ParlGov, or V-Party

Note: [a] For some countries, the data go further back in time. The information refers to the data set version 'VAPS Veto Points Data Set (v1)', published in March 2022.
Source: Immergut et al. (2022).

Table 8.7 *Comparative National Elections Project*

Data type	Individual-level
Method of data collection	Survey data; archival data; experimental data; observational data
Country coverage	30 countries
Time coverage	1990–2023
Potential focus of analysis	● Electoral systems and elections ● Public opinion (e.g., satisfaction with democracy and democratic participation, political attitudes such as party identifications, political knowledge, and subnational political identities)
Particularly notable features	● Merges the common core questions from 69 individual surveys across many of the world's democracies and several non-democracies holding elections in a single file

Note: The information refers to the data set version published in December 2023.
Source: Beck and Gunther (2024).

Table 8.8 *Varieties of Party Identity and Organization*

Data type	Multidimensional and disaggregated; party-election year units
Method of data collection	Expert coding (e.g., of project managers, research assistants, country coordinators, and 711 country experts)
Country coverage	3,467 political parties across 3,151 elections in 178 countries
Time coverage	1900–2019
Potential focus of analysis	● Party systems and party system institutionalization
Particularly notable features	● Allows examination of both the policy positions and organizational structures of political parties across the world ● Time series of countries and/or variables can be explored online for pre-analysis purposes through 'V-Party Explorer' ● Provides a harmonized English name (*v2paenname*) that may also be used for English-language translations of party names

Note: The information refers to the data set version 'V-Party (v2)', published in February 2022.
Source: Lindberg et al. (2022).

Table 8.9 ParlGov

Data type	Party-election year units
Method of data collection	Data infrastructure that combines official data sources (e.g., from national statistical offices)
Country coverage	1,700 political parties, 1,000 elections, and 1,600 cabinets across all EU and most OECD democracies (37 countries)
Time coverage	1945–[a]
Potential focus of analysis	● Party systems and party system institutionalization ● Governments and political executives ● Parliaments
Particularly notable features	● Enables visualization of ParlGov data through an interactive dashboard ● Provides a comprehensive GitHub repository with R snippets for reproducing up-to-date figures such as cabinet maps illustrating left/right positions across Europe ● Assists in identifying official sources for each country through the reference in the codebook ● May be merged with the 'Parties and Elections in Europe' data set to incorporate regional elections (Nordsieck, 2024)

Note: [a] For some countries, the data go further back in time. The information refers to the 'ParlGov development version', published in November 2023.
Source: Döring et al. (2023).

Table 8.10 WhoGov

Data type	Country-year (effective month: July)
Method of data collection	Expert coding
Country coverage	56,063 cabinet members in 177 countries (i.e., all countries with a population of more than 400,000 citizens)
Time coverage	1966–2023
Potential focus of analysis	● Political executives ● Bureaucracy and public administrations
Particularly notable features	● Provides biographic information, such as gender and party affiliation, on cabinet members that allows for the exploration of, e.g., the share of female cabinet members, cabinet turnover, and ministerial experience ● Offers potential merging with the 'Paths to Power' data set to incorporate additional variables on social background, education, and occupation of cabinet members ● May be merged with the 'Global Dataset on Political Leaders (1945–2020)' (Herre, 2023) that identifies the economic ideologies and political parties of heads of government

Note: The information refers to the data set version 'WhoGov (v2)', published in July 2022.
Source: Nyrup and Bramwell (2020, 2022).

Table 8.11 Parline

Data type	Country-month, complemented with other aggregation levels
Method of data collection	Official data sources with the cooperation of national parliaments providing and checking the data through a network of 'Parline correspondents' for the Inter-Parliamentary Union
Country coverage	180 countries (all Inter-Parliamentary Union members)
Time coverage	1933–[a]
Potential focus of analysis	● Legislatures and legislative politics
Particularly notable features	● Provides data on every parliamentary chamber, with the data presented in different sections covering the different aspects of each parliamentary chamber's composition, structure, and functioning (e.g., parliamentary mandate, law-making, oversight, budget competences, working methods, administration, and specialized bodies) ● Facilitates online comparison of parliaments through an interactive visualization tool, allowing users to explore and analyse various aspects across different parliamentary systems

Note: [a] For some countries, the data goes further back in time. The information refers to the data set version 'Parline' (last accessed in February 2024).
Source: Inter-Parliamentary Union (2024).

Table 8.12 Comparative Constitutions Project

Data type	Country-year (chronology of constitutional events[a])
Method of data collection	Expert coding
Country coverage	All independent states
Time coverage	1789–2021
Potential focus of analysis	● Regime types ● Federalism and federal systems ● Constitutions and the rule of law ● Judiciary
Particularly notable features	● Enhances the chronology of constitutional events by incorporating a core data set detailing general characteristics of national constitutions (e.g., length and formal amendment procedure) ● Includes the full texts of constitutions and nearly every constitutional event, offering valuable resources for text-as-data techniques ● Enables the construction of various categorical measures, including those related to state structure (e.g., federalism) or colonial past, often necessary as controls

Note: [a] A 'constitutional event' is defined as 'any formal change to a country's constitution, including adoption, amendment, suspension, or reinstatement' (Elkins & Ginsburg, 2022, p. 4). The information refers to the data set version 'Chronology of Constitutional Events (v4)', published in October 2022.
Source: Elkins and Ginsburg (2022).

Table 8.13 Regional Authority Index

Data type	Country-year and region-year[a]
Method of data collection	Expert coding
Country coverage	96 countries and respective subnational government levels with an average population of 150,000 or more
Time coverage	1950–2021
Potential focus of analysis	● Federalism and federal systems ● Regional political institutions
Particularly notable features	● Available in five aggregations, offering annual scores for each region or regional tier in a country, the most authoritative regional tier, each country, annual scores for each metropolitan regions, or indigenous governance ● Presents a comprehensive measure of regional authority across ten dimensions; this measure can be disaggregated into a 'self-rule' score (assessing subnational autonomy) or a 'shared rule' score (gauging subnational participation in upper-level decision-making processes) ● May be merged with the 'Observatory on Regional Democracy' (Schakel, 2024), collecting a swath of data on elections, parties, parliaments, governments, and political institutions at the subnational level

Note: [a] Note that the unit of analysis in the data set is the individual region, defined as 'a jurisdiction between national government and local government' (Hooghe et al., 2016, p. 14). Readers interested in local governments and the local level may refer to the 'Local Autonomy Index' instead, applying the comprehensive and adjusted RAI measurement framework to 57 countries (1990–2020; see Ladner et al., 2019, 2023). The information refers to the data set version 'Regional Authority Index (v3)', published in April 2021.
Source: Hooghe et al. (2016); Schakel (2024).

CONCLUSION

The OGD movement, coupled with the political science community's strong commitment to openly share data and full replication materials in response to the 'replication crisis', has created an environment that is both incredibly exciting and, to some extent, overwhelming. It is exciting because the ever-growing availability of off-the-shelf data sets catalyses the comparative study of political institutions. Novel techniques in computational social sciences provide increasingly sophisticated tools to handle vast amounts of diverse data, enabling investigations into the dynamics, interactions, and effects of political institutions from a deliberate comparative perspective. However, the proliferation of political science data sets can also be overwhelming, presenting new challenges for (graduate) students, researchers, and other colleagues in the field. Navigating the data selection process becomes complex, requiring informed choices regarding data quality and the reliability and validity of the measurements used.

In order to guide fellow comparativists in their data selection choices, the chapter therefore presented brief data set profiles of a total of 13 data sets that experts deem particularly relevant for the current and future comparative study of political institutions. Tabular overviews help familiarize readers at a glance with how the data sets differ in terms of data type, method of data collection, number of variables, country and temporal coverage, and the analyses of political institutions they allow for. It is the author's hope that these brief data set profiles serve as an invitation for readers to leverage the available data resources and embark on their own cutting-edge research endeavours that push the boundaries of the field.

NOTES

1. This chapter has been published Open Access by courtesy of the Open Access Publication Fund at the University of Bern (Switzerland).
2. I would like to thank Tarik Abou-Chadi, Julian Erhardt, Martina Flick Witzig, Mirco Good, Carl Henrik Knutsen, Sarah Kuhn, Lucas Leemann, Pierre Lüssi, Jonas Schmid, Isabelle Stadelmann-Steffen, and Adrian Vatter for providing excellent advice and helpful suggestions on what data sets on the comparative study of political institutions should be included in this chapter.
3. Neither generalist handbooks on the entire political science discipline (e.g., Goodin, 2009), including those specifically dealing with political institutions (e.g., Gandhi & Ruiz-Rufino, 2015; Rhodes et al., 2006), nor handbooks on political science methodology (e.g., Box-Steffensmeier et al., 2008) have featured a chapter introducing and comparing a selection of state-of-the-art data sets.
4. https://github.com/erikgahner/PolData.
5. Note that these questions guiding data selection processes have general relevance but are intentionally tailored to the data selection process based on the PolData GitHub repository. In other instances additional considerations may be relevant, such as ethical considerations, cost and resource considerations, and data governance and compliance issues (e.g., issues related to data ownership, rights, or privacy).
6. Needless to say, the forward-looking peer assessment to sampling data sets comes with its own disadvantages, such as homogeneity, limited representativeness, availability bias, or limited scope. For the purpose at hand, however, the advantages associated with such an approach outweigh the downsides.

REFERENCES

Armingeon, K., Engler, S., Leemann, L., & Weisstanner, D. (2024). Comparative Political Data Set 1960–2022. *University of Zurich/Leuphana University/University of Lucerne*. https://cpds-data.org/

Beck, R., & Gunther, R. (2024). *Comparative National Election Project*. Ohio State University. https://u.osu.edu/cnep/

Bernauer, J., & Vatter, A. (2019). *Power Diffusion and Democracy: Institutions, Deliberation and Outcomes*. Cambridge University Press.

Bernauer, J., & Vatter, A. (2024). *Power Diffusion and Democracy*. GitHub. https://github.com/julianbernauer/powerdiffusion/tree/master

Box-Steffensmeier, J. M., Brady, H. E., & Collier, D. (Eds.). (2008). *The Oxford Handbook of Political Methodology*. Oxford University Press.

Cheibub, J. A., Gandhi, J., & Vreeland, J. (2024). *The Democracy and Dictatorship Dataset (DD/PACL/ACLP/CGV)*. GitHub. https://xmarquez.github.io/democracyData/reference/pacl.html

Coppedge, M., Gerring, J., Knutsen, C. H., Lindberg, S. I., Teorell, J., Altman, D., … Ziblatt, D. (2024). *Dataset v14* (Country-Year/Country-Date). V-Dem. https://v-dem.net/data/the-v-dem-dataset/country-year-v-dem-fullothers-v14

Döring, H., Quaas, A., & Manow, P. (2023). *Parliaments and Governments Database (ParlGov)*. ParlGov. www.parlgov.org/data-info/

Elkins, Z., & Ginsburg, T. (2022). Chronology of Constitutional Events/Characteristics Data. *Comparative Constitutions Project*. https://comparativeconstitutionsproject.org/download-data/

Gahner Larsen, E. (2024). *PolData*. GitHub. https://github.com/erikgahner/PolData

Gandhi, J., & Ruiz-Rufino, R. (Eds.). (2015). *Routledge Handbook of Comparative Political Institutions*. Routledge.

Goodin, R. E. (Ed.). (2009). *The Oxford Handbook of Political Science*. Oxford University Press.

Herre, B. (2023). Identifying Ideologues: A Global Dataset on Political Leaders, 1945–2020. *British Journal of Political Science*, *53*(2), 740–8.

Herre, B., Ortiz-Ospina, E., & Roser, M. (2024). *Democracy*. Our World in Data. https://ourworldindata.org/democracy#introduction

Hooghe, L., Marks, G., Schakel, A. H., Chapman Osterkatz, S., Niedzwiecki, S., & Shair-Rosenfield, S. (2016). *Measuring Regional Authority*. Oxford University Press.

Immergut, E., Abou-Chadi, T., Burlacu, D., Orlowski, P., Roescu, M., & Wegemann, M. (2022). *VAPS Veto Points Dataset*. Cadmus European University Institute Research Repository. https://cadmus.eui.eu/handle/1814/73926

Inter American Development Bank. (2024). *Database of Political Institutions 2020 (DPI2020)*. Inter American Development Bank. www.iadb.org/en/sharing-knowledge/research-idb/research-datasets/database-political-institutions

Inter-Parliamentary Union. (2024). *Parline*. https://data.ipu.org/content/about-open-data-platform

Knutsen, C. H., Marquardt, K. L., Seim, B., Coppedge, M., Edgell, A. B., Medzihorsky, J., Pemstein, D., Teorell, J., Gerring, J., & Lindberg, S. I. (2024). Conceptual and Measurement Issues in Assessing Democratic Backsliding. *PS: Political Science & Politics*, *57*(2), 166–77.

Korbmacher, M., Azevedo, F., Pennington, C. R., Hartmann, H., Pownall, M., Schmidt, K., … Evans, T. (2023). The Replication Crisis Has Led to Positive Structural, Procedural, and Community Changes. *Nature Communications Psychology*, *1*(1), 1–13.

Ladner, A., Keuffer, N., Baldersheim, H., Hlepas, N., Swianiewicz, P., Steyvers, K., & Navarro, C. (2019). *Patterns of Local Autonomy in Europe*. Palgrave Macmillan.

Ladner, A., Keuffer, N., & Bastianen, A. (2023). Local Autonomy around the World: The Updated and Extended Local Autonomy Index (LAI 2.0). *Regional & Federal Studies*. https://doi.org/10.1080/13597566.2023.2267990

Lijphart, A. (2012). *Patterns of Democracy: Government Forms and Performance in Thirty-Six Countries*. Yale University Press.

Lindberg, S. I., Düpont, N., Higashijima, M., Berker Kavasoglu, Y., Marquardt, K. L., … Seim, B. (2022). *Varieties of Party Identity and Organization (V-Party; v2)*. V-Dem. https://doi.org/10.23696/vpartydsv2

Little, A. T., & Meng, A. (2024). What Do We Know about Democratic Backsliding? *PS: Political Science & Politics*, *57*(2), 149–61.

Nordsieck, W. (2024). *Parties and Elections in Europe*. Parties and Elections. http://www.parties-and-elections.eu/

Nyrup, J., & Bramwell, S. (2020). Who Governs? A New Global Dataset on Members of Cabinets. *American Political Science Review*, *114*(4), 1366–74.

Nyrup, J., & Bramwell, S. (2022). *Who Governs? A New Global Dataset on Members of Cabinets (2.0)*. Nuffield College, University of Oxford. https://politicscentre.nuffield.ox.ac.uk/whogov-dataset/download-dataset/

Rainey, C., Roe, H., Wang, Q., & Zhou, H. (2024). *Data and Code Availability in Political Science Publications from 1995 to 2022*. SocArXiv Papers. https://osf.io/preprints/socarxiv/a5yxe

Rhodes, R. A. W., Binder, S. A., & Rockman, B. A. (Eds.). (2006). *The Oxford Handbook of Political Institutions*. Oxford University Press.

Schakel, A. (2024). *Regional Authority Index*. University of Bergen. www.arjanschakel.nl/index.php/regional-authority-index

Teorell, J., Sundström, A., Holmberg, S., Rothstein, B., Alvarado Pachon, N., Dalli, C. M., Lopez Valverde, R., & Nilsson, P. (2024). *The Quality of Government Standard Dataset*. Quality of Government Institute. www.gu.se/en/quality-government/qog-data/data-downloads/standard-dataset