The Oxford Handbook of Comparative Judicial Behaviour

(In Progress)

Lee Epstein (ed.) et al.

https://doi.org/10.1093/oxfordhb/9780192898579.001.0001

Search in this book

CHAPTER

Observational Databases 🔓

Benjamin G. Engst, Thomas Gschwend

Abstract

Why do we need databases in research on comparative judicial behaviour? This chapter argues that comprehensive assessments of common models on judicial behaviour require data on decisions, judges, and environmental characteristics. An expert survey shows that data on these characteristics was often published in rectangular datasets focusing on courts in the United States or specific International Courts mostly allowing assessments of the attitudinal model of judicial behaviour. Databases on courts in other regions that allow the assessment of judicial behaviour have been published more recently. They allow for the modelling of different entities—such as information on decisions and information on judges—and establishing the relationships between them; for example, linking specific judges to specific decision outcomes. The advantage of designing databases is to summarize clearly specified concepts in parsimonious and flexible ways without producing redundancies when collecting data. Comparative judicial databases including information on the action of multiple courts are scarce. However, scholars of judicial politics can learn from existing comparative projects such as the Comparative Study of Electoral Systems (CSES) to design a Comparative Study of Judicial Behaviour (CSJB).

Keywords: datasets, databases, judicial behaviour, parsimony, flexibility, specificity, non-redundancy, Constitutional Court Database

Subject: Legal System and Practice, Comparative Law, Law

Series: Oxford Handbooks

Collection: Oxford Handbooks Online

1. Introduction

Why do we need databases in research on comparative judicial behaviour? Databases facilitate the assessment of various key aspects of democracy, such as judicial independence and the judicialization of politics. In order to evaluate competing behavioural models, data of a similar nature but partially different content is necessary. Data comprises 'systematically collected elements of information about the world' (King, Keohane, and Verba 1994, 23) commonly structured in rectangular datasets. Consider the legal model of judicial behaviour which is built on the premise that judges are constrained by constitutional norms and their legal training when making decisions (Parcelle Jr, Curry, and Marshall 2011, 29–32). To assess this model, we need to understand the ways in which opinions are written which requires data on decision characteristics. On the contrary, the attitudinal model assumes that the individual preferences of the judges matter when they make decisions (Parcelle Jr, Curry, and Marshall 2011, 34–39). Data on decision characteristics is essential to the attitudinal model as well but a comprehensive assessment requires identifying the relationship between decision characteristics and characteristics of individual judges, for example, their potential partisan leaning or voting pattern. Databases are helpful tools to establish such relationships and enable scholars and practitioners to link datasets of different content and complexity. Specifically, databases model real-world objects—both entities (e.g. judges, the actor who nominated them, and judicial decisions) and relationships (e.g. 'Chief Justice Roberts' who was 'nominated by President George W. Bush' voted 'with the majority in *Dobbs v Jackson Women's Health*')—and capture structure between them (Foster et al. 2021, 71).

This chapter is no technical guide for computing databases but the research cited here provides a point of departure for scholars eager to develop comprehensive and comparative databases in judicial politics. This is of great importance because courts do not operate in a vacuum. To understand the wider implications of judicial behaviour we require data structures that situate the judiciary within different societies and polities.

In this chapter, we argue that databases in judicial politics are necessary to connect different characteristics to assess competing models of judicial behaviour efficiently. To make this argument we outline the characteristics that we require data to cover in order to assess these behavioural models. The results from an expert survey show that existing data structures already include these characteristics to varying degrees (section 2). However, the data structures are developed for single cases and are seldom comparative in nature. Thus, they do not allow for an analysis of judicial behaviour across countries or regions. Therefore, we continue outlining important features of databases using one of the projects as an example (section 3). This allows us to present ideas on how to develop an international comparative data infrastructure project, the 'Comparative Study of Judicial Behaviour' (CSJB), accounting for the identified features (section 4). The concluding section summarizes our discussion.

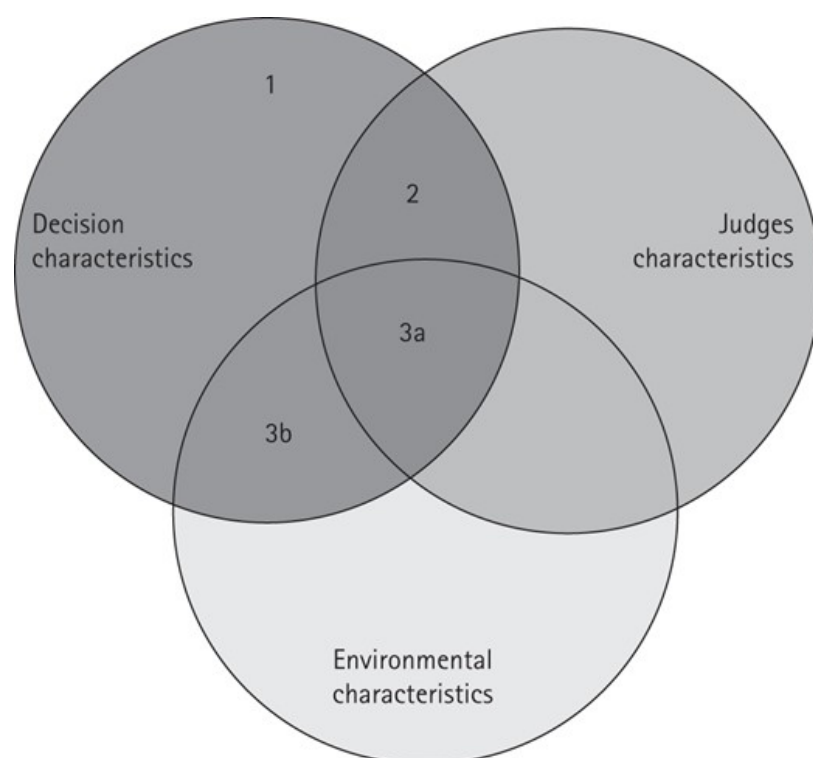## 2. Datasets and Databases to Study Judicial Behaviour

In this section, we argue that the different models of judicial behaviour require data summarizing different characteristics of the judicial process. An expert survey we conducted shows that there are different data structures that already link multiple of those characteristics. However, databases that allow to study judicial behaviour comparatively across countries or regions are scarce.

The most common models used to explain judicial behaviour are the legal, attitudinal, and strategic model (see part on Approaches to Judging in this volume; Epstein, Sǎdl, and Weinshall 2022). Each model is built on different premises which is why assessments require different information. The legal model assumes that judges make decisions based on their legal training using jurisprudential methods (Parcelle Jr, Curry, and Marshall 2011, 29–32). Accordingly, judges are constrained by the legal system (George and Epstein

1992). Therefore, if judges wish to incorporate their preferences into their decisions, then they can only do so through the means of decision characteristics; for example, citing certain legal sources to derive at a particular verdict. Decision characteristics are a necessary precondition to contextualize judicial behaviour. Without decisions there is no judicial behaviour. Thus, we argue that a comprehensive assessment of the models of judicial behaviour always requires decision characteristics. The attitudinal model assumes that judges are unconstrained actors who follow their sincere preferences when making decisions (Schubert 1965; Rohde and Spaeth 1976; Segal and Cover 1989; Parcelle Jr, Curry and Marshall 2011, 34–39). It is plausible to assume that attitudes are shaped by a number of formative events, such as judges' age or the places judges were educated. This requires to also collect a number of judges' characteristics in addition to various preference measures in order to assess the attitudinal model. Finally, the strategic model extends upon the attitudinal model adding the assumption that judges are constrained by the decision-making environment (Murphy 1964; Epstein and Knight 1998; Parcelle Jr, Curry, and Marshall 2011, 39–49). Thus, judges account for fellow judges, political, or societal actors as well as the institutional setting when making decisions. These environmental characteristics structure judicial behaviour and constrain judges in voicing their preferences; for example, when judges confront declining public support for their decisions and alter the clarity of writing in their decisions (Staton and Vanberg 2008).

Figure 1 visualizes the considerations made. Decision characteristics are necessary to assess the legal model of judicial behaviour (zone 1). These characteristics are also necessary to assess all other models. When individual judicial characteristics are added then assessments of the attitudinal model become possible (zone 2). Finally, analysing the strategic model requires environmental characteristics either coupled with judicial characteristics (zone 3a) or not (zone 3b).

Figure 1

Overlapping entities necessary to assess judicial behaviour.

Figure 2 highlights examples of data structures that include the different characteristics. The first column indicates the behavioural models that can be studied with the characteristics included in the data. The numbers next to the models speak to the overlapping zones mentioned in Figure 1. The examples in Figure 2

are from a non-representative expert survey conducted in April 2022 among contributors to this *Handbook* and among scholars that are mentioned in the acknowledgments in Garoupa, Gill, and Tiede (2022, vii).[1] The experts mentioned different datasets and databases and we reviewed the characteristics included in the data. The list makes no claim to be complete. We exclude datasets composed of tailor-made data collected for individual research projects that address very specific research questions. This includes for example Vanberg's (2005) data on Germany or Staton's (2010) data on Mexico. In line with our argument, we also excluded data that does not include decision characteristics, such as Hamann's (2019) data on German justices. What do we learn from Figure 2 on data available to study comparative judicial behaviour?

Figure 2

Models	Type	Example	Access	Regions	Authors
1 - Legal	Database	PITAD.beta on investment arbitration cases	OA	various countries	Behn et al. 2019
	Datasets	The High Courts Judicial Database	OA*	various countries	Haynie et al. 2007
	Database	European Court of Justice Data	OA	Europe	Carrubba / Gabel 2011
	Datasets	Comparative Constitutional Reasoning Project	OA	various countries	Jakab et al. 2015
	Database	Inter-American Court of Human Rights Project	OA	South America	Loyola Law School
	Database	Websites by courts publishing decisions	OA	various countries	courts
	Database	Proprietary databases such as West Law	CA	various countries	often special publishers
2 - Attitudinal	Datasets /-base	Supreme Court Database	OA	United States	Spaeth et al. 2017
	Datasets /-base	Israeli Supreme Court Database	OA	Israel	Weinshall et al. 2018
	Database	CJEU Database (IUROPA Project)	OA	Europe	Brekke et al. 2023
	Database	Doranoh Database	LA	Norway	Grendstad et al. 2015
	Datasets	UK Supreme Court Data	OA	Great Britain	Hanretty 2020
	Datasets	U.S. Courts of Appeals Databases	OA	United States	Songer et al. 2008
	Datasets	DDL Judicial Data Portal	OA	India	Ash et al. 2021
	Datasets	Swedish Supreme Court Database	OA	Sweden	Lindholm et al. 2023
	Datasets	European Court of Human Rights Database (ECHRdb)	OA	International	Cichowski / Chrun 2017
	Datasets	EEOC Litigation Database	OA	United States	Kim et al. 2013
	Database	Fee Law Project	OA	United States	Lissner et al.
3a - Strategic	Datasets	Inter-American Court of Human Rights Database	OA	South America	Stiansen et al. 2020
	Datasets /-base	Constitutional Court Database	OA	Germany	Engst et al. 2020
	Datasets	European Court of Human Rights Database	OA	International	Stiansen / Voeten 2019
	Dataset	Judicial Review of Congress Database	OA	United States	Whittington 2019
		Comparative Law Project	NP	various countries	Gabel
3b - Strategic	Dataset	Compliance with Human Rights Tribunals Dataset	OA	International	Hillebrecht 2016

Note: Access OA is open access; LA is limited access either requiring registration or contacting the author(s); NP is data not published; CA is commercial access which requires a purchase. * Indicates that data is open access but downloaded files cannot be processed. Source: Non-representative expert survey among judicial politics scholars.

Examples of data structures to assess models of judicial behaviour.

Overall, most data structures mentioned by the experts include judges' characteristics and are suitable for assessments of the attitudinal model. Among these is the Supreme Court Database (Spaeth et al. 2017), which is one of the most popular data structures in judicial politics. The Supreme Court Database is a prime example of data often not being published in a database format. Databases are flexible and allow for the modelling of entities and relationships to capture the structure between them (Foster et al. 2021, 71). In the originally published Supreme Court Database, entities and relationships are already connected in a case-centred or justice-centred dataset. Only recently an online tool was added that allows the browsing of the data in a flexible database format. Publishing datasets readily available for analysis instead of underlying databases seems not as uncommon as the column 'Type' in Figure 2 may indicate.

Comparing the publication dates mentioned in the column 'Author(s)' across the column 'Models' and the 'Regions' uncovers further patterns in the data available to study judicial behaviour. First, datasets and databases that allow for assessments of the legal and attitudinal model are longer established than data structures suitable for analyses of the strategic model. Second, older data structure mostly cover courts in the United States from Courts of Appeal (Songer 2008), certain decisions by Federal Courts (Kim, Martin, and Schlanger 2013) to the Supreme Court (Spaeth et al. 2017; Whittington 2019) as well as popular highest international courts such as the European Court of Justice (Carrubba and Gabel 2011) or the European Court of Human Rights (Cichowski and Chrun 2017). Data on courts from individual countries such as the Supreme Court of the United Kingdom (Hanretty 2020), the German Constitutional Court (Engst et al. 2020), or Indian Courts (Ash et al. 2021) has been published more recently. Moreover, while there are only a few data structures that collect information on various different countries such as the PluriCourts Investment Treaty

Arbitration Database (PITAD) (Behn et al. 2019), the High Courts Judicial Database (Haynie et al. 2007), or the Comparative Constitutional Reasoning Project (Jakab, Dyevre, and Itzcovich 2017) most data structures focus on a single international court or national courts, for example the Inter-American Court of Human Rights (Stiansen, Naurin, and Bøyum 2020), the Norwegian (Grendstad, Shaffer, and Waltenburg 2015), or Israeli Supreme Court (Weinshall and Epstein 2020). Finally, participants in our expert survey mostly mentioned European and US courts.

The column 'Access' in Figure 2 indicates that the overwhelming majority of data is open access, albeit some projects do not get updated anymore. In this regard, one should note that proprietary databases such as Thomson Reuters *Westlaw* or Germany's equivalent *Juris* provide numerous original legal documents that permit the collection of information for datasets but only as commercial licences. The proprietary databases do not provide ready-to-use data structures for behavioural analyses. Instead, these databases are a collection of documents that allow deriving legal characteristics in raw form. The same holds true for official websites of individual courts that make decisions and legal documents searchable.

In sum, Figure 2 shows that data structures that include decision and justice characteristics are well established in judicial politics. Moreover, in recent years scholars tend to add environmental characteristics to such data structures. Thus, data to assess the different models of judicial behaviour is available in open access. Nevertheless, while the available data includes international courts the majority of data does not allow for comparative research across individual countries or regions. This is why in the next section we outline how databases for comparative judicial behaviour can be designed.
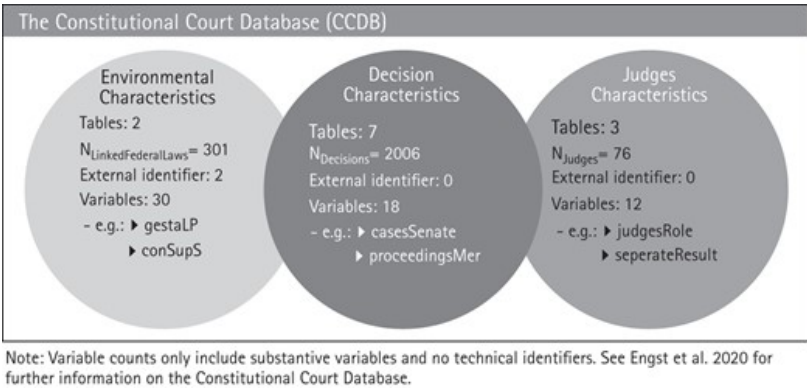
3. Developing Comparative Databases

In this section, we identify features essential to the design of databases. To do so we present the Constitutional Court Database (CCDB) (Engst et al. 2020) as an example of a database computed for a single country to assess judicial behaviour. Afterwards we outline how to transfer the features important to the design of the CCDB to a larger international and comparative project.

3.1 Features of a Database for a Single Court

The CCDB is a database recently published that includes decisions, judges, and environmental characteristics on three different layers. The database falls in zone 3a of Figure 1. It summarizes information on 2,006 decisions that include 3,284 proceedings initiated by 4,088 plaintiffs who referred 6,790 issues to the German Federal Constitutional Court between 1972 and 2010 (*decision characteristics*). The database also includes information on separate opinions and background information on seventy-six judges who served on the court (*judges characteristics*). Finally, the database includes a layer that allows linking the 301 federal laws referred to the constitutional court to a dataset on Germany's legislative proceedings (*environmental characteristics*). Figure 3 visualizes the CCDB. The circles in Figure 3 represent the different layers of the database including summary statistics on each layer. The layers reflect on the characteristics we require to assess judicial behaviour in accordance with Figure 1. What can we learn from the design of the CCDB?

Figure 3

Note: Variable counts only include substantive variables and no technical identifiers. See Engst et al. 2020 for further information on the Constitutional Court Database.

The Constitutional Court Database as a tool to study judicial behaviour.

First, the database consists of twelve different tables similar to rectangular datasets. The tables can be linked through identifiers internal to the databases and each table has at least one of these identifiers; for example, the 2,006 decisions can be linked to the seventy-six judges to compute a dataset that permits identification of the (normally) eight judges that make each of the 2,006 decisions. Instead, of coding the judges' information 2,006 times when creating the initial dataset, it is only necessary to code the judges' information seventy-six times. Obviously, seventy-six entries can be evaluated by the coders much easier than 2,006 entries. When an error is identified for information on a single judge then the correction applies equally to all decisions to which this judge is linked. Therefore, a database *reduces redundancies* in coding information which makes the data collection less prone to error. If an error on one of the tables is identified it can be corrected easier than when reviewing a long dataset.

Second, a legitimate claim of importance to all data-driven projects is to resist irrational 'data exuberance' (Weinshall and Epstein 2020, 429–30). Reducing redundancies in coding is one way to achieve this. Moreover, linking information across tables in a database allows us to stay focused on essential information of relevance to a data table. It is much easier to answer the question of the characteristics of judges one is interested in when looking at judges independent from decisions or votes. Thus, while the CCDB might appear complex at first sight, Figure 3 shows that there are only about sixty substantive variables that allow to cover many characteristics necessary to assess judicial behaviour. In our opinion that make the database *parsimonious*. This also makes its future maintenance easier.

Third, the different tables of the CCDB can be linked internally in multiple ways. For instance, the date a decision was published—which is a feature of a table summarizing information on a case—can be linked to a date included in a table summarizing public opinion data on Germany's political parties. This way one can identify the support for a political party, say the Social Democrats (variable conSupS), on the date a decision was made. Equipped with this data scholars can begin to assess how political support influences judicial decision-making. Similar questions have been studied with tailor-made datasets in the past (e.g. Vanberg 2005). Moreover, Figure 3 shows that the CCDB also includes external identifiers as part of the environmental characteristics. These identifiers allow links to preexisting data external to the CCDB, for instance to Germany's legislative proceedings. In sum, internal identifiers permit linking data within the CCDB in multiple ways while the external identifiers provide even more opportunities to add or create a number of datasets for a variety of specific research questions. This *flexibility* is a major advantage of databases over single datasets.

Finally, in order for a database to be parsimonious and flexible one has to be clear about the information coded. This is not only a question of the number of variables included in a database but also a question of how to transfer latent judicial concepts into relatable concepts understandable to a wider audience. Take, for

example, the complex way in which judges write an opinion and argue a case on the merits. The majority of people are probably not interested in the legal details but a simplifying measure on the merits. In the CCDB the decision on the merits has been coded per proceeding in a trichotomous variable that indicates whether a referred proceeding is not justified, partially justified, or justified on the merits (variable proceedingsMer). Similarly, the German Constitutional Court seldom publishes individual judicial votes which are frequently used in analyses of the attitudinal model in the US context (Epstein and Knight 1998; Martin and Quinn 2002). However, in rare instances, the German judges publish separate opinions (Wittig 2016) and while specific details on those opinions are not included in the CCDB, the results are coded in a trichotomous variable. This variable summarizes whether a separate opinion is written as a concurring, partially dissenting or fully dissenting opinion (variable separateResult). The way the dissenting opinions are coded should be compelling to a wide audience. Therefore, not all information can be coded in a parsimonious database and it is essential to simplify information while being *conceptually clear*.

In sum, databases are powerful tools to account for variables on each characteristic necessary to assess models of judicial behaviour. Databases help to reduce redundancies, permit parsimonious data collection, and provide for high flexibility in dataset production. To enable a large audience to work with databases, scholars need to make sure that the concepts underlying the data collection efforts are clearly specified. The CCDB is an example of a database designed for the German Constitutional Court as a single case, though. Indeed, Figure 2 shows that the overwhelming majority of databases are designed for single courts. This is why in the next section we outline how features of the CCDB can be transferred to a comparative project.

3.2 Conceptualizing a Comparative Judicial Behaviour Database

What can we learn from developing the CCDB when conceptualizing other data structures for the analysis of courts (within and) across countries? We identified four features of importance to high-quality data infrastructures: reducing redundancies, parsimony, flexibility, and conceptual clarity. In this section, we outline the implications of these features when designing a comparative judicial behaviour database as part of an international project.

3.2.1 Reducing Redundancies

Our first argument concerns the *reduction of redundancies*. As a field, we would benefit if we took advantage of opportunities to systematically learn from each other's expertise on particular courts and judicial systems in order to conceptualize and implement a comparative data infrastructure project. The project could be a driver of novel comparative research particularly *across* courts and legal systems.

It is reasonable to expect that such a data infrastructure can be a game-changer for the questions we aim to answer and for how we conduct comparative judicial research in the future. Currently, such research is often merely equated with studying courts and their decision-making behaviour outside the United States. Recent summaries of the comparative literature by regional specialists highlight that the prevailing research design in the reported studies consists of country-level or court-level case studies (see the respective chapters in Garoupa, Gill, and Tiede 2022).

Driscoll (2022) conducted a content analysis of articles published in top peer-reviewed political science journals between 2000 and 2018 (*American Political Science Review*, *American Journal of Political Science*, *The Journal of Politics*, *British Journal of Political Science*, *Comparative Political Studies*, and the *Journal of Law and Courts*). The analysis provides evidence that only a minority of studies explicitly implement a cross-national design.

Cross-national comparative work encompasses various approaches. There is research that tries to test general theories based on a country-by-country analysis. An example of this is the work by Brouard and

Hönnige (2017). The authors utilize veto player theory to derive observable implications that can be tested comparatively. They gather context-specific measures of the positions of institutions (such as the highest court, parliament, government, and president) in the country-specific policy spaces of France, Germany, and the United States to examine how frequently the highest courts are absorbed by veto players. Since their key measures are context-specific, they cannot simply aggregate their data but instead rely on conducting a country-by-country analysis. They select their countries because system design differs greatly (Przeworski and Teune 1970).

In a similar vein, Alarie and Green (2017) adopt a country-by-country analysis of five high courts: the US Supreme Court, the UK Supreme Court (including its predecessor, the House of Lords), the Supreme Court of Canada, the High Court of Australia, and the Indian Supreme Court. The study demonstrates that these courts vary along two dimensions and that the design differences correlate with how the courts decide. While the authors do compile some country-specific measures, their main source of data is the publicly available High Courts Judicial Database (HCJD) assembled by Haynie et al. (2007).

Finally, there is research that pools data for analysis when relevant measures are comparable across different contexts. Take, for instance, Skiple, Bentsen, and McKenzie (2021) analysis of dissent in tax decisions in the Norwegian and Danish Supreme Courts. Recognizing that courts may respond differently to increasing caseloads by exercising discretionary docket control to varying degrees, the authors employ a most similar design (Przeworski and Teune 1970) and pool the data from both courts. Their aim is to examine whether discretionary dockets are associated with higher dissent and reversal rates compared to mandatory dockets. The data for this analysis is primarily sourced from available databases, namely the Danish Supreme Court Database (McKenzie, Bentsen, and Skiple 2016) and the Norwegian Supreme Court Database (Grendstad, Shaffer, and Waltenburg 2015).

To summarize, even though cross-national comparative work comes in different flavours, one obvious benefit of a comparative database is the provision of data across countries (or courts). Moreover, a second benefit lies in the development of such an infrastructure project. It requires scholars to coordinate their efforts to conceptualize the design of the database. This reduces redundancies (beyond mere technical aspects) by incentivizing scholars to contribute to a common good rather than duplicating each other's data-collection efforts.

3.2.2 Parsimony

Our second argument concerns *parsimony*. What does a *parsimonious data collection strategy* imply when developing a comparative database? To enable scholars to eventually use the most common explanatory models of judicial behaviour for their analysis, any data infrastructure should provide information not only on decisions and judge characteristics but also on environmental characteristics. As with children just before their birthday though, the wish lists quickly become quite long. The data of the variables on the wish list has to be not only compiled at some point in time but also needs to get updated to maintain such a project over time.

Thus, scholars involved in designing such an infrastructure project need to resist data exuberance. Less is more! After taking stock of existing resources—for instance the Israeli Supreme Court Database (Weinshall and Epstein 2020), which in many respects is modelled after the US Supreme Court Database—scholars need to define 'core characteristics' that should be collected consistently for all courts and countries included in the comparative database.

Such characteristics describe decisions, judges as well as the environment. They obviously live on different observation levels. While having a master plan is definitely a good thing, scholars are luckily not simply a bunch of Trotskyist internationals. They can allow themselves to change plans—for instance when updating

the data across countries and courts. Scholar might therefore distinguish cross-sectional core characteristics, which are only measured once, from longitudinal core characteristics that are repeatedly measured. This allows for the possibility that at some point also new characteristics can make it in the set of 'core characteristics' when starting to update the initial data collection plan. However, relevant characteristics should apply to a large enough subset of cases to allow for a cross-country comparative analysis of those characteristics.

Ideally, one can develop pipelines to automate the updating process for some of these characteristics. The less human judgment is involved in coding decisions, the more reliable is the data collection strategy. The more data can be collected automatically or the fewer characteristics need to get updated, the more room there is to implement ideas for new characteristics that should get coded in subsequent data collection phases. Additionally, alongside the core characteristics, there can be specific characteristics that apply only to a subset of courts or countries.

In summary, each data collection phase to maintain the data infrastructure requires scholars to have a predefined list of characteristics that should be collected. However, it is important to recognize that this list may undergo partial changes when transitioning from one data collection phase to the next.

3.2.3 Flexibility

Our third argument concerns *flexibility*. Scholars should leverage the inherent flexibility offered by databases in contrast to standalone datasets. Therefore, in addition to determining what should be collected in a comparative manner and at which level of observation, scholars also need to consider how newly collected data can be systematically integrated with existing data infrastructure projects. Databases offer internal links that enable efficient combination of data stored in separate datasets. When designing a comparative database, scholars should also consider incorporating external links to facilitate the incorporation of specialized information to enhance the richness of the data structure.

By incorporating valid external links, some of the burden of compiling environmental characteristics can be alleviated. By including a time variable for court decisions for instance, scholars could collect their own information to better characterize the political environment in which the respective decisions are published, such as election results or the popularity of the respective government.

Another possibility would be that such time and decision identifiers allow tracking down (potentially using external data) the actual decision text to conduct an analysis of who cites whom among the apex courts and which potentially common other legal sources have been cited. In addition to the decision text itself one could also include accompanying text material such as briefs, press releases, or news reports related each decision. Moreover, one could also include qualitative interview data, survey marginals but also potentially variables derived from verbatim notes, audio or video files from hearings of the cases, or results from crowd-sourced coding or lab experiments that are related to a decision as well. The sky is the limit.

Likewise, by including court and country identifiers as an external link, for instance, scholars could integrate existing (or newly collected) country-level datasets about institutional characteristics to the analysis. These additional country-level datasets provide valuable information on the system of governance or the legal system.

In summary, by proactively considering how different data sources can be linked to the comparative database through external links and maintaining a flexible data structure also through internal links, the data infrastructure becomes not only highly beneficial to those who conceptualized it but also to a wide range of additional users.

3.2.4 Conceptual Clarity

Our fourth argument concerns *conceptual clarity*. This is arguably the most complex feature to address when designing a comparative database. It necessitates consensus among scholars regarding the relevant theoretical concepts to be used, how they should get operationalized into variables, and the methods of measuring them at which appropriate level of observation. Building on the expertise of a diverse group of experts becomes crucial in overcoming challenges related to translating relevant terms and concepts, as well as ensuring that concept specifications are applicable across different countries or legal systems.

It is well known that the meaning of even the most fundamental concepts, such as case outcome, a judge's ideology, or case characteristics, can vary across countries (e.g. Gill and Zorn 2022, 312–18). Consequently, achieving direct comparability of these concepts is not straightforward. This challenge is common in the field of comparative politics. The literature suggests (e.g. van Deth 1998) addressing such challenges by specifying 'functional equivalent' rather than identical concepts. Equivalent concepts are related to other concepts in an analogous way across contexts. This often requires 'increas[ing] the level of abstraction of the concepts' (van Deth 1998, 6) until context-specific differences seem not to matter anymore. This is not a free lunch though, as increasing abstraction could easily lead to concept stretching (Sartori 1970, 1041) such that the functional equivalent concepts lose their analytical power.

Think about how one could conceptualize decision outcomes. The Supreme Court Database, for example, focuses on the ideological direction of a decision, that is, whether an outcome is liberal rather than conservative. Outcomes of decisions in other countries might vary along different substantive dimensions— for instance, more or less European Union integration in the case of the Court of Justice of the European Union (CJEU). By increasing abstraction of the concept 'outcome of a decision' one can avoid any substantive characterization of outcomes that hardly travel across contexts.

An outcome could be simply specified in terms of whether the plaintiff wins, partially wins, or whether the defendant wins. Operationalizing outcomes with regard to those facts we may even end up with a more abstract operationalization, for example plaintiff wins the case or not. Thus, we avoid defining what a 'partial win' implies. This more abstract, yet valid and reliable measurement strategy travels well across different contexts. Another strategy is used by Skiple, Bentsen, and McKenzie (2021). When conceptualizing their non-unanimous decision variable, they distinguish between disagreement about outcome (dissents) and disagreement about the rationales (concurrences). Given that they like to pool the data across different contexts they realize that this can only be done for dissents but not for concurrences. Hence, they only code and analyse dissents as a decision outcome.

Most importantly though, the validity of cross-national measurement strategies must be established rather than assumed. Additionally, the respective coding protocols should minimize the use of human judgment in order to enhance reliability. Reliability is a crucial criterion for defining a reproducible strategy that facilitates the updating of observations within a new time frame in a given context. It is also essential for ensuring that observations from new contexts can be added to the existing infrastructure.

Finally, conceptual clarity is increased by mapping out a precise case selection strategy that defines the criteria for including an observation and determines the appropriate level of observation for each characteristic to be coded. Taking the example of decision outcomes again, we can consider the decision-level outcome or a more refined strategy that examines how the court disposes of each legal issue per case. The latter approach provides more nuanced information that would otherwise be overlooked. However, refined coding strategies are not necessarily superior. They require more effort and potentially more human judgment during the coding process.

Furthermore, if the underlying theory or other variables do not vary across different legal issues within a case, it becomes questionable why one should invest additional coding effort to obtain a more detailed

picture.

To summarize, ensuring conceptual clarity is crucial but it can be a challenging task. Designing a comparative database involves establishing case selection criteria, conceptualizations, and coding protocols that are applicable across different contexts. These procedures should be reproducible, allowing others to add new countries and courts in the future or update the existing data for different time periods.

4. An Institutional Proposal for Action: The 'Comparative Study of Judicial Behaviour'

How should we move forward? While it may have been controversial decades ago (e.g. Tate 1989, 1992), by now it seems to be common sense for everyone working on courts comparatively that the field needs to find ways to compile, regularly update, and maintain a high-quality data infrastructure for analysing courts and judicial behaviour within and across countries.

In order to make real progress and allocate our time and resources towards data compilation efficiently, our field needs to change its approach to overcome the dominant data collection strategy known as the 'one-off' approach (e.g. Weinshall and Epstein 2020; Epstein and Weinshall 2021). Currently, scholars often collect data and measures that specifically align with their particular research questions, theories, and chosen institutional and political contexts to test their hypotheses. While this is individually rational—scholars need to use their time and research funds effectively—as a field we would greatly benefit from finding ways to avoid duplicative scholarly efforts. Furthermore, we should also learn from failed attempts to build a database. Why were projects ultimately unsuccessful in accomplishing their intended goals?

How can we establish a high-quality data infrastructure and set it on the right track? We propose looking at how scholars of comparative political behaviour have done this previously. We can learn from them because scholars in this field face similar challenges than we do when studying judicial behaviour comparatively. Conducting comparable election studies, which involves surveying voters in different electoral systems, poses similar challenges. Implementing the same survey instrument requires case-specific knowledge. Scholars need to understand the political context in which the survey is conducted. They also need to understand the nature of party competition in a country and how its institutional context, such as the electoral system, pre-structures various party strategies and voter behaviour. Additionally, scholars need to have sufficient language proficiency to ensure the comparability of different implementations of the same survey items in different languages. To achieve this, scholars must collaborate and pool their resources and expertise.

We propose the establishment of a collaborative research programme called the Comparative Study of Judicial Behaviour (CSJB) with research teams from around the world. The aim is to create and maintain a high-quality data infrastructure for comparative judicial research, similar to the '*Comparative Study of Electoral Systems*' (CSES) (see https://cses.org/ for more information). The CSES is a collaborative programme for election studies that was institutionalized by scholars of comparative political behaviour two decades ago. Participating teams include a common module of survey questions in their post-election studies to measure respondents' voting behaviour and political attitudes during national elections. In addition to individual survey data, the resulting data includes contextual variables that characterize features of, for example, the electoral district, the overall polity, and the electoral system, providing additional context for understanding respondents' voting decisions. All this election-specific and country-specific information is merged with the survey data to create a single, publicly available dataset. Every five years, a new CSES module is designed with a different substantive theme, selected to address relevant questions in comparative political behaviour.

The research design, the respective required comparative data as well as the study design, for example inclusion criteria, could be developed by a module planning committee, an international committee of scholars of judicial behaviour in political science, law, and neighbouring fields such as sociology and economics. The design is then implemented in each participating country by a scholar or a team of scholars —the country collaborators. Similar to the CSES operations, those country collaborators possess enough case-specific knowledge to organize the data collection within their country. This requires the planning committee to lay out conceptualizations and coding protocols that travel across legal systems and are reproducible (Weinshall and Epstein 2020).

The respective planning committee needs to decide what the first data module should look like, that is which case-level and judge-level data should get collected systematically. Moreover, in addition to case-level and judge-level data the module planning committee should define macro-level characteristics such as features about the court and its internal decision-making processes, as well as characteristics of the legal system together with variables (or at least external identifiers for them) that define the political context (e.g. public opinion data, election results) in which the judicial decisions are made. Such characteristics are on the macro level in the sense that they apply, for instance, to each case-level variable in the same way.

Actively resisting data exuberance is of utmost importance for the committee to motivate scholars participating in this endeavour and to help update already compiled data. Thus, not every characteristic of judges, courts, decisions, and the political context can be included, but whatever characteristic should get coded (at whatever level of observation) requires a detailed coding protocol to make sure that the data are valid and country collaborators can reliably collect (by minimizing human judgment) the data.

Similar to the CSES, the planning committee of the first CSJB module could initially establish a common set of baseline variables and measures that would be expanded and updated in subsequent modules. The sources from which the data is generated are manifold. In addition to the decisions themselves, including accompanying textual material such as briefs, press releases, or news reports, qualitative interview data, survey data, and also potential variables from verbatim notes, audio or video files of hearings on the cases, or results of crowd-sourced coding or laboratory experiments could be included.

While potentially not all variables can be updated, nevertheless a master plan is required that defines which key variables should get updated from one module to the next (or even automatically if such pipelines can be established). Such a plan ensures that the infrastructure is sustainable while not growing too big such that new country teams can still join this collaborative program at a later stage without hesitation. The CSES is meanwhile coordinating its sixth data module.

Why should scholars participate in this comparative infrastructure project? One incentive to participate in the CSJB research programme could be to get access to a (centralized) tech support infrastructure that develops and adapts data collection pipelines to automatically and, thus, reliably code the respective characteristics within a given context. Hand-coding should be avoided as much as possible. This would enable researchers with less knowledge and expertise in automatic data collection strategies (e.g. web-scraping, regular expressions, etc.) and programming expertise to collect more than the required data to further their personal research agenda as well. In addition to defining particular characteristics, the tech support infrastructure could provide tools to automatically use the text of a decision to categorize it depending on the agenda defined by the module's planning committee. The resulting data should be open and accessible. Therefore, it should get deposited along with coding protocols of the macro data to the project's headquarter for data checking and to prepare for its immediate release to the research community. Even if a physical comparative judicial database may not grow as fast as we all would like, it would be nevertheless instrumental that the datasets used in research—as a way of establishing a best-practice norm —become available in various data repositories. If they include external identifiers, such data can be used for new research projects of scholars who might be willing to extend or update existing data.

5. Conclusion

Why do we need databases in research on comparative judicial behaviour? The assessment of common models on judicial behaviour requires knowledge of decisions, judges, and environmental characteristics. Databases allow us to summarize information on all these characteristics in parsimonious and flexible ways without producing redundancies in collected data. An expert survey we conducted shows that while common rectangular datasets were originally used to study the attitudinal model of judicial behaviour, relational databases addressing the strategic model have been published in recent years. However, most existing data structures address single national courts or specific international courts. In order to design an international database for a CSJB, existing comparative projects, such as the CSES, can function as role models.

The development of judicial databases is of major importance to modern research on democracy. Databases allow us to embed courts and judicial action within political systems and societies. Thus, they portray courts as what they are: key institutions in the separation of powers. In order to understand developments within political regimes, open access data on the judiciary is as essential as data on the executive and the legislative branches. Thus, it is important that we as a community are able to engage in large scale comparative data driven projects. Developments that limit the opportunity to analyse data structures (e.g. the French Article 33 LOI n° 2019-222 du 23 mars 2019) eventually hinder scientific progress. After all, databases in judicial politics allow to evaluate the independence of judges, the judicialization of politics as well as the politization of the judiciary and the protection of individual rights as essential features of democracy.

References

Alarie, Benjamin, and Andrew J. Green. 2017. *Commitment and Cooperation on High Courts: A Cross-Country Examination of Institutional Constraints on Judges*. Oxford: Oxford University Press.
Google Scholar   Google Preview   WorldCat   COPAC

Ash, Elliot, Sam Asher, Aditi Bhowmick, Sandeep Bhupatiraju, Daniel Chen, Tanaya Devi, Christoph Goessmann, Paul Novosad, and Bilal Siddiqi. 2021. 'In-group bias in the Indian judiciary: Evidence from 5 million criminal cases'.
https://www.devdatalab.org/judicial-data
WorldCat

Behn, Daniel, Malcolm Langford, Ole Kristian Fauchald, Runar Lie, Maxim Usynin, Taylor St John, Laura Letourneau-Tremblay, Tarald Berge, and Tori Loven Kirkebø. 2019. *PITAD Investment Law and Arbitration Database: Version (2019-02-12T09:28:49.184Z)*. Oslo: University of Oslo, Pluricourts Centre of Excellence.
Google Scholar   Google Preview   WorldCat   COPAC

Brouard, Sylvain, and Christoph Hönnige. 2017. 'Constitutional courts as veto players: Lessons from the United States, France and Germany'. *European Journal of Political Research* 56 (3): 529–52.
Google Scholar   WorldCat

Carrubba, Clifford J., and Matthew J. Gabel. 2011. *European Court of Justice Data*.
https://polisci.emory.edu/home/people/biography/carrubba-clifford.html
Google Scholar   Google Preview   WorldCat   COPAC

Cichowski, Rachel, and Elizabeth Chrun. 2017. *European Court of Human Rights Database (ECHRdb), Version 1.0 Release 2017*.
https://depts.washington.edu/echrdb/
Google Scholar   Google Preview   WorldCat   COPAC

Driscoll, Amanda. 2022. Comparative Studies of Judicial Behavior. In *High Courts in Global Perspective: Evidence, Methodologies, and Findings*, edited by Nuno Garoupa, Rebecca D. Gill, and Lydia B. Tiede, 275–306. Charlottesville: University of Virginia Press.
Google Scholar   Google Preview   WorldCat   COPAC

Engst, Benjamin G., Thomas Gschwend, Christoph Hönnige, and Caroline E. Wittig. 2020. *The Constitutional Court Database: Conceptualizing a Relational Database*. https://ccdb.eu/
Google Scholar   Google Preview   WorldCat   COPAC

Epstein, Lee, and Jack Knight. 1998. *The Choices Justices Make*. Washington: CQ Press.
Google Scholar   Google Preview   WorldCat   COPAC

Epstein, Lee, and Keren Weinshall. 2021. *The Strategic Analysis of Judicial Behavior: A Comparative Perspective*. Cambridge: Cambridge University Press.
Google Scholar   Google Preview   WorldCat   COPAC

Epstein, Lee, Urška Šadl, and Keren Weinshall. 2022. 'Symposium: The Role of Comparative Law in the Analysis of Judicial Behavior'. *The American Journal of Comparative Law* 69 (4): 689–719.
Google Scholar   WorldCat

Foster, Ian, Rayid Ghani, Ron S. Jarmin, Frauke Kreuter, and Julia Lane. 2021. *Big Data and Social Science. Data Science Methods and Tools for Research and Practice*. 2nd ed. Boca Raton: CRC Press.
Google Scholar   Google Preview   WorldCat   COPAC

Garoupa, Nuno, Rebecca D. Gill, and Lydia B. Tiede. 2022. *High Courts in Global Perspective: Evidence, Methodologies, and Findings*. Charlottesville: University of Virginia Press.
Google Scholar   Google Preview   WorldCat   COPAC

George, Tracey E., and Lee Epstein. 1992. 'On the Nature of Supreme Court Decision Making'. *American Political Science Review* 86 (2): 323–37.
Google Scholar   WorldCat

Gill, Rebecca D., and Christopher Zorn. 2022. Overcoming the Barriers to Comparative Judicial Behavior Research. In *High Courts in Global Perspective: Evidence, Methodologies, and Findings*, edited by Nuno Garoupa, Rebecca D. Gill, and Lydia B. Tiede, 307–30. Charlottesville: University of Virginia Press.
Google Scholar   Google Preview   WorldCat   COPAC

Grendstad, Gunnar, William R. Shaffer, and Eric N. Waltenburg. 2015. *Policy Making in an Independent Judiciary: The Norwegian Supreme Court*. Colchester: ECPR Press.
Google Scholar   Google Preview   WorldCat   COPAC

Hamann, Hanjo. 2019. 'The German Federal Courts Dataset 1950–2018: From Paper Archives to Linked Open Data'. *Journal of Empirical Legal Studies* 16 (3): 671–88.
Google Scholar   WorldCat

Hanretty, Chris. 2020. *A Court of Specialists: Judicial Behavior on the UK Supreme Court*. Oxford: Oxford University Press.
Google Scholar   Google Preview   WorldCat   COPAC

Haynie, Stacia L., Reginald S. Sheehan, Donald R. Songer, and C. Neal Tate. 2007. *High Courts Judicial Database*. Accessed at the University of South Carolina Judicial Research Initiative. http://www.songerproject.org/national-high-courts.html
Google Scholar   Google Preview   WorldCat   COPAC

Jakab, András, Arthur Dyevre, and Giulio Itzcovich. 2017. *Comparative Constitutional Reasoning*. Cambridge: Cambridge University Press.
Google Scholar   Google Preview   WorldCat   COPAC

Kim, Pauline T., Andrew D. Martin, and Margo Schlanger. 2013. *EEOC Litigation Database*. https://eeoclitigation.wustl.edu/
Google Scholar   Google Preview   WorldCat   COPAC

King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.
Google Scholar   Google Preview   WorldCat   COPAC

Martin, Andrew D., and Kevin M. Quinn. 2002. 'Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953–1999'. *Political Analysis* 10 (2): 134–53.
Google Scholar   WorldCat

McKenzie, Mark, Henrik Litleré Bentsen, and Jon Kåre Skiple. 2016. '*The Danish Supreme Court Database*'. Texas Tech University.
Google Scholar   Google Preview   WorldCat   COPAC

Murphy, Walter F. 1964. *Elements of Judicial Strategy*. Chicago: Chicago University Press.
Google Scholar   Google Preview   WorldCat   COPAC

Parcelle Jr., Richard L., Brett W. Curry, and Bryan W. Marshall. 2011. *Decision Making by the Modern Supreme Court*. Cambridge: Cambridge University Press.
Google Scholar   Google Preview   WorldCat   COPAC

Przeworski, Adam, and Henry Teune. 1970. *The Logic of Comparative Social Inquiry*. New York: Wiley Sons.
Google Scholar   Google Preview   WorldCat   COPAC

Rohde, David W., and Harold J. Spaeth. 1976. *Supreme Court Decision Making*. San Francisco: Freeman.
Google Scholar   Google Preview   WorldCat   COPAC

Sartori, Giovanni. 1970. 'Concept Misformation in Comparative Politics'. *American Political Science Review* 64 (4): 1033–53.
Google Scholar   WorldCat

Schubert, Glendon A. 1965. *The Judicial Mind. Attitudes and Ideologies of Supreme Court Justices, 1946–1963*. Evanston: Northwestern University Press.
Google Scholar   Google Preview   WorldCat   COPAC

Segal, Jeffrey A., and Albert D. Cover. 1989. 'Ideological Values and the Votes of U.S. Supreme Court Justices'. *American Political Science Review* 83 (2): 557–65.
Google Scholar   WorldCat

Skiple, Jon Kåre, Henrik Litleré Bentsen, and Mark Jonathan McKenzie. 2021. 'How Docket Control Shapes Judicial Behavior: A Comparative Analysis of the Norwegian and Danish Supreme Courts'. *Journal of Law and Courts* 9 (1): 111–36.
Google Scholar   WorldCat

Songer, Donald R. 2008. *U.S. Courts of Appeals Databases—Phase II*. http://www.songerproject.org/us-courts-of-appeals-databases.html
Google Scholar   Google Preview   WorldCat   COPAC

Spaeth, Harold J., Lee Epstein, Andrew D. Martin, Jeffrey A. Segal, Theodore J. Ruger, and Sara C. Benesh. 2017. *Supreme Court Database, Version 2017 Release 01*. http://scdb.wustl.edu/
Google Scholar   Google Preview   WorldCat   COPAC

Staton, Jeffrey K. 2010. *Judicial Power and Strategic Communication in Mexico*. Cambridge: Cambridge University Press.
Google Scholar   Google Preview   WorldCat   COPAC

Staton, Jeffrey K., and Georg Vanberg. 2008. 'The Value of Vagueness: Delegation, Defiance, and Judicial Opinions'. *American Journal of Political Science* 52 (3): 504–19.
Google Scholar   WorldCat

Stiansen, Øyvind, Daniel Naurin, and Live Standal Bøyum. 2020. 'Law and Politics in the Inter-American Court of Human Rights'. *Journal of Law and Courts* 8 (2): 359–79.
Google Scholar   WorldCat

Tate, C. Neal. 1989. 'Scientific Comparative Judicial Politics: An Assessment and a Call to Action'. *Law Courts Newsletter* 7 (1): 1–8.
Google Scholar   WorldCat

Tate, C. Neal. 1992. 'The Development of Comparative Judicial Politics'. *Perspectives on Political Science* 21 (3): 138–44.
Google Scholar   WorldCat

van Deth, Jan W. 1998. 'Equivalence in Comparative Political Research'. In *Comparative Politics. The Problem of Equivalence*, edited by Jan W. van Deth, 1–19. London: Routledge.
Google Scholar   Google Preview   WorldCat   COPAC

Vanberg, Georg. 2005. *The Politics of Constitutional Review in Germany*. Cambridge: Cambridge University Press.
Google Scholar   Google Preview   WorldCat   COPAC

Weinshall, Keren, and Lee Epstein. 2020. 'Developing High-Quality Data Infrastructure for Legal Analytics: Introducing the Israeli Supreme Court Database'. *Journal of Empirical Legal Studies* 17 (2): 416–34.
Google Scholar   WorldCat

Whittington, Keith E. 2019. *Repugnant Laws: Judicial Review of Acts of Congress from the Founding to the Present*. Lawrence: University Press of Kansas.
Google Scholar   Google Preview   WorldCat   COPAC

Wittig, Caroline E. 2016. *The Occurrence of Separate Opinions at the Federal Constitutional Court. An Analysis with a Novel Database*. Berlin: Logos Verlag.

Google Scholar   Google Preview   WorldCat   COPAC

Note

1   We like to thank all participants of our expert survey for their valuable input.