Taking Law Seriously: The Challenges of Law as Research Data in Socio-legal Scholarship

Audrey M. Plan

Sutherland School of Law, University College Dublin, Ireland

Abstract

'Law-as-data' is often understood in terms of computational data. This article reframes it around a research-centred notion of data, broadening the methodological possibilities available to empirical legal research. This approach transcends the qualitative/quantitative divide, identifying common challenges that arise when law is treated as research data. Methodological tools from the social sciences and examples from recently published socio-legal scholarship are used to demonstrate how law presents unique challenges at the stages of conceptualisation, operationalisation and coding. These challenges stem from factors such as the non-commensurability of legal texts, the fluidity and contested meaning of legal notions, the difficulty of systematically gathering legal data and the risk of bias towards data-driven research rather than law- or reality-centred inquiry. However, strategies and best practices can be developed to address or mitigate these risks. Legal academics are particularly well suited to this task, given the high level of expertise required to effectively leverage law as research data. The article argues that a true methodological naturalisation of law as research data, which embraces the unique characteristics and limitations of law as a data source, would greatly benefit legal studies by diversifying its methodological toolbox while preserving its specificities.

Keywords: Law-as-data; empirical legal research; interdisciplinarity; research methods; text-as-data; qualitative methods.

Introduction

Conceptualising the law as data presents an opportunity for the legal academy on multiple levels.[1] It is a chance to communicate across disciplines with a commonality of empirical methods; a chance to provide a more empirical basis to key research streams in the field; and a chance to rediscover law in a new light, identify new patterns, rethink old debates and spark new ones.

Yet 'law-as-data' is often reduced to large-scale, quantitative data analysis and data processing.[2] This article argues that this understanding of *data* is influenced by computer science through big data-driven, computer-assisted social sciences, where the goal is to handle large amounts of data, normalised and systematised, to be analysed using computational methods. Meanwhile, social science research in general has long had a much broader understanding of data: *a datum is a* piece of reality against, or through, which researchers can improve their understanding of socio-legal phenomena. It does not correlate with being quantifiable or being fit to be handled by an algorithm. Still, most of the scholarly discussions on law-as-data have been informed by a conception of law as computational data, as opposed to research data: by explicitly defining law-as-data as meaning computational data,[3] by not including works treating law as *qualitative* data in the discussion of empirical legal studies[4] or by explicitly bypassing them for convenience.[5]

The goal is not to push for law to always, and only, be *research data*. Data should be a means to an end – depending on the end, treating law as 'data' in the way data is approached in social science scholarship is not always pertinent or advisable. Law

---

[1] The author expresses her sincere gratitude to the reviewers and editors for their insightful comments and valuable suggestions. Their constructive feedback greatly contributed to the improvement of the initial manuscript.

[2] Livermore, Law as Data.

[3] Frankenreiter, "Computational Methods in Legal Analysis."

[4] Chin, "Replicability in Empirical Legal Research."

[5] Cope, "Empirical Studies of Human Rights Law."

can be *turned into* data, but it also exists in its own state of *law* and can be analysed as such. A contribution to legal scholarship, doctrinal research and other normative legal analysis can be produced in a scholarly way without having to rely on law-as-data. The goal of this article, by focusing on an *academic research* framework rather than a computer science one, is to put the *law* front and centre of *law-as-data*. How law fits into this understanding (or multiple understandings!) of data is the crux of this discussion. And while this is particularly relevant for legal and socio-legal research, the question of how to turn law into usable data for research purposes is relevant to any field where the law at large matters: economics, sociology and political science first and foremost.

The contribution of this article is therefore twofold. First, it offers a research- and social science-based alternative to the assumption that law-as-data is necessarily text-as-data for quantitatively minded researchers interested in large-N analysis and computational methods.[6] The article presents a conception of law-as-data relevant to both qualitative and quantitative empirical research, on the basis that they both share at least partially similar ontologies and epistemologies. Moreover, this approach avoids placing the emphasis on the methods at the expense of the idiosyncrasies of *law* in empirical research, overlooking both complexities and opportunities that will be raised in this article.

Second, the article highlights very specific challenges in turning law into data with which boundary-pushing works developing and encouraging robust and diverse empirical methodology for law have not yet grappled.[7] The underlying assumption is that law is very well placed to become data due to being embedded in the tangible medium of 'text', a medium well suited to being data (including qualitative data).[8] Breakdowns of how to conduct empirical legal research, however, often jump from 'what are the existing data sources' to 'data analysis' without covering the intermediary steps that are commonly explored in similar social sciences methods scholarship.

To make these contributions, the article is structured as follows. In section 1, I present *what is meant by 'law as data'*. This differs from both data in general for empirical and socio-legal studies, but also law as a normative concept as it is familiar to lawyers. This will set up one of the throughlines of the discussion on law-as-data: that of the standpoint of the researcher. Approaching law as data requires the researcher to engage in a constant dialectic between their social scientist-self and their lawyer-self.[9] Acknowledging this allows us to better understand the challenges presented in the next section and identify potential solutions.

Section 2 therefore delves into the challenges of turning law into data: how law proves to be a very idiosyncratic object to data-ify for empirical research. This is broken down into two quandaries, which can sometimes be overcome or mitigated, and sometimes must simply be acknowledged as limits. The first is the *conceptualisation/operationalisation problem*: legal concepts are often fuzzy, multi-faceted and in a constant state of being defined and redefined. Turning these concepts into data requires finding a way to transcend the constant scholarly legal debates over definitions, one way or another, without being overly extensive or overly restrictive. There is no data without engaging in theoretical conceptualisation: law-as-data does not bypass legal debate. The second is the *coding problem*: the law is inherently open to interpretation and scholarly debate about its very meanings, making it difficult to capture some objective empirical reality about its content that can be turned into data. Section 3 takes a step back to look at the *'hammer problem'*: law-as-data is a useful tool, but it cannot and should not be used to solve all questions falling under the ambit of legal scholarship. To do otherwise impoverishes the legal discourse rather than enriching it.

Section 4 offers a few reflections and suggestions on how to encourage best practices in law-as-data research. Far from being exhaustive, it seeks to encourage a dialogue on the wider scholarly environment in which such research is conducted.

Although I have endeavoured to focus on questions about data that would cut across as many methodological divides and research areas as possible, examples will be drawn predominately from my own research area (international courts, human rights and EU law). This should not be taken as circumscribing the arguments put forward to these fields and methods only; instead, this article invites further reflection on how the many sub-fields of law can be combined with the many approaches and understandings of *research data*.

---

[6] Frankenreiter, "Computational Methods in Legal Analysis."

[7] Kritzer, Advanced Introduction to Empirical Legal Research; Bos, Empirical Legal Research; Bhat, Idea and Methods of Legal Research.

[8] Mitchell, "Analyzing the Law Qualitatively"; Dale, "Law and Word Order." See also the 2019 Special Issue on natural language processing for legal texts: Robaldo, "Introduction for Artificial Intelligence and Law."

[9] The author thanks the anonymous reviewers for pointing out the importance of discussing the standpoint of the researcher throughout this article.

1. From Data to Law and Back

Law-as-data is not *all data* used by empirical legal studies and socio-legal studies. As stated by Webley, data 'may be cases and statutes or a range of other types of documents (law firm annual reports, arbitral awards, curricula, and training materials, etc.). It may extend to in-person observations of sites of legal engagement, whether court, lawyers' offices or law school clinics or classrooms';[10] law-as-data, within the boundaries of this article, is only the first part. Put differently, if socio-legal scholars investigate law with its 'text, context and subtext',[11] then law-as-data is only the text. This means that this article is interested in turning 'the lawyer's law' into data.[12]

We can understand why law-as-data can indeed fit text-as-data computational methods but does not need to be limited to them. Law-as-data can be text, or it can be legal information of which the text is only the medium. Law-as-data is agnostic *vis-à-vis* its potential computation. Before being transformed into a bag of words fed through *quanteda*,[13] law-as-data is its text as much as it is its legal information (law as a normative object with which lawyers are familiar). Legal researchers, including doctrinal researchers, are very well placed to exploit law *as data*. We can leave (at least some of) the law to lawyers, not out of turf wars[14] but because there is legal *information* that is fundamentally human-made that they can turn into a usable research datapoint/datum. To retrieve this information, which exists within the text (and sometimes, as will be seen later, trickily exists *without a text*), researchers must position themselves as lawyers.

What this leaves is simply 'legal data'.[15] It is law turned into data, to become a piece of reality to be inputted into research that may or may not be about the law itself. In and of itself, a law can be a data point, or it can be a data source for its content, which is the true datum the researcher is interested in (a specific article? multiple articles?). For a more extensive example of the many ways to look for potential data: a ruling, in and of itself, can be a datapoint (although stopping at its existence or non-existence borders on metadata). The entire text of the ruling could be treated as data,[16] as can the topic(s) covered,[17] a standard used,[18] the presence of cross-reference to other jurisdictions,[19] the vocabulary and expressions used,[20] the precedents cited[21] and the actual outcome of a case.[22] All can be data, but the researcher must switch their lawyer's hat for their social scientist cap to pinpoint *what* exactly will be the datum, in this broad data source understood by the lawyer – what piece of reality are they looking for? It is at this exact interstice between law and social science, where the frontier between the lawyer and the social scientist can (and must) blur, that the complexity of law as data lies.

2. Fitting Law into Research Data: Round Peg for a Five-dimensional Hole

This section identifies the first set of challenges: the complexities of transforming law into data. This requires untangling the knot of an often messy and iterative process. The goal is to break down which difficulties can appear, and how they may call for different solutions or mitigation strategies.

It is important here to recall that social sciences, contrary to natural sciences, also battle with the issue of 'turning' a raw material into data, whether this material is explicitly textual or not. To grapple with it and connect data to theory, social scientists rely on three steps: conceptualisation, operationalisation, and coding. Each presents specific challenges for academics working with law as research data.

---

[10] Webley, "The Why and How of Conducting a Sociolegal Empirical Research Project," 63.
[11] McKnight, "The Fourth Act in Socio-Legal Scholarship," 111.
[12] The 'official law' as opposed to the 'unofficial law'; see Bens, "Ethnographic Legal Studies."
[13] Benoît, "Quanteda: An R Package for the Quantitative Analysis of Textual Data."
[14] Frankenreiter and Livermore, "Computational Methods in Legal Analysis."
[15] Bhat, Idea and Methods of Legal Research, 10.
[16] Aletras, "Predicting Judicial Decisions of the European Court of Human Rights"; Medvedeva, "Using Machine Learning to Predict Decisions of the European Court of Human Rights"; Sadl, "Can Quantitative Methods Complement Doctrinal Legal Studies?"
[17] Helfer, "Legal Integration in the Andes."
[18] Abrusci, Judicial Convergence and Fragmentation in International Human Rights Law; Webb, International Judicial Integration and Fragmentation.
[19] Law, "The Limits of Global Judicial Dialogue"; Law, "Judicial Comparativism and Judicial Diplomacy"; Plan, "A Theory of International Strategic Judicial Dialogue."
[20] Genovese, "Administering Harm."
[21] Alschner, "The Growing Complexity of the International Court of Justice's Self-Citation Network"; Larsson, "Speaking Law to Power."
[22] Carrubba, "Judicial Behavior Under Political Constraints."

2.1 Measuring the Immeasurable

The first challenge of using law as data is the difficulty of developing proper *conceptualisation and operationalisation* of a key legal concept or phenomenon of interest. The conceptualisation is the definition on which the researcher settles; the operationalisation is what this definition means in terms of constitutive elements of the concept. Legal concepts and phenomena can have flexible – and heavily debated – *definitions*. But in the context of empirical work, a definition must allow for concrete operationalisation. Two options are open to the researcher at this point, both with their own advantages and pitfalls: thin conceptualisations or thick conceptualisations.

First, the researcher can decide to adopt a narrow conceptualisation, and therefore a narrow operationalisation. One should not consider 'narrow' here as being pejorative in any way: better an explicit, transparent, narrow definition with clear operationalisation and subsequent coding than an ambitious definition that ends up not being suitable for empirical goals. One issue, however, can be the narrowing or hyper-specification of concepts usually considered broader in doctrinal debates, leading to difficulties in discipline-wide debate or even between empirically minded researchers.

The illustration here will be drawn from research on 'judicial dialogue', at a global scale. More doctrinal literature tends to have a broad, quasi-sociological understanding of the notion: covering cross-referencing between independent courts from different jurisdictions (both national and international),[23] procedures linking courts institutionally through preliminary references,[24] patterns of overall convergence in their case law, the practice of meetings and the exchange of information,[25] all the way to more network-based approaches involving not just the judges but also experts working with the court,[26] a 'transnational judicial public sphere'[27] and often a mix of all of these.

But this very broad approach, while incredibly relevant and capturing a real socio-legal *je ne sais quoi*, is of course very tricky to properly operationalise. This has led authors to use narrower definitions to empirically explore the existence, scale, relevance and consequences of such 'judicial dialogue'. For example, Law and Chang attempt to define it as the use of external citation, and 'judge to judge' exchanges – the former explored quantitatively in the case law of the Taiwanese Constitutional Court and the latter through interviews.[28] Webb has a close approach, where judicial dialogue is 'the citation, discussion, evaluation, application, or rejection of decisions'.[29] Almeida operationalises specifically that 'the dialogue between international courts occurs predominantly via external citation' to find that the dialogue between the International Court of Justice (ICJ) and the Inter-American Court of Human Rights (IACtHR) is heavily asymmetric – in favour of the ICJ.[30] On the other hand, Mauès et al[31] break down judicial dialogue between domestic courts and the Inter-American human rights system into three components: 'Hierarchy and Direct Effect of International Treaties', 'principle of Consistent Interpretation' and 'judicial postures' – resistance, engagement and convergence. Abrusci finds that the dialogue between International Courts is what *leads* to convergence, thereby excluding convergence from the operationalisation of said dialogue.[32] Plan similarly adopts the posture of the dialogue being the process leading to an outcome in the case law;[33] both works focus on the operationalisation of the outcome, rather than the more elusive *process* of the dialogue. This results in a spectacularly rich literature from which drawing conclusions, even at the national or regional level, is difficult: a fragmented literature that struggles to connect within its own branches and with broader doctrinal works on judicial dialogue.

Researchers can then be tempted to take the opposite approach and adopt maximalist definitions. As mentioned, these are more difficult to operationalise, necessitating more resources to collect and code relevant data. Moreover, it is difficult to find a maximalist definition that is not subject to ample debate. One reason why there is a piecemeal approach to judicial dialogue in the socio-legal literature may be that there is no consensus definition. It does not sit alone at the table of 'debated yet fundamental socio-legal concepts with a blurry definition' and fuzziness is not a good quality for a concept to have when it comes to tractability for operationalisation.[34] How could one agree on a definition of 'rule of law' that would lend itself to a *measurement*? As stated by Botero and Ponce, who developed the World Justice Project's Rule of Law Index, 'While the

---

[23] Slaughter, "A Typology of Transjudicial Communication"; Slaughter, A New World Order.
[24] Baldinger, Vertical Judicial Dialogues in Asylum Cases.
[25] Allard, Les juges dans la mondialisation, La nouvelle révolution du droit; Allard and Waeyenberge, "De la bouche à l'oreille ?"
[26] Claes, "Are You Networked Yet?"; Jackson, Constitutional Engagement in a Transnational Era.
[27] Traisbach, "A Transnational Judicial Public Sphere as an Idea and Ideology."
[28] Law, "The Limits of Global Judicial Dialogue."
[29] Webb, International Judicial Integration and Fragmentation.
[30] Almeida, "The Asymmetric Judicial Dialogue Between the ICJ and the IACtHR."
[31] Maués, "Judicial Dialogue Between National Courts and the Inter-American Court of Human Rights."
[32] Abrusci, Judicial Convergence and Fragmentation in International Human Rights Law.
[33] Plan, "A Theory of International Strategic Judicial Dialogue."
[34] Gerring, Finding Your Social Science Project, 240.

principles embedded in [scholarly and historic] definitions [of the Rule of Law] provide a common ground for discussion, they cannot be easily employed by practitioners and policymakers.'[35] A solution could be to adopt a flexible definition, where the researcher can then operationalise different components independently, as needed.[36] But while satisfying from a lawyer's perspective, this is an issue for the social scientist: as with 'judicial dialogue', this is a coherent solution at the scale of the research(er), but not at the scale of the broader field, where flexibility of definition hinders scholarship-wide discussion. There is a pressing need to establish a way to measure the rule of law, particularly in a context where there is broad consensus that the rule of law is threatened in multiple countries, and indicators can be very compelling for socio-legal researchers. Indeed, they allow for efficient cross-country comparison and analysis over time and can be more convincing when engaging with policy-makers.[37] There is a need for concrete data from which the magnitude of a rule of law backslide can be assessed, as well as the manner in which this is done and the trends that can possibly be identified, with the promise of objectivity.

There are, however, significant obstacles in the way of operationalising a large, broadly defined concept and this ends up blurring the boundaries with other neighbouring but conceptually different notions. Continuing with the theme of the rule of law, an illustration can be provided through the comparison of different broadly used indicators of *rule of Law* and *democracy*, and specifically their respective codebooks. Polity5 is an indicator of democracy, but one of the components of the score a state obtains is 'the existence of institutionalized constraints on the exercise of power by the executive',[38] which is more closely related to the rule of law than democracy itself (although it would be relevant for *liberal* democracy, showing the importance of precise and transparent conceptualisation). The World Justice Project includes fundamental rights in its measure of the rule of law – not problematic if one endorses a thick definition of the rule of law, but more so for researchers embracing a thinner definition. Freedom House, for its part, proposes a score for civil liberties, of which 'rule of law' is a part; however, rule of law is not part of Freedom House's political rights measurement.[39] Of course, fundamental rights, civil liberties, democracy and the rule of law are all neighbouring concepts, which explains this intrusion of one into the measurement of the other. But still, it is problematic from the perspective of conceptual clarity, and how we may want more interaction between normative, doctrinal work and empirical approaches for these notions. Additionally, these small differences, added to other methodological choices that the institutions creating the indicators adopted, can lead to true differences in how well they can capture actual changes in the rule of law and democracy, including in Europe today.[40] It is therefore fundamental to always consider which methodological and conceptual trade-offs were weighted – in other words, where the balance between the perspective of the social scientist and of the lawyer was struck – to tackle concepts that are still doctrinally debated today.

2.2 Coding the Uncodable: Legal Data and the Myth of Legal Categories

Whereas, in the previous section, the issue laid in a deliberate choice to look at data from very different and potentially mutually exclusive angles (conceptualisation and measurement), here there is a potential disagreement about the interpretation of similar data with the same analytical frame (coding).

Coding raw data involves the researcher engaging in data reduction,[41] transitioning either from conceptualisation to data or the other way around, from data to concept-building. This often takes the form of a codebook or analytical code: a set of rules which will aid the researcher with the justification of their interpretation of the raw data. Lawyers are very familiar with this practice and with its difficulties: applying law often consists of having ready-made categories and finding the one most relevant to a situation. 'Does this situation fulfil all the required elements to qualify as manslaughter?' is not intellectually far from 'Does this ruling fulfil all the required elements to qualify as progressive?' or 'Does this ruling fulfil the required elements to qualify as a convergence with another court?' in situations where the researcher has done the prerequired work of defining exactly what will count as 'a progressive ruling' or 'judicial convergence'. But coding the law is vulnerable to both random and systematic errors, raising issues regarding the validity and reliability of the coding.[42]

This leads to the first main challenge of coding law as data: that of categorising the uncategorisable. Reality does not always *fit* the codebook. Menkel Meadow identifies two strands of this issue. The first is the risk that 'coding and quantification may tend to assimilate in coding boxes and classifications that which is not really uniform', compounded by the second risk that

---

[35] Botero and Ponce, "Measuring the Rule of Law," 8.

[36] See, for example, Boon, Lawyers and the Rule of Law.

[37] Rothenberg, "Field-based Methods of Research on Human Rights Violations."

[38] Marshall, "POLITY5: Dataset Users' Manual," 14.

[39] Freedom House, "Freedom in the World 2024 Methodology Questions."

[40] Rohlfing, "Death by a Thousand Cuts."

[41] Miles, Qualitative Data Analysis.

[42] Drost, "Validity and Reliability in Social Science Research"; Chin, "Improving the Credibility of Empirical Legal Research"; Lombard, "Content Analysis in Mass Communication."

'empirical claims made in socio-legal studies that codify and classify phenomena, behaviors, and people … are far more complex and context variable than empirical "measurements" permit in many studies'.[43] This is a problem of limited commensurability. While comparing legal systems is perfectly appropriate, what about comparing laws in different legal systems?[44] Comparative law as a field is flexible enough to account for this variation when exploring specific legal concepts, questions and problems across jurisdictions; however, law-as-data is a more rigid approach: if laws adopted by the legislature themselves do not have a similar role in each legal system, would it make sense to look at them similarly with a similar codebook if one were to conduct a cross-country study? Should one compare a federal Act in Canada and a Gesetz in Germany? How can constitutional jurisdictions be compared through a strict, unified codebook when France has not one, but three, Supreme – and Constitutional – Courts, fully independent from each other? How is it possible to compare rulings of the Court of Justice of the European Union – short, terse, in the French tradition – and those of the European Court of Human Rights (ECtHR) – dozens and dozens of pages of facts, relevant law and legal analysis – with the same codebook? These are not impossible tasks, but the very nature of law – its diversity across jurisdictions, or even within a given jurisdiction – makes it particularly challenging. We can look for inspiration in content analysis coding, where 'latent coding' will get the most out of any text by looking at the meaning of the words, whereas 'manifest coding' only looks at the surface level of the words.[45] Law-as-data can be the object of a new 'legal coding', which demands of the coder nuance, transparency and often legal expertise, truly embracing the *law* in law-as-data.

But even with sufficient expertise and transparency, the law is simply, by its nature, often open to scholarly debate and interpretation. Law *is not* code. Coding for a complex concept, even when directed by a thorough codebook, is still a human activity and always attributes an 'interpreted meaning to each individual datum'.[46] The lawyer's expertise and awareness of the inherent normative value of the law are here both their best tool and their Achilles' heel. On the one hand, lawyers are best placed to systematically review and code case law through (latent) content analysis,[47] or even notice the absence of a feature as being noteworthy. On the other hand, this exercise must be done with the perspective of a social scientist, guided first and foremost by the need to work from, or towards, operationalisation and conceptualisation.

A very helpful illustration was provided in a recent back-and-forth discussion following an article originally published by Helfer and Voeten (HV).[48] The authors investigated whether the ECtHR was 'walking back human rights' – for example 'whether ECtHR case law is shifting in a rights-restrictive direction' in a manner that is not made explicit in the majority rulings. The authors explained that their codebook was asking specifically:

> whether the separate opinion asserts that the majority overturns prior case law in one of three ways: (i) by *explicitly overturning* prior judgments (either in a progressive or conservative direction); (ii) by *implicitly or tacitly overturning* prior case law; or (iii) by *construing* prior case law too narrowly or too broadly, *ignoring* prior case law or *failing to apply* it.

and

> whether a separate opinion (i) disagrees with the majority over the application of one or more key legal doctrines *and, in addition*, (ii) asserts that the doctrine had been applied more broadly or narrowly in prior case law.[49]
> They found, overall, that the ECtHR was 'walking back' human rights, given the rise in the number (both absolute and relative) of such opinions since 1999.

This prompted a response from Stone Sweet, Sandholtz and Adenas (SSA),[50] who disagreed on multiple accounts. There was first a disagreement on the details of the codebook, with SSA deciding stricter rules that, according to them,[51] better reflected the operationalisation of what 'walking back human rights' means. For example, multiple disagreements come from a dissent not stating exactly which precedent the Court would be 'walking back' – something that matters to responders but not to the initial authors. Arguably, this could have been solved by ensuring closer coherence between what walking back was in theory (conceptualisation), what this entailed in a ruling or dissent (operationalisation) and what specifically the researcher should then look for in the rulings (coding).

---

[43] Menkel-Meadow, "Uses and Abuses of Socio-Legal Studies," 448–49.
[44] Especially in mapping projects such as PACE Committee on Equality and Non-Discrimination, "Discrimination Against Transgender People in Europe Report Doc. 13742"; Marguery, Mutual Trust Under Pressure.
[45] Neuman, Social Research Methods, 374–75.
[46] Saldana, The Coding Manual for Qualitative Researchers, 4.
[47] Hall, "Systematic Content Analysis of Judicial Opinions."
[48] Helfer, "Walking Back Human Rights in Europe?"
[49] Helfer, 811.
[50] "Stone Sweet, Dissenting Opinions and Rights Protection in the European Court."
[51] Stone Sweet, "Dissenting Opinions and Rights Protection in the European Court," Appendix B.

But even then there were coding-specific challenges within each team. HV mentioned that on some sub-questions of the coding, there was only a 77 per cent consistency between coders (research assistants). The issue raised here is that of intercoder reliability (ICR). HV noticed the issue beforehand and had conducted an initial check on 33 rulings to improve the codebook first. They then solved any remaining cases of inter-coder disagreement by recoding the litigious cases. Yet this seems like a lawyer's solution to a social science method problem. From a lawyer's perspective, the solution to a disagreement is to reread the law more carefully, but this is treating law as law, rather than law as data. Law as a research artefact means embracing a wide array of more sophisticated tools to evaluate ICR, which can account for potential systematic biases of coders, accidental agreement, the use of categorical or ordinal coding or the number of potential categories.[52] Dealing with persistently low ICR then calls for a potential diversity of solutions, of which reliance on the most experimented coder is only one.[53] SSA, in the document recording their coding of each case, demonstrate that in multiple instances, the three authors disagreed on their assessment between themselves, and with HV. For example, in *Van Der Heijden v. The Netherlands*, one coder identified *no Walking Back Dissent*, another found '*No – tossup*', and the other '*Yes – tossup*'. HV found that this was, in fact, a walking back dissent. When five highly established professors, specialising in international law and international courts, could not agree on what to do with six pages of dissent, giving the final decision on coding to the most experimented coder may not have been the solution – at least when wearing a social scientist's hat.

3.   When You Have a Hammer: From Data-based Research to Data-driven Research

The last issue potentially facing empirical legal researchers is that not all can be data and data cannot be all there is to legal research. This is far from being specific to legal research, as other social sciences have had to grapple previously with this issue, particularly in the 1970s behaviouralist turn of the field, all the way to the 1990s, or even today's, push for data-driven research in political science. Law presents its own kind of challenges when it comes to giving law-as-data a proper place in empirical scholarship. But the challenges themselves remain: an excessive focus on research agenda, fields or even regions where data exists, creating dark spots in the scholarship; attempts to make data where there is no (sufficient) material to be turned into data; and trying to solve moral, ethical or otherwise philosophical questions through data.

The first potential pitfall of an excessive focus on law-as-data as the only valid way to conduct socio-legal research is the risk of deciding on the worthiness of a research agenda or project based on whether a way exists to rely on law-as-data to conduct that research. This would mean deciding which issue, which country and which population, should be the focus of extensive research based only on data availability. For example, does a country have an easily accessible database of its legislation? Can it be scraped automatically? Are the judgments delivered by its judiciary easily accessible? How easy would the rulings be to code?

The issue here is that data availability is not equally distributed around the globe. The previous section has already covered how global indicators can struggle with this, but it is even more prevalent when researchers themselves need to collect law as raw data and turn it into usable data. Under the pressure of producing research outputs, publishing and defending the need to fund a research project, humanities or research itself, the socio-legal researcher may be pushed towards areas and countries where it is easier to turn law into data. But this means overlooking states, often from the Global South, that do not have a properly collected, easy-to-scrape and navigable database or repository of all legislations, decisions and other relevant legal acts. This can also lead to relying on data that exists, but is simply not of adequate quality for the research to be conducted – yet it is data, so it will be used. Other pitfalls of hyper-reliance on unreliable, but obtainable data for socioeconomic research have already been stressed, including how they maintain inequality of knowledge, and therefore policies furthering socioeconomic inequalities based on gender,[54] poverty[55] or countries belonging to the Global South.[56]

For example, researchers very rarely rely on the CJEU database's classification of topic/issue areas, as it seldom captures appropriately the issues under discussion, and can sometimes be too broad or too narrow. Therefore, many researchers have taken to use text-as-data machine learning techniques to identify the relevant issues and organise cases in a manner most appropriate for their research. But this can be done because the CJEU database, Curia, is very easily scrapable, and turning the text of each individual ruling into data while retaining all relevant metadata for each ruling is an easy task. All the rulings are present, the metadata on each ruling is extensive and the text is baked into the HTML code. On the other hand, the Economic Community of West African States (ECOWAS) Court database provides limited metadata on each ruling (for example, only the year of the ruling rather than a full date) and is much more challenging to accurately scrape as the rulings are not baked in

---

[52] Lombard, "Content Analysis in Mass Communication."
[53] O'Connor, "Intercoder Reliability in Qualitative Research."
[54] Perez, Invisible Women.
[55] Römer, Inequality Knowledge.
[56] Jerven, Poor Numbers.

the HTML code of a webpage. Instead, a web scraper would need to be combined with an OCR algorithm to get the text from a downloadable PDF (actually a scanned) document for each ruling individually. While doable, this is a much more challenging task for a legal researcher who has limited time to invest in developing computer science skills. Another illustration is provided by Hildebrandt,[57] commenting on Aletras et al's use of the 'facts' section of ECtHR rulings. Aletras[58] used this section as a neutral proxy for the facts of the case themselves – data available and ready to be exploited by a socio-legal researcher. But as Hildebrandt points out, the recollection of facts by the Court is not neutral – while presenting itself as such, it is still a section written by the Court, baked into the judgment itself. A presentation of the facts in the submission of the parties would be different. As she notes:

> This is an example of opting for 'low hanging fruit' (easy to obtain training data), which raises some issues, as it implies that the system draws its conclusions based on the Court's articulation of the facts of the case. As the authors note, the Court probably formulates the facts in a way that is conducive to fit the conclusion.[59]

Then, the issue can lie in the unavailability of a document. The CJEU is notoriously secretive about the archives of its own case law, with the Observations of the Commission unavailable as a matter of principle, as are the Observations of intervening states. Amicus Curiae submitted to the ECtHR are similarly not disclosed by the Court (although the submitting organisation can decide to make them public). Working and preparatory documents for specific legislation, hearings transcripts and preparatory documents for all sorts of procedures, judgments and legislation, are sometimes unavailable despite being of vital interest to socio-legal research.

The second challenge when it comes to over-reliance on data is that in law, perhaps more than in other fields of social sciences, there is sometimes simply no actual law to turn into data. In this category is, first, all *unwritten* law. International law is full of rules and sources of law which are, by definition, unwritten and constantly debated: *jus cogens*, customary law, general principles of international law, and so on. Their value as *law, from the lawyer's perspective,* is undisputed, but one would be hard-pressed to collect them all systematically. The same goes for many constitutional conventions in UK constitutional law or the French *Principes Fondamentaux Reconnus dans les Lois de la République*[60] – with a clear source, but an ever-expanding list. The researcher can survey relevant case law to at least identify all items of a particular type of unwritten law; Petersen did so when exploring rulings where the ICJ *did* identify customary international law,[61] for example. This may not always be satisfying from a lawyer's perspective, reducing unwritten law to what has been written, in the end, in a ruling. After all, the ICJ does not create *jus cogens* norms, it merely identifies them. Some methods drawn from legal ethnographers, used to working with 'living customary law',[62] could help work around this lack of pre-existing clear data sources for international customary law, for example. Ethnographers have already relied on fieldwork such as focus groups, interviews and participant observation to gather data about norms in African custom-based legal systems to great success, for instance.[63]

Still, this would not apply to all unwritten law. Additionally, not all unwritten law is based on sociological practice – General Principles of European Law are discovered by the CJEU, but their acknowledgement does not follow the *de facto* behaviour of actors, only their implicit presence in EU law until then. There are limits to what a switch in perspective from lawyer to social scientist can accomplish because the unwritten law of the ethnographer is not always the unwritten law of the lawyer.

The urge to try to fill in this blank instead of going around it can lead to the use of a proxy for this non-existent data. For example, Aletras et al tackled the unavailability of briefs submitted by parties to the ECtHR by working with the assumption that 'the text extracted from published judgments of the Court bears enough similarities with [the briefs] and can therefore stand as a (crude) proxy for, applications lodged with the Court as well as for briefs submitted by parties in pending cases', for the law section of the judgment. However, as Hildebrandt once again points out, at the very least this does not hold across all cases: when a case is deemed inadmissible, the 'relevant law' section of the case is empty.[64]

Additionally, *more* data will not always solve the issue of legal data needing to be contextualised to be leveraged. The legal debates that both law-makers at large and all gravitating around them are engaged in are steeped into their own legal and political contexts, which are not transferable across courts and cannot be explained in terms of data. Yet the researcher must

---

[57] Hildebrandt, "Algorithmic Regulation and the Rule of Law."

[58] Aletras, "Predicting Judicial Decisions of the European Court of Human Rights."

[59] Hildebrandt, "Algorithmic Regulation and the Rule of Law," 7. See also Hall, "Systematic Content Analysis of Judicial Opinions."

[60] Fundamental Principles Acknowledged in the Laws of the Republic.

[61] Petersen, "The International Court of Justice and the Judicial Politics of Identifying Customary International Law."

[62] Zenker, *The State and the Paradox of Customary Law in Africa*.

[63] Bierschenk, "The Everyday Functioning of an African Public Service"; Otieno Ngira, "Understanding Children's Rights from a Pluralistic Legal Context."

[64] Hildebrandt, "Algorithmic Regulation and the Rule of Law."

know them to make the best of law-as-data. How is it possible to leverage a US Supreme Court ruling, an article written by RBG as a lawyer or a speech by Clarence Thomas, as data, without understanding the legal and philosophical questions raised by gender-based discrimination, affirmation action or the textualism–originalism debate. How can the rulings of the CJEU, the speeches of its President or joint declarations with the ECtHR be used as data without the un-datable understanding of integration, autonomy of EU law and subsidiarity? As stated by Sadl and Olsen, leading researchers in the field of machine-learning law-as-data:

> Corpus linguistic analysis does not provide any conclusive answers as to the legal importance and relevance of individual texts. Nor does it distinguish between more and less important cases and more and less legally relevant concepts … To answer any questions about why specific patterns occur, the findings must be interpreted.[65]

Finally, one must contend that no matter how effective a hammer law-as-data is, some questions in the legal scholarship require a screwdriver. More data, or even better data, will not answer all questions, just as more data has not put an end to ethical, moral and philosophical debates. Data will not save us, because some fundamental legal debates do not grow from epistemological differences, but from metaphysical, moral and ontological ones. Normative doctrinal questions are also the essence of legal research. They include the aforementioned debates on originalism and textualism in US constitutional law, or the proper content to give to the requirement of 'intent' for the crime of genocide and how to actually define and use the principle of subsidiarity in EU law. Of course, law-as-data can give more insight, and widen the empirical ground on which we can build a richer jurisprudential discourse. Tan et al recently conducted a qualitative exploration of the constitutional case law of the New Zealand High Court, based on an updated categorisation of Bobbit's modalities. They were able to identify the preferred mode of this Court, with specific and transparent data to back up this claim, demonstrating how useful bridges can created across the various strands of legal research.[66] But if one day the Hart–Dworkin debate reaches an end, if we all rally behind either the interest theory or the will theory, if we agree on what is truly the role of the judge, it will not be thanks to having enough data at last.

Furthermore, *more* data will not save us because more empirical information is not always the most relevant addition to the state of knowledge. Taking a page of critical legal studies and interpretivist approaches at large, it may not always be the most useful strategy. Exploring the racial health disparities in the Trump administration's response to the COVID-19 pandemic, Hatch paraphrases Jackson by asking, 'What has the data done for us, lately?'[67] More data may simply be fed into the current power structures to safeguard inequalities as opposed to addressing them, especially when the underlying conceptualisation on which they rely may carry with it a legacy of domination.[68] For Jones:

> If we accept a subjectivity of knowledge, and if we seek it out through our research, then we can highlight the multiple positionalities of those who speak, those who listen, and those who reply … [Without it] we would miss an opportunity for learning through such dialogue and, furthermore, run the risk of doing harm through the marginalisation of certain positionalities and the rendering invisible of certain experiences.[69]

4. The Way Forward: From the Researcher to the Research Environment

Legal scholarship is currently at a crossroads. Similar to economics and political science a few decades ago, it is confronted by the possibility of now using a new lens to approach its usual object of study: law, but as *data*. The challenge, as it has been for other disciplines, lies in making the best of this opportunity. While drafting this article, a timely academic debate broke out around an article seeking to empirically establish what the current landscape of international legal scholarship was like.[70] With the findings very puzzling for non-US scholars,[71] what followed was an in-depth analysis of the data source (which over-represented US student-run journals simply due to their availability on HeinOnline), the conceptualisation (a 'top author' was defined as one with a high number of citations) and the operationalisation (citations were counted only if in Bluebook style). In other words, these were the exact questions that any data-based research, but especially *law-as-data*-based research, must answer to avoid any bias. At the heart of discussions was not the principle of using empirics, but rather how these three steps (although not necessarily characterised as such) had seemingly been glossed over or had been signalled in the article without the necessary consequences being established.

---

[65] Sadl and Olsen, "Can Quantitative Methods Complement Doctrinal Legal Studies," 332.
[66] Tan, "How Does the High Court Interpret the 'Constitution'?"
[67] Hatch, "The Data Will Not Save Us."
[68] Mutua, Human Rights.
[69] Jones, "Qualitative Data and the Challenges of Interpretation in Transitional Justice Research," 229.
[70] Hathaway, "International Law Scholarship."
[71] Milanovic, "Horrible Metrics, Part Deux."

Since there is indeed an appetite for empirical work, how is it possible to avoid repeating these missteps? How can we address head on the challenges discussed in this article? A complete research (and, more broadly, academic) agenda goes beyond the scope of this article but a few fundamental building blocks can already be identified.

Many of the challenges come down to finding a balance between the perspective of the lawyer and that of the social scientist – and which of these should inform which step of the research process. Perhaps more than finding a definitive answer as to how this balance is to be found, it is important for the researcher to make a conscious and transparent decision about it, knowing which hat should be worn for which step. The social scientist hat should be worn when coming up with the research design, troubleshooting the method, deciding on a case selection strategy and assessing reliability or even statistical significance. The social sciences have a long tradition of developing the relevant tools and addressing the pitfalls, biases and challenges that can come up when conducting empirical studies. On the other hand, the lawyer's perspective is a powerful one for actually conducting the research: developing a codebook appropriate for legal data sources and legal text; identifying *what is* a deviant case, a typical case or an extreme case; conducting latent content analysis rather than manifest content analysis.[72] But how is it possible to ensure that socio-legal researchers can navigate this complex identity in their own research activities?

First, law-as-data fits within the greater question of how to conduct legal research in the first place, and what the role of empirical socio-legal research is within this greater family. A stronger methodological basis on law-as-data therefore starts by properly addressing why one should use a law-as-data approach for a given project. While bringing strong credentials from other disciplines, there are limits to the focus on data that need to be considered. Questions such as *why data* and *what does it mean to properly treat law as data*, as well as *what does one hope to accomplish*, are key steps to help parse out when to use and not use this approach – and hopefully not try to end the Hart–Fuller debate by throwing data at it. This also includes embracing law-as-data as having both a quantitative and a qualitative branch and understanding that there is more to qualitative epistemologies than simply being 'non-quantitative'. Making these decisions then lets the legal scholar know when and how to adopt the standpoint of a social scientist without sacrificing the idiosyncrasies of working with law.

Second, these individual decisions about the use of law-as-data must be supported by an appropriate academic environment. Rigorous social science methods also require proper training and technical skills, which are time-consuming to learn and use. These, however, are tremendously important for the next generation of legal researchers interested in law-as-data to be well equipped and find this balanced standpoint between social science and law. Appropriate training to give young researchers the tools to face the quandaries identified in the previous sections needs to include questions of ontology and epistemology in social science research, research design training and hands-on practice in skills such as manifest coding, latent coding, statistics and/or relevant computer coding skills. Lack of such training impedes the use of social science methods by legal researchers who would like to leverage them,[73] even as Universities in the United States and United Kingdom are showing interest in recruiting researchers with this skillset.[74] Additionally, as van Dijck et al argue, there is a need to develop a more general infrastructure facilitating law-as-data research.[75] The goal is here to broaden the availability of diverse data to avoid the issue of research biased towards some geographic or disciplinary area, as mentioned previously. Some methods require datasets in specific formats, which can be time-consuming and expensive to collect and tidy. Thorough law-as-data research may therefore require more substantial funding[76] for research assistance, technical support, data storage or software costs, which may be unusual for most funding streams tailored to legal research.

Conclusion

In a nutshell, as Perez has done previously for political theory before, I make a case for methodological naturalisation.[77] To retain the idiosyncrasies of law and truly leverage this opportunity, let us take *law*, rather than *data*, as a definitive starting point.[78] Let us take the risk to leave (at least some of) the law to lawyers in socio-legal research, or at least accept that they are equipped with unique tools to make the best of law as research data. This, however, means that socio-legal researchers using law-as-data must openly reckon with the challenges identified in this article, including conceptualisation, operationalisation and coding of law, as well as the associated questions of reliability, transparency and the pitfalls of data-driven research. Doing this requires individual self-reflection, with a constant dialectic between identities as a social scientist and as an expert jurist.

---

72 Hall, "Systematic Content Analysis of Judicial Opinions."
73 Siems, "Why Do We Do What We Do?"
74 Adler, "Stepwise Progression."
75 van Dijck, "Empirical Legal Research in Europe."
76 van Dijck, "Empirical Legal Research in Europe," 18.
77 Perez, "The Case for Methodological Naturalisation."
78 Godzimirska, "Taking Texts Seriously."

The importance of this dialectic must not be downplayed; instead, it must be accompanied and supported by the entire academic ecosystem, from postgraduate training up to research funding.

The goal is not to simply import methods from other disciplines, but to truly develop an approach to law-as-data that is tailored to what formal law (written or unwritten) *is*, how to work with it and what legal research aims to accomplish. We should strive to avoid bias, promote transparency and adopt a communitarian approach to knowledge production.[79] Much comes down to transparency at every step,[80] to allow for collective, collaborative growth of the discipline and development of best practices around working with law as data. At last, let us rejoin Shera who, reflecting on the turn to data and data management of information sciences in the 1970s, concluded that, 'The computer is here to stay; therefore it must be kept in its proper place as a tool and a slave, or we will become sorcerer's apprentices, with data data everywhere and not a thought to think.'[81]

---

[79] Merton, The Sociology of Science.
[80] Chin, "Improving the Credibility of Empirical Legal Research."
[81] Shera, "Librarianship and Information Science."

Bibliography

Abrusci, Elena. *Judicial Convergence and Fragmentation in International Human Rights Law: The Regional Systems and the United Nations Human Rights Committee.* Cambridge: Cambridge University Press, 2023.

Adler, Michael and Jonathan Simon. "Stepwise Progression: The Past, Present, and Possible Future of Empirical Research on Law in the United States and the United Kingdom." *Journal of Law and Society* 41, no 2 (2014): 173–202. https://doi.org/10.1111/j.1467-6478.2014.00663.x.

Aletras, Nikolaos, Dimitrios Tsarapatsanis, Daniel Preoțiuc-Pietro, and Vasileios Lampos. "Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective." *PeerJ Computer Science* 2 (2016): e93. https://doi.org/10.7717/peerj-cs.93.

Allard, Julie and Antoine Garapon. *Les juges dans la mondialisation, La nouvelle révolution du droit. La République des idées.* Paris: Seuil, 2005.

Allard, Julie and Arnaud Van Waeyenberge. "De la bouche à l'oreille ? Quelques réflexions autour du dialogue des juges et de la montée en puissance de la fonction de juger." *Revue interdisciplinaire d'etudes juridiques* 61, no 2 (2008): 109–129.

Almeida, Paula Wojcikiewicz. "The Asymmetric Judicial Dialogue Between the ICJ and the IACtHR: An Empirical Analysis." *Journal of International Dispute Settlement* 11, no 1 (2020): 1–19. https://doi.org/10.1093/jnlids/idz015.

Alschner, Wolfgang and Damien Charlotin. "The Growing Complexity of the International Court of Justice's Self-Citation Network." *European Journal of International Law* 29, no 1 (2018): 83–112. https://doi.org/10.1093/ejil/chy002.

Baldinger, Dana. *Vertical Judicial Dialogues in Asylum Cases: Standards on Judicial Scrutiny and Evidence in International and European Asylum Law.* Amsterdam: Brill Nijhoff, 2015.

Benoît, Kenneth, Kohei Watanabe, Haylan Wang, Paul Nulty, Adam Obeng, Stefan Müller and Akitaka Matsuo. "Quanteda: An R Package for the Quantitative Analysis of Textual Data." *Journal of Open Source Software* 3, no 30 (2018): 774. https://doi.org/10.21105/joss.00774.

Bens, Jonas and Larissa Vetters. "Ethnographic Legal Studies: Reconnecting Anthropological and Sociological Traditions." *The Journal of Legal Pluralism and Unofficial Law* 50, no 3: 239–254. https://doi.org/10.1080/07329113.2018.1559487.

Bhat, Author P. Ishwara. *Idea and Methods of Legal Research*. Oxford: Oxford University Press, 2019.

Bierschenk, Thomas. "The Everyday Functioning of an African Public Service: Informalization, Privatization and Corruption in Benin's Legal System." *The Journal of Legal Pluralism and Unofficial Law* 40, no 57: 101–139. https://doi.org/10.1080/07329113.2008.10756619.

Boon, Andrew. *Lawyers and the Rule of Law*. Oxford: Hart, 2022.

Bos, Kees van den. *Empirical Legal Research: A Primer*. Cheltenham: Edward Elgar, 2020.

Botero, Juan Carlos and Alejandro Ponce. "Measuring the Rule of Law." World Justice Project Working Paper 001 (2011). https://worldjusticeproject.org/our-work/publications/working-papers/measuring-rule-law.

Carrubba, Clifford J., Matthew Gabel and Charles Hankla. "Judicial Behavior Under Political Constraints: Evidence from the European Court of Justice." *American Political Science Review* 102, no 4 (2008): 435–452. https://doi.org/10.1017/S0003055408080350.

Chin, Jason M, Alexander C DeHaven, Tobias Heycke, Alexander O Holcombe, David T Mellor, Justin T Pickett, Crystal N Steltenpohl, Simine Vazire and Kathryn Zeiler. "Improving the Credibility of Empirical Legal Research: Practical Suggestions for Researchers, Journals and Law Schools." *Law, Technology and Humans* 3, no 2 (2021): 107–132. https://doi.org/10.5204/lthj.1875.

Chin, Jason M. and Kathryn Zeiler. "Replicability in Empirical Legal Research." *Annual Review of Law and Social Science* 17 (2021): 239–260.

Claes, Monica and Maartje de Visser. "Are You Networked Yet? On Dialogues in European Judicial Networks." *Utrecht Law Review* 8, no 2 (2012): 100–114. https://doi.org/10.18352/ulr.197.

Cope, Kevin L, Cosette D Creamer and Mila Versteeg. "Empirical Studies of Human Rights Law." *Annual Review of Law and Social Science* 15 (2019): 155–182. https://doi.org/10.1146/annurev-lawsocsci-101317-031123.

Dale, Robert. "Law and Word Order: NLP in Legal Tech." *Natural Language Engineering* 25, no 1 (2019): 211–217. https://doi.org/10.1017/S1351324918000475.

Drost, Ellen A. "Validity and Reliability in Social Science Research." *Education Research and Perspectives* 38, no 1 (2011): 105–123.

Frankenreiter, Jens and Michael A Livermore. "Computational Methods in Legal Analysis." *Annual Review of Law and Social Science* 16 (2020): 39–57.

Freedom House. "Freedom in the World 2024 Methodology Questions." Washington, DC: Freedom House, 2024. https://freedomhouse.org/sites/default/files/2024-02/FIW_2024%20MethodologyPDF.pdf.

Genovese, Emma. "Administering Harm: The Treatment of Trans People in Australian Criminal Courts." *Current Issues in Criminal Justice* 36, no 2 (2023): 177–196. https://doi.org/10.1080/10345329.2023.2231112.

Gerring, John and Jason Seawright. *Finding Your Social Science Project: The Research Sandbox – Strategies for Social Inquiry*. Cambridge: Cambridge University Press, 2022.

Godzimirska, Zuzanna and Anne Lise Kjaer. "Taking Texts Seriously: The Language of International Law." *Nordic Journal of International Law* 93, no 1 (2024): 68–89. https://doi.org/10.1163/15718107-bja10074.

Hall, Mark A and Ronald F Wright. "Systematic Content Analysis of Judicial Opinions." SSRN Scholarly Paper, 2006. https://papers.ssrn.com/abstract=913336.

Hatch, Anthony Ryan. "The Data will Not Save Us: Afropessimism and Racial Antimatter in the COVID-19 Pandemic." *Big Data & Society* 9, no 1 (2022): 20539517211067948. https://doi.org/10.1177/20539517211067948.

Hathaway, Oona A and John Bowers. "International Law Scholarship: An Empirical Study." *Yale Journal of International Law* 49 (2024). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4817645.

Helfer, Laurence and Karen J Alter. "Legal Integration in the Andes: Law-making by the Andean Tribunal of Justice." *European Law Journal* 17 (2011): 701–715. https://doi.org/10.1111/j.1468-0386.2011.00574.x.

Helfer, Laurence R and Erik Voeten. "Walking Back Human Rights in Europe?" *European Journal of International Law* 31, no 3 (2020): 797–827. https://doi.org/10.1093/ejil/chaa071.

Hildebrandt, Mireille. "Algorithmic Regulation and the Rule of Law." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, no 2128 (2018): 20170355. https://doi.org/10.1098/rsta.2017.0355.

Jackson, Vicki. *Constitutional Engagement in a Transnational Era*. Oxford: Oxford University Press, 2013.

Jerven, Morten. *Poor Numbers: How We are Misled by African Development Statistics and What to Do About It*. Ithaca, NY: Cornell University Press, 2013.

Jones, Briony. "Qualitative Data and the Challenges of Interpretation in Transitional Justice Research." In *Routledge Handbook of Socio-Legal Theory and Methods*, edited by Naomi Creutzfeldt, Mark Mason and Kirsten McConnachie, 220–231. London: Routledge, 2019.

Kritzer, Herbert M. *Advanced Introduction to Empirical Legal Research*. Cheltenham: Edward Elgar, 2021.

Larsson, Olof, Daniel Naurin, Mattias Derlén and Johan Lindholm. "Speaking Law to Power: The Strategic Use of Precedent of the Court of Justice of the European Union." *Comparative Political Studies* 50, no 7 (2017): 879–907. https://doi.org/10.1177/0010414016639709.

Law, David S. "Judicial Comparativism and Judicial Diplomacy." *University of Pennsylania Law Review* 163, no 4 (2015): 927–1036.

Law, David and Wen-Chen Chang. "The Limits of Global Judicial Dialogue." *Washington Law Review* 86, no 3 (2011): 523.

Livermore, Michael A and Daniel N Rockmore, eds. *Law as Data: Computation, Text, & the Future of Legal Analysis*. Santa Fe, NM: Santa Fe Institute of Science, 2019.

Lombard, Matthew, Jennifer Snyder-Duch and Cheryl Campanella Bracken. "Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability." *Human Communication Research* 28, no 4 (2002): 587–604. https://doi.org/10.1111/j.1468-2958.2002.tb00826.x.

Marguery, Tony, ed. *Mutual Trust Under Pressure, the Transferring of Sentenced Persons in the EU: Transfer of Judgments of Conviction in the European Union and the Respect for Individuals' Fundamental Rights*. Oisterwijk: Wolf, 2018.

Marshall, Monty G and Ted Robert Gurr. *POLITY5: Dataset Users' Manual*. Polity Project, 2020.

Maués, Antonio Moreira, Breno Baía Magalhães, Paulo André Nassar and Rafaela Sena. "Judicial Dialogue Between National Courts and the Inter-American Court of Human Rights: A Comparative Study of Argentina, Brazil, Colombia and Mexico." *Human Rights Law Review* 21, no 1 (2021): 108–131. https://doi.org/10.1093/hrlr/ngaa047.

McKnight, Janet. "The Fourth Act in Socio-Legal Scholarship: Playing with Law on the Sociological Stage." *Qualitative Sociology Review* 11 (2015): 108–124.

Medvedeva, Masha, Michel Vols and Martijn Wieling. "Using Machine Learning to Predict Decisions of the European Court of Human Rights." *Artificial Intelligence and Law* 28, no 2 (2020): 237–266. https://doi.org/10.1007/s10506-019-09255-y.

Menkel-Meadow, Carrie. "Uses and Abuses of Socio-Legal Studies." In *Routledge Handbook of Socio-Legal Theories and Methods*, edited by Naomi Creutzfeldt, Mark Mason and Kirsten McConnachie, 35–57. London: Routledge, 2019.

Merton, Robert K. *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago: Chicago University Press, 1973.

Milanovic, Marko. "Horrible Metrics, Part Deux." *EJIL: Talk!* (blog), 9 May 2024. https://www.ejiltalk.org/horrible-metrics-part-deux.

Miles, Matthew B, A Michael Huberman and Johnny Saldaña. *Qualitative Data Analysis: A Methods Sourcebook*, 3rd ed. Thousand Oaks, CA: Sage, 2013.

Mitchell, Matthew. "Analyzing the Law Qualitatively." *Qualitative Research Journal* 23, no 1 (2022): 102–113. https://doi.org/10.1108/QRJ-04-2022-0061.

Mutua, Makau. *Human Rights: A Political and Cultural Critique*. Philadelphia: University of Pennsylvania Press, 2002.

Neuman, W Lawrence. *Social Research Methods: Qualitative and Quantitative Approaches*. 7th ed. Boston: Pearson, 2009.

O'Connor, Cliodhna and Helene Joffe. "Intercoder Reliability in Qualitative Research: Debates and Practical Guidelines." *International Journal of Qualitative Methods* 19 (2020). https://doi.org/10.1177/1609406919899220.

Otieno Ngira, David. "Understanding Children's Rights from a Pluralistic Legal Context: Multi-Legalities and the Protection of the Best Interests of the Child in Rural Kenya." *The Journal of Legal Pluralism and Unofficial Law* 53, no 3 (2021): 545–569. https://doi.org/10.1080/07329113.2021.1982170.

PACE Committee on Equality and Non-Discrimination. *Discrimination Against Transgender People in Europe Report*. https://assembly.coe.int/nw/xml/xref/xref-xml2html-en.asp?fileid=21736.

Perez, Caroline Criado. *Invisible Women: Exposing Data Bias in a World Designed for Men*. London: Chatto & Windus, 2019.

Perez, Nahshon. "The Case for Methodological Naturalisation: Between Political Theory and Political Science." *The British Journal of Politics and International Relations*, 25, no 4 (2022). https://doi.org/10.1177/13691481221113218.

Petersen, Niels. "The International Court of Justice and the Judicial Politics of Identifying Customary International Law." *European Journal of International Law* 28, no 2 (2017): 357–385. https://doi.org/10.1093/ejil/chx024.

Plan, Audrey Mathilde. "A Theory of International Strategic Judicial Dialogue: Convergence and Divergence Between the Court of Justice of the European Union and the European Court of Human Rights." PhD thesis, Trinity College Dublin, 2024. http://www.tara.tcd.ie/handle/2262/104842.

Robaldo, Livio, Serena Villata, Adam Wyner and Matthias Grabmair. "Introduction for Artificial Intelligence and Law: Special Issue 'Natural Language Processing for Legal Texts.'" *Artificial Intelligence and Law* 27, no 2 (2019): 113–115. https://doi.org/10.1007/s10506-019-09251-2.

Rohlfing, Regitze Helene and Marlene Wind. "Death by a Thousand Cuts: Measuring Autocratic Legalism in the European Union's Rule of Law Conundrum." *Democratization* 30, no 4 (2023): 551–568. https://doi.org/10.1080/13510347.2022.2149739.

Römer, Felix. *Inequality Knowledge: The Making of the Numbers about the Gap Between Rich and Poor in Contemporary Britain*. Berlin: De Gruyter Oldenbourg, 2023.

Rothenberg, Daniel. "Field-Based Methods of Research on Human Rights Violations." *Annual Review of Law and Social Science* 15 (2019): 183–203. http://dx.doi.org/10.1146/annurev-lawsocsci-102612-133939.

Sadl, Urska and Henrik Palmer Olsen. "Can Quantitative Methods Complement Doctrinal Legal Studies? Using Citation Network and Corpus Linguistic Analysis to Understand International Courts International Legal Theory." *Leiden Journal of International Law* 30, no 2 (2017): 327–350. https://doi.org/10.1017/S0922156517000085.

Saldana, Johnny. *The Coding Manual for Qualitative Researchers*, 3rd ed. Thousand Oaks, CA: Sage, 2015.

Shera, Jesse H. "Librarianship and Information Science." In *The Study of Information: Interdisciplinary Messages*, edited by Fritz Machlup and Una Mansfield, 379–388. Chichester: Wiley-Interscience, 1983.

Siems, Mathias and Daithi Mac Sithigh. "Why Do We Do What We Do? Comparing Legal Methods in Five Law Schools Through Survey Evidence." In *Rethinking Legal Scholarship: A Transatlantic Interchange*, edited by Rob van Gestel, Hans Micklitz and Edward L Rubin. New York: Cambridge University Press, 2015. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2625473.

Slaughter, Anne-Marie. *A New World Order*. Princeton, NJ: Princeton University Press, 2005.

Slaughter, Anne-Marie. "A Typology of Transjudicial Communication." *University of Richmond Law Review* 29, no 1 (1994): 99–137.

Stone Sweet, Alec, Wayne Sandholtz and Mads Andenas. "Dissenting Opinions and Rights Protection in the European Court: A Reply to Laurence Helfer and Erik Voeten." *European Journal of International Law* 32, no 3 (2021): 897–906. https://doi.org/10.1093/ejil/chab057.

Tan, David, Tamsin Phillipa Paige, Despina Hrambanis and Joseph Green. "How Does the High Court Interpret the 'Constitution'? A Qualitative Analysis between 2019–21." *University of New South Wales Law Journal* 47, no 1 (2024): 177–210.

Traisbach, Knut. "A Transnational Judicial Public Sphere as an Idea and Ideology: Critical Reflections on Judicial Dialogue and Its Legitimizing Potential." *Global Constitutionalism* 10, no 1 (2021): 186–207. https://doi.org/10.1017/S2045381720000295.

van Dijck, Gijs, Shahar Sverdlov and Gabriela Buck. "Empirical Legal Research in Europe: Prevalence, Obstacles, and Interventions." *Erasmus Law Review*, no 2 (2018): 105–119. https://doi.org/10.5553/ELR.000107.

Webb, Philippa. *International Judicial Integration and Fragmentation*. Oxford: Oxford University Press, 2013.

Webley, Lisa. "The Why and How of Conducting a Socio-legal Empirical Research Project." In *Routledge Handbook of Socio-Legal Theory and Methods*, edited by Naomi Creutzfeldt, Mark Mason and Kirsten McConnachie, 58–69. London: Routledge, 2019.

Zenker, Olaf and Markus Virgil Hoehne, eds. *The State and the Paradox of Customary Law in Africa*. London: Routledge, 2019.