5-2021

Quantitative legal research in Germany

Dirk HARTUNG
Singapore Management University, dirkhartung@smu.edu.sg

12. Quantitative legal research in Germany

Dirk Hartung

INTRODUCTION

This chapter introduces Germany as a place for quantitative legal studies. A country-wide study of big data law might be conducted in three ways, depending on focus: Big data *law* would focus on regulation and creation of, and interaction with, large amounts of data, as well as related legal domains, including data protection, privacy, industrial and intellectual property rights law.[1] *Big* data law would require the object of study to meet the definition of big data. Most of the currently relevant legal data in Germany lacks one or more properties customarily ascribed to big data,[2] as the datasets are often rather limited in volume, slow-growing and uniform. Lastly, big *data* law would put legal data and its science first, and suggest principles and methods for legal data creation, cleaning, analysis, interpretation and visualization. That is what this chapter is about.

While many of the ideas presented are generally applicable, they are discussed within the context of quantitative legal studies in Germany, where the discipline is at an early stage, and does not always provide sufficient examples to illustrate the ideas presented. Where possible, publications from Germany are used, complemented by papers from around the globe. Germany provides an interesting context for studying quantitative legal research for two reasons: it is a highly developed industrial country with an extensive justice system and a long tradition of jurisprudence, making it comparable to the United States, and therefore worth studying with established methods. It is, however, also a civil law country, one with a traditionally very doctrinal approach to the study of the law. As such it differs sufficiently from the United States so that a proper introduction seems necessary to guide scholars from other jurisdictions. Finally, the current, nascent phase of quantitative legal studies in Germany still allows for standards-setting and the development of best practices. Thus, Germany could profit immensely from attracting quantitative legal scholars from elsewhere and engaging in a global dialogue.

The underlying assumption of this chapter is that understanding the legal system as a complex adaptive system (CAS) might help researchers to better describe, analyse and grasp both its internal causes and effects as well as its place within the greater social system. Complex adaptive systems theory has been previously suggested as suitable for the legal system.[3] Subsequently, metrics and methods for studying legal complexity have been developed, including measuring variation over time.[4] They have also been applied to specific areas of law.[5] At last, this understanding has been accepted as a useful approach by the wider scientific community.[6]

To operationalize these insights for quantitative legal studies in Germany, researchers should endeavour to understand its legal system with data as a central theme. This chapter aims to provide starting points and define general best practices for such an endeavour, practices which, to produce meaningful insights, should consider a variety of sociocultural factors. The next section provides just such context.

DATA-BASED LEGAL RESEARCH IN GERMANY

A data-focused approach to jurisprudence is a well-established concept in Germany. Sociologists with a particular interest in the legal system include Karl Marx, Eugen Ehrlich, Max Weber, Theodor Geiger and Niklas Luhmann. Luhmann's main work on sociological systems theory shares many characteristics with the complex adaptive systems approach. It describes an almost identical concept but does not contain a mathematical and quantitative foundation – possibly because of a lack of computing power and data, and the undeveloped state of computational methods in social sciences in general, when Luhmann introduced his work in the early 1980s.

Since the 1970s, *Rechtssoziologie* (legal sociology or law and society) has become an independent sub-discipline of legal research, with dedicated professorships and research institutes at several German universities and law schools. Pursuant to the customary subdivision of doctrinal legal research into civil, criminal and public law, the empirical parts of legal sociology are known as *Rechtstatsachenforschung* (factual legal research), *Kriminologie* (criminology) and *Verwaltungswissenschaft* (administrative studies).[7] Relevant research can be found under these German terms in older publications.

While empirical legal research in the US has enjoyed a meteoric rise since the days of Roscoe Pound and Oliver Wendell Holmes Jr,[8] German scholars have been generally reluctant to adopt this perspective, holding that law is subject to its very own principles and should be studied only with its own set of methods.[9] Thus, an unfamiliarity with quantitative methods, and their irrelevance to both a career in legal academia and in law practice, have kept the field from catching on in Germany so far.[10] Though there exists a periodical, *German Journal of Law and Society*,[11] and a regular convening of the German Association for Law and Society,[12] legal sociology is currently in a state of crisis and reorganization nonetheless, with little mainstream attention, its representatives slowly vanishing from the academic stage.[13] Its core research questions, however, are not only of continuing interest, but also regularly examined – albeit often under different labels, the most popular of which seems to be "legal sociology" and "law and society".[14] The role of data in legal scholarship is vigorously discussed beyond the field of legal sociology.[15] While some scholars have argued in favour of a quantitative paradigm shift of legal research as a whole,[16] others have opposed – or at least strongly advocated for limiting the scope of – such a movement.[17] The discussion is fierce and in part ideological, bringing to mind early discussions around legal sociology and Marxism.

For those from countries in which empirical and quantitative legal studies are significant or even dominant parts of legal research, the situation in Germany might seem unusual. Empirical legal research is far from typical for German legal scholars. The vast majority of German legal scholarship is deliberately normative and doctrinal: many German legal scholars regard the word "dogma," which generally carries a negative connotation in other scientific contexts, as the very characterization of their work. They would describe a scientifically rigorous approach to law as very *dogmatisch* (dogmatic). Given this perspective, the existence of strong opinions on the merits of empirical legal research is far less surprising. Broaching the subject all but guarantees controversy, and is likely to trigger responses beyond the academic realm. The most recent wave of discussions was caused by the 2014 publication of a dissertation arguing for evidence-based jurisprudence.[18] The dissertation's author received a number of scientific awards, and the work was both highly praised and reasonably criticized.[19] Despite

this attention, the discussion has not yet led to perceivable changes to the greater landscape in legal academia.[20]

Another quantitative approach examines the law from a linguistic perspective.[21] The resulting discipline, legal linguistics, can be traced back to the 1970s, when it started to appear at some universities. By the late 1990s and during the early 2000s, it was established as a proper discipline, had defined a core of research interests and co-founded a research association.[22] Important research groups exist at the universities in Halle-Wittenberg, Heidelberg and Regensburg.[23]

Among the prolific research in the field of legal linguistics, scholars with a quantitative and technical focus would probably find studies on computer-assisted legal linguistics the most interesting.[24] These studies analyse corpora of various sizes with statistical methods to discover language patterns yielding insights into legal semantics. The research interest is fundamentally legal in nature, as these works aim to improve understanding of traditional, normative legal questions. As an auxiliary discipline for legislators, legal linguistics goes beyond dogmatic questions, attempting to understand ways legal texts can become more comprehensible.

Finally, and most recently, data-based legal research has been reintroduced to the German discussion under the label of "quantitative legal research."[25] In contrast to the slightly earlier work, this technology-heavy line of research does not strongly argue for a particular approach towards legal research in general. Instead, it focuses on studies and best practices for technology use in legal scholarship. Methods stem from data mining, text analytics, network science[26] and natural language processing.[27] It has generated some early and mildly positive reception.[28] Most recently, Bucerius Law School, a private school in Hamburg, has founded a Centre for Legal Technology and Data Science to provide an institutional framework for this line of research.

As this approach extensively uses advanced methods from computer science, it links the field to another well-established academic discipline – *Rechtsinformatik* (legal informatics). In Germany, legal informatics has its roots in the 1960s in mathematical structure theory applied to law by Herbert Fiedler,[29] and, over the next 50 years, developed into an interdisciplinary field.[30] Today, scholars from both law and computer science occupy professorial chairs dedicated to research at the intersection of law and computer science, with a broad research agenda including formalization and rule-based legal reasoning, XML specification, IT security and distributed ledger technologies. There are established research institutes for legal informatics at the Universities of Hannover[31] and Münster,[32] and Saarland in Saarbrücken.[33] A more technical focus on NLP and software engineering can be found among individual researchers at institutes in Hamburg,[34] Munich[35] and Heidelberg,[36] with Heidelberg's being first to establish an explicitly interdisciplinary graduate school in 2017.[37] Beyond these, there is an established society for law and informatics,[38] a specialist group within the general computer science society[39] and regular conferences.[40]

Within this community, research is produced for computer science, legal and interdisciplinary audiences. For this chapter's focus on quantitative studies, only part of the research on legal informatics is relevant: projects and publications on rule-based systems, formalizations and legal aspects of distributed ledger technologies as well as IT law in general do not fall under the above definition of big data law. Natural language processing within the legal domain, network science concerned with legal networks and other forms of quantitative legal studies do.

In summary, quantitative legal research can look back at a long tradition of using quantitative methods for legal research and examining the use of technology for the law. Yet, the application of methods from computer science and statistics is far from the mainstream within German legal scholarship. While the academic environment is not ideal, recent publications in highly renowned journals (Archivzeitschriften) indicate that quantitative methods are increasingly gaining popularity and sparking interest.[41] The most promising way for quantitative methods to continue on this trajectory is the publication of conclusive and relevant findings. Therefore, researchers need access to a precious resource – legal data. The following section examines in detail which data is available and how to unlock more.

ACCESS TO LEGAL DATA

This section introduces the current state of legal data, that is, the data produced by all three branches of government – legislative, executive and judicial – on federal, state and municipal levels, complemented by commentary from legal scholars. Availability of legal outputs such as statutes, regulations and court decisions – and corresponding data – differ significantly enough to address the topic for each branch separately. In general, data at the federal level is the most available and easiest to work with; data becomes more fragmented and sparse on the state and municipal levels. Somewhat reflecting the interest of quantitative legal scholarship and related academic discussion, as measured by number of publications, this section addresses judicial data first, followed by legislative, and finally by data generated by the executive branch related to administrative proceedings. It provides the current academic perspective on relevant legal impediments such as copyright or privacy restrictions, so that researchers are aware of their options when seeking access.

Access to Judicial Data

Legal scholars are naturally interested in judicial data, specifically in two types of documents, which differ in the amount of information contained: final court decisions and entire case files.[42] Decisions, long the focus of doctrinal legal research, also seem more relevant for points of law in quantitative studies, and are therefore discussed first. Discussed second are case files, which, while generally harder to access, often contain more socio-economic information about the parties, potentially of great interest for projects examining the relationship between law and society.

Court decisions

Given the civil law system in Germany, where precedent has historically been less of a consideration, the interest in high volumes of court decisions is relatively recent. Before the advent of computer-based text processing, handling several millions of files, let alone making sense of the relations among and within them and their legal content, easily overburdened individual researchers and even entire teams. In the past, most legal scholars were more interested in the legal holdings in individual cases than in general patterns over massive amounts of decisions. Consequently, judges determined on a case-by-case basis whether their decision merited publication, mostly publishing decisions if they subjectively deemed these to contain novel points of law or interesting legal questions. This approach strongly influences both the availability of

case decisions and the rules regarding the publication of and access to court decisions. Court operations are scaled to provide a relatively small number of decisions on a case-by-case basis mainly for their specific ruling.

Researchers may wish to note that final judgments and all other materials created by judges or clerks in their official capacity – such as guiding principles, excerpts or summaries – are not protected by copyright. While these pass the threshold of originality, constituting protectable creations, they also fall under the "official text" exemption of Section 5 *Urhebergesetz* (UrhG) (Copyright Act).[43] As is presented in more detail below, several databases containing judicial data are available in Germany. While they technically fall under a specific copyright rule set of *Datenbankurheberrecht* (database rights), the exception for "official texts" also applies to databases, according to the Federal Court of Justice – although this position is questionable from a European law perspective.[44] Additionally, the specific database rights regime contains a dedicated exemption for research projects in Art. 87c UrhG. As a result, copyright laws do not hinder access to court decisions residing within databases, though researchers might encounter copyright-based arguments by those unwilling to provide judicial data.

More importantly, the protection of personality rights and specifically privacy has far-reaching effects on the availability of judicial data. The German constitution combines its protection of human dignity in Art. 1 *Grundgesetz* (GG) (Basic Law) with the free expression of one's personality in Art. 2 GG into a general right of personality. From this basis, the Federal Constitutional Court created an individual right to informational self-determination,[45] which is of enormous public importance, and the foundation for the great importance of data protection in Germany. As a result, it is the dominant legal position that personal data must be deleted from judicial documents before they can be published. This has drawn criticism, as court proceedings in general are explicitly public, pursuant to Section 169 *Gerichtsverfassungsgesetz* (Court Constitutional Act). As parties are aware of this, it could be argued that they consent to their personal information being made public in the decision. And, if legal or even state entities are a party to the proceedings, it is particularly hard to see how a protection of their "personality rights," rights whose legal basis is arguably questionable for such entities, might be deemed mandatory. Still, with the recent boost of privacy laws under the *Datenschutzgrundverordnung* (General Data Protection Regulation, or GDPR),[46] it seems unlikely that this position will soon change.

Also, while courts have generally held that their decisions must be published, and granted the public an explicit right to access, in a recent criminal case, the Federal Court of Justice ruled that this general public right can be outweighed by personality and privacy rights of the parties.[47] While only applicable to criminal matters, this limitation substantially deviates from the decisions of the Federal Constitutional Court,[48] the Federal Administrative Court[49] and the civil department of the Federal Court of Justice.[50] This deviation further illustrates the uncertainty and extensive leeway around the publication of court decisions.

Thus, quantitative legal research is limited both by the constraints imposed by anonymization and the relatively low number of published decisions – and, as we introduce below, additionally by costs to access. As a result, studies requiring redacted personal information, e.g. social network analyses of parties or lawyers, are currently impossible. These limitations extend beyond academia to impede economically valuable activities related to legal practice, such as analysis of decision-making processes and patterns of judges or performance analysis of lawyers.

First, the considerable challenges posed by anonymization requirements are described in more detail. As published decisions generally must not contain personal information, the names, addresses and other identifiers of parties and other participants in the proceedings, such as witnesses and even participating legal counsels, are deleted. Given that the overall process is designed to provide access to individual decisions for their legal content upon individual request, and therefore not designed for scale, it is hardly surprising that anonymization is mostly manual. The author has encountered an astonishing array of anonymization methods, ranging from replacing or deleting names and designations in open-text documents to redacting personal information in a PDF file by simply concealing it with a black rectangle positioned over the text in the plane above. Guidelines vary from court to court, sometimes within a single institution. Oftentimes, the anonymizations are handled by clerks but need to be signed off by judges, adding processing time. All in all, the process is very labour-intensive and prone to errors and inconsistencies. Once the decision is anonymized, it is sent to the requestor, but may not be published without explicit consent of the deciding judges.

Even for those judges who are willing or even eager to publish decisions, anonymization still creates a bottleneck they can hardly overcome by themselves, and that quantitative legal research may have to help solve. As an example, the author is currently part of a three-year research project to develop a compliant approach to federated, human-in-the-loop machine learning, to train anonymization models for mass deployment.[51] A significant grant by the German Ministry of Research and Education indicates that the federal government takes the issue seriously. Hopefully, the resulting technical solutions provide a way to considerably speed up the process while reducing oversights and mistakes.

Interestingly, many public court decisions actually contain names of judges, as the proceedings are deemed to fall within these judges' professional, instead of private, spheres. While it is hard to see how the very same court proceedings are deemed private with respect to the parties, this framework at least enables analyses of the judges' decision patterns and related statistics. Alas, any such analysis would require sufficiently large and consistent datasets, and the anonymization of parties severely impedes publication in total, as mentioned above.

The GDPR and the *Bundesdatenschutzgesetz* (Federal Data Protection Act) privilege scientific research insofar as the data processing does not necessarily require consent if the public interest in the research results outweighs the individual's need for privacy. This exception, however, applies only to research institutions. As a result, it does not help the general publication of unredacted decisions, but could provide an option for research projects seeking decisions directly from the relevant court. For projects specifically dealing with decisions by individual courts, this might be the most promising way forward. Still, the administration of said courts might not want to risk a possible privacy breach, and may prefer to anonymize the decisions before handing them over.

In addition to the challenges posed by anonymization, the total amount of published decisions is relatively low. The publication rate is highest for federal courts, but varies significantly even among them. A recent analysis of the decisive bodies of the Federal Court of Justice has found publication rates between 10.0 and 31.2 per cent.[52] Given the much higher case volume and more fragmented structure of regional, district and local courts, their rates are likely lower, but at present no comprehensive study exists.[53] For perspective: there are fewer publicly available decisions for the entire history of Germany than were decided in ordinary law matters in 2017, 2018 or 2019.[54]

As observed above, these low numbers are not solely caused by data protection requirements. Many courts and judges might not see the benefit of publishing individual judgments, and, therefore, refuse to make them public. While some may fear transparency and accountability, most probably assess their work purely on the merits of individual legal analysis. As a result, decisions which a judge deems uniform or standard are classified as unworthy of publication. Here, quantitative legal scholars can raise awareness by explaining the importance of large amounts of data for their work, for example, in detecting the emergence of complexity.

Requesting decisions directly from the courts introduces a new obstacle: cost. Fees for a single decision requested directly from the respective court easily reach up to EUR 1.50 per decision for a digital transmission and up to EUR 0.50 per page for a paper copy. Despite the higher cost, most files requested directly from the courts are provided in paper form and either sent by mail or collected in person. Legal research projects are often exempted from these fees in principle,[55] but, again, these exemptions are designed for requests of single or small numbers of decisions. As fulfilment of large volumes requires substantial effort, these requests are likely to trigger associated caps and limitations, holding requestors liable at minimum for actual fulfilment costs.[56] Clearly, as most courts are unable to export their entire case collection, requests for higher volumes, in digital or machine-readable format, can most likely only be obtained from judicial databases.

In the 1980s, federal courts started to address this problem of scale by founding a special-purpose vehicle which traces its origins to a division of the Federal Ministry of Justice, and has today evolved into the private corporation *juris GmbH* (*Juris*).[57] The German state owns a majority, while the rest is owned by the state of Saarland, the Federal Lawyers Association, the Federal Bar Association and publishing industry investors. Juris entered a public contract with the Federal Republic and various states, granting them exclusive access to all federal court decisions.[58] In exchange, Juris provides the technical infrastructure for collecting the decisions and their metadata.[59] Juris currently maintains a database of more than 1.5 million cases, to which it provides access via a subscription-based business model. Current company revenue is more than EUR 51 million, with an impressive annual net profit of over EUR 8 million (16 per cent yield on turnover).[60] The exclusivity of the agreement has been contested in first- and second-instance courts.[61] The parties settled out of court before the Federal Administrative Court could take a final decision,[62] leading to the de facto availability of edited versions of federal case decisions from 2010 on the web (RII)[63] free of charge. Earlier decisions and those of lower-instance courts are not part of the settlement agreement, and therefore not available via the portal.

While Juris has voluntarily granted access to select parts of its collection to researchers in the past, there is currently no standard procedure, and requests are granted on a case-by-case basis. Data delivery may take several months, and interested researchers should plan accordingly. Bigger datasets require distinct terms, including a non-disclosure agreement. Depending on the nature of the research project, this agreement should address different forms of publication, as the standard NDA does not include language regarding cloud-based data processing and archival publication of datasets. Juris has been open to addressing these issues on an individual level in the past, although this requires additional expenditure of time.

The largest private provider of legal information in digital form is the specialist publisher *Verlag C.H. Beck*, whose database, *beck-online*, contains approximately 3.5 million cases, and several hundred thousand pages of statutes and legal literature, as well as commentary from its more than 12,000 exclusive authors. Its case collection is largely independent from Juris and is

acquired through a variety of channels, including its monthly magazines, and an incentive programme for lawyers, who are compensated if they contribute a decision in one of their cases.

In the early 2000s, federal courts started to publish a selection of their decisions on their websites free of charge; some lower-instance courts have since followed.[64] Most of these databases use Juris technology, which is helpful to know, as the data structure is identical to the one used in the Juris proprietary database. As a result, data pipelines developed for the latter can be used for data from the state and lower-instance courts with minor adjustments. The most extensive collection, RII, currently contains approximately 100,000 decisions. However, its dataset is not fully congruent with constituent files at the individual source websites. As an example, RII contains about 52,000 decisions by the Federal Court of Justice (*Bundesgerichtshof*), whose own website provides access to several thousand more. This discrepancy is likely due to different selection criteria, as both collections are explicitly curated; the exact criteria are unfortunately unknown. As a recent research project has shown, the data in these official collections also differs from exports provided directly by individual senates of the court.[65]

In summary, court decisions are poorly organized and only partially available. This is certainly not a desirable state, but individual researchers can do little about it. While a global quantitative legal studies community should place not only access to but also organization of legal information among its core concerns, readers of this chapter likely seek a more immediate and practical solution. As such, the only principled approach currently seems to be to clearly describe the data source and guard against naïve comparison of results among different studies, as they could have used differently curated datasets. More on best and current practices in Germany can be found in the related section below.

Case files

While all codes of procedure (civil, criminal and special) grant access to the entire court files to the parties involved,[66] the details of the scope of this access differ. For example, in criminal procedures, the defendant can access not only court but also administrative files from the prosecutor's office, including even relevant police reports – though some documentation of the actual trial, in particular judges' personal notes, do not have to be released. In civil cases, the parties have full access to the files; however, the judgment's preparatory documents, its draft and other internal communication of the court, cannot be requested. As German civil procedural law only grants very limited discovery rights, the parties rarely exercise these. In administrative court matters, however, individual claimants regularly make use of their right to access court files as these oftentimes contain the administrative file of the decision they are fighting. German law ensures that their rights to access legal information and the administration's interest to keep certain circumstances secret are balanced. For researchers, these rights would only be relevant if they sought information about proceedings they are actually a part of. In the context of big data law, these cases would be rare.

Under certain circumstances, procedural laws grant access to case files to third parties. In criminal matters, this right to access files mostly applies to the victim and is of little use for researchers. In civil matters, a legitimate legal interest is required. Research can constitute a legitimate interest in individual cases, but there is no general rule granting access to case files purely because they are the object of a research project.[67] It might make sense to pursue this claim if the research questions are closely related to a clearly defined set of cases. More general research questions relating to a more diverse set of cases lower chances of gaining

access to individual files. Under administrative court procedure law, there is no way for third-party researchers to access case files.

Finally, only criminal procedure rules provide for access to full case files for legal research.[68] This framework exists to fulfil the requirement of a foundational legal principle under data protection law addressing processing and release of personal information within these files. The requirements are strict, but legitimate research projects should be able to satisfy them. The procedure can be very long; researchers not only have to provide extensive project information, including plans for securing and managing the data, particularly personal information, but also prove that research goals could not be achieved with anonymized data. This provision both grants an actual right to access the information and allows the court or the prosecutor's office to release relevant personal data for this purpose.[69]

In summary, access to case files is even further restricted than access to mere decisions. Given their more extensive nature, judicial files may still be worth the effort. Particularly those projects concerned with argument mining and the interaction between the parties and the court or requiring extensive sociological or demographic data – such as with criminology studies – might find case files a useful resource.

Access to Legislative Data

The legislature produces statutes and material containing draft laws, legislative proposals, parliamentary protocols, reports, recommendations and resolutions generated by different parts of the legislative process. These materials are available in German from the documentation and information system of parliament on its website.[70] The *Bundesrat* (Federal) also publishes a complete collection of its materials on its website, beginning in 2003.[71] All state parliaments follow a similar approach. Their joint initiative *Parlamentsspiegel* provides access to the databases maintained by the individual states[72] (as of the date accessed, 14 of 16 links were working).[73] Additionally, most of the more than 11,000 municipal councils and respective entities make their materials public, too. Unfortunately, there is no centralized resource to access this data, and listing all sources exceeds the scope of this chapter.[74] As an ever-growing number of legal rules in Germany are based on European legislation, European legislative material is becoming increasingly important as a data source.[75] While the data is generally available in a variety of file formats, and despite the existence of explicit data models, the data is typically not available in a machine-readable format, but rather in PDF or Microsoft Word containers – except for European statutes and their materials.[76]

Federal statutes and amendments are made publicly available via publication in the *Bundesgesetzblatt* (Federal Gazette).[77] This repository includes records of changes to or abrogation of an existing law, and the full-text initial announcement of a new law. While there is yet no publicly available tool to compile the current version and/or historical versions of a law from the publications in the Federal Gazette, at least the current versions of all federal laws can be obtained in a machine-readable format from an official website of the Federal Ministry of Justice, "Statutes on the Web" (Gesetze im Internet (GII)).[78] Researchers will want to note: this database does not archive previous versions of these statutes. And, just like court decisions, the statutes are not protected by copyright, as the same exceptions for official texts apply – as laid out above, this exception currently applies to databases under German law, too, possibly in violation of European law. Finally, the technical infrastructure supports a download of the entire database, with regular updates available via RSS feed.

In addition, numerous third parties provide access to laws and statutes;[79] most importantly, Juris stores all federal laws since the 1950s. Though not official versions, they are compiled by the Federal Office of Justice for the official federal law database used by parliament and federal ministries, using Juris technology.[80]

Similarly, state statutes are published in that state's official gazette. All states run websites containing state legislation, similar to GII and running on the same infrastructure.[81] As a result, their data follows the Juris models and can be processed by pipelines developed for it.

Legislative texts are among the easiest of the available legal corpora to obtain. Depending on project design and focus, they may be valuable sources for quantitative legal research, e.g. for understanding how the volume of statutes develops over time.[82]

Access to Administrative Data

The third important source of data is federal, state and municipal administrative proceedings. This data was traditionally regarded as an official secret, and access was rarely granted. As noted above, only those who were an immediate part of the proceedings received relevant information. As government administration became more transparent over time, this restrictive approach developed into a general right to government information for everyone, and, finally, into a legal obligation for government entities to provide data, where possible, even in machine-readable form. This section describes those developments and their implications for researchers.

Originally, access to information was only provided under explicit rules for hearings. As a general principle, everyone who might suffer a detriment under a public act must receive a hearing, and relevant information to prepare their case. In specific areas such as environmental law or zoning, the administration must provide extensive materials, including general goals, public findings relevant to the matter, and considered courses of action. Large-scale projects and matters of procurement fall under specific disclosure and public communications rules applicable throughout the process. As a result, researchers – typically not directly concerned – can claim access to this data. If the public authority fails to comply with requests for information, various administrative law acts provide legal means to either annul the act or stop its implementation with an injunction by an administrative court. As a result, authorities are likely to comply with requests.

This framework, however, is primarily designed for affected individuals; while researchers might attempt to gain access via this vehicle, as discussed earlier, the general design of individual access rights does not lend itself to quantitative legal research endeavours. For example, many authorities either supply information only on paper, and/or require that it must be retrieved in person, often lacking the infrastructure to easily transfer larger data volumes.

As indicated above, since the late 1990s, individual states have abandoned the idea of official secrets and introduced freedom of information laws, creating a general right to administrative information unless specific exceptions apply, such as for national security, or diplomatic or third-party rights. Copyright is typically not an issue,[83] but privacy and the resulting anonymization of personal information are – similar to judicial data.

The Federal Government followed suit in 2006, introducing the *Informationsfreiheitsgesetz* (IFG) (Federal Freedom of Information Act) to provide such a right on a federal level. As of today, 13 of the 16 German states have comparable legislation in place. The remaining three make a limited amount of data available upon request under different legal rule sets, e.g. data

protection laws.[84] The above-mentioned case databases contain only about 1,200 court cases dating from 2006 dealing with these freedom of information laws. Given that Germans filed nearly 70,000 requests up until 2017 on the federal level alone,[85] the authorities appear to generally comply, or at least provide acceptable reasons for dismissals. A more detailed analysis presents an excellent opportunity for a quantitative legal research project. The degree to which public bodies must release data proactively varies widely, with states such as Hamburg, Rhineland-Palatinate and Thuringia requiring all public authorities to automatically make all their data available on the web.

This is significant for researchers, as requests for access to administrative information typically require a fee of EUR 15 to 500 per case. In contrast to the legal framework governing judicial data explained above, applicable fee schedules do not allow general exceptions for research purposes. In practice, administrative authorities are nonetheless often willing to provide larger quantities of data for research without a significant charge. But then again, even though operations are designed to accommodate a high volume of requests, they are not built to transfer large amounts of data. As a result, even when authorities are willing to provide access, it may take significant time to compile larger datasets, and these datasets, in turn, may require extensive pre-processing. Fortunately, even though there are generally restrictions on the use of data acquired under the IFG,[86] there are no restrictions that would limit the use of the data for research purposes.

These operational challenges are increasingly being addressed for some categories of administrative information. For instance, electronically stored and formalized raw data (excluding text) and metadata from IT systems used by public authorities fall under the E-Government-Gesetz (EgovG) (E-Government Act).[87] All 16 states have enacted similar laws.[88] Essentially, these require authorities to provide data in a machine-readable format to the public. This is a particularly consequential development as it drastically reduces the need for pre-processing, rendering unnecessary the costly, labour-intensive digitization of paper copies and information extraction from PDFs or image formats. Further, changing the structure for analytical purposes becomes much easier as the data already follows a standard model.

Although many legal research projects may require information far exceeding the scope of this law, its mere introduction forces public authorities to build infrastructure which in future could be used for providing data under other legal rule sets, perhaps even judicial data. The IT Planning Council[89] manages the implementation of these laws at different levels and coordinates efforts between the federal government, states and municipalities. Its open government data platform currently provides over 70,000 individual datasets.[90] The near doubling of this amount in the past 18 months reflects extensive effort by administrative entities. The platform's adoption is not equally distributed among levels of administration: interestingly, it is the lowest level, municipalities – especially those with the highest population – that seem to lead the effort.[91] At any rate, government must regularly monitor progress, and currently seems committed to improving access to administrative metadata.

As this section has laid out, legal data is neither completely impossible nor particularly easy to come by at present. Quantitative legal scholars therefore have a strong interest in influencing the policy debate regarding the legal framework that controls access to this information. Their most promising approach is to demonstrate the usefulness of this data to the academic community and to the public, and so it is of vital importance that their publications are as insightful and robust as possible. The next section suggests best practices for achieving this outcome.

BEST PRACTICES

At least in the wider academic legal community in Germany, quantitative legal studies face an uphill battle. While some may find this frustrating,[92] it should actually motivate interested scholars to work thoroughly and comprehensively, and keep their claims modest. One can only convince critics by surpassing their expectations and by addressing their concerns head on. The following proposes a means to achieve this goal.

As mentioned, the field of quantitative legal research in Germany is in an early stage. With incomplete information, hypothesizing can be rather difficult. As most traditional legal research is concerned with how the law and the world should be, forming a testable hypothesis can often be impossible at the beginning. Exploratory data analysis as a mindset first and a toolset second provides a solution.[93] This is an important distinction within more established fields of quantitative analysis such as law and economics, in which theoretical foundations are more advanced and selected datasets are better understood.

Picking up from the previous section's conclusion, the next section describes ideal qualities of legal datasets, suggests analytical approaches and finally recommends the most productive audiences to address to meet researchers' aims.

Datasets

The dataset is the foundation for all following steps and should be of the highest possible quality. It is the easiest starting point for critics, as finding flaws in the dataset often requires neither deeper knowledge of the methods nor extensive analysis of the results.

Accessible publication of datasets is paramount. Researchers should work hard to overcome obstacles, real or perceived, to publishing data, and associated methodologies or protocols, in a manner that is as accessible and reproducible as possible.[94] As long-term availability and immutability are crucial, and with professional and reliable data archiving available for free,[95] archiving the dataset should be standard practice. Researchers should resolve to invest their time and resourcefulness to overcome obstacles that often prevent publication, such as non-disclosure agreements. As a recent example, the author encountered a situation in which a valuable dataset was available only under a strict non-disclosure agreement. Instead of simply accepting this and writing a data availability statement for the paper, the research team engaged in a productive discussion with the provider. It turned out that confidentiality of only the documents' textual content was desired, while their structure could be made available. Since the paper mainly focused on the latter, a comprehensive dataset could be released once the text was removed.

Ideally, datasets should be as complete as possible. For example, if the decisions of a court are studied, the dataset should include all decisions by that court. Far too frequently, partial samples are taken, as these are more convenient to procure, raising an elemental risk of availability heuristics. As discussed extensively above, many providers are not equipped to fulfil requests for large datasets, prompting the emergence of shadow providers. The author and other quantitative legal scholars have been approached on more than one occasion by unofficial or unauthorized providers of legal datasets. While this option may be tempting, it creates a serious problem. If the origin of the data is doubtful, so is its composition.

If no complete dataset can be compiled, the reasons for a partial analysis need to be stated openly, and the resulting limitations for analysis addressed explicitly. Again, too often

researchers devise after-the-fact justifications for incomplete data. While an understandable and expedient response, it conceals the actual problem of a lack of available data. Unless researchers openly address this issue, they will miss the opportunity to apply the political pressure needed to release the required information.

The quality of datasets involves more than completeness, even though this might be the most straightforward indicator of quality. Ideally, datasets should also be as rich and reliable as possible. This requires discipline at all stages of data handling – processing, transformation, dimensionality reduction, etc. – as the original data may contain information deemed unnecessary for the analysis at hand. Often after the project's conclusion, the original data fails to be preserved in its entirety, making it substantially harder for others to reproduce and/or build on top of these results. In addition, as familiarizing oneself with a new dataset requires substantial amounts of time, researchers are encouraged to take their analyses as far as possible. Even if the amount of newly discovered knowledge required for publication is already reached, researchers should be mindful that it will never be easier to produce more insights than in the current project, and make a best practice out of preserving datasets beyond the requirements for the project at hand.

Efforts to create datasets should also be community-based. As outlined below, the procurement and cleaning of these datasets requires extensive pipelines and a lot of time; to avoid inefficiencies, these compilations and pipelines should be reused whenever possible. As best practices or widely accepted document standards do not exist yet, the pipelines are prone to programming errors, many of which can be avoided when the code is constantly tested, and these tests made public with the dataset. Some issues, however, require human intervention and can only be resolved when researchers in the field use each other's pre-processing code. This is easier said than done, as it often seems faster to build a new, rather than adapt an existing, solution. In the medium term, joint development and reuse will lead to standard packages, whose accuracy can be increasingly trusted over time, though hopefully not unconditionally.[96]

Theoretically, German courts and administrative bodies within the judiciary have adopted XJustiz – a fully defined, document type definition for structured legal data and their exchange between public and private parties. It currently contains 22 modules of XML schemas for legal procedures. The official specifications[97] are freely available and maintained by the *Bund-Länder-Kommission für Informationstechnik in der Justiz* (Joint Federal and States Commission for Information Technology in the Judiciary). In practice, these standards are not widely adopted outside of electronic legal communication and public registers. Juris uses its own document type definition, which is available upon request. As Juris provides technical infrastructure for many court websites, and the only official digital collections of court decisions[98] and statutes[99] available to the public, its specifications are of major importance. As laid out above, pipelines which take Juris particularities into account have a high chance of being useful for various sources of legal data. That said, where a specification deviates from accepted standard practices and paradigms, it should not be followed, as it is likely to cause problems for other researchers reusing the code.

Inconsistencies such as these provide further reason to extensively document datasets. Where possible, datasets should be introduced in a separate publication from their analysis to provide sufficient space for their description. Research publications increasingly accept datasets as a form of publication equivalent to other contributions. As the composition of these sets is work-intensive and requires painstaking accuracy, it should be regarded as a valuable

contribution to the community in its own right. At times, such efforts can transcend data compilation to become toolsets and resources that unlock other pools of information.[100]

Toolsets

Keeping the goal of rich, well-documented and widely available open-source datasets in mind, this section lays out thoughts on a suitable organizing or methodological approach, advocating for a combination of methods from data mining, text analysis and network science.

First, there are no right or wrong methods for quantitative legal studies; they depend heavily on the dataset and research question at hand. Applications of existing tools to new problems and applications of new tools to old problems can both lead to meaningful insights. There are, however, general considerations.

As quantitative legal research is by nature interdisciplinary, as observed earlier, ideally this quality would be reflected in the individual researcher as well; a productive team for this type of research comprises individuals with education in or familiarity with both a quantitative field, such as statistics or computer science, and law. Teams of individuals thus doubly qualified benefit from several advantages. For one, such individuals help to minimize limitations in the choice of methods, as those foreign to a field often do not grasp the full depth of its available methods. For another, projects progress significantly faster, as the group is likely to suffer fewer misunderstandings generated by members rooted in monolithic scientific perspectives. Yet another advantage is that these types of interdisciplinary teams are more likely to ensure that preconditions for their applied methods are met and that their results are not only valid, but also understandable. Whatever the team composition, researchers should understand the legal phenomena examined well enough to make informed, expert judgement calls, for example, determining the set of hyperparameters for machine learning models.

Many research projects require the extraction of knowledge from large quantities of data which is not specifically structured for the endeavour. This is a task generally performed using data mining methods. As best practices in data mining fill whole books of exceptional quality,[101] there is no need to go into details in this chapter. Oftentimes, the data source is a corpus of legal documents containing text. To enable its analysis, methods from text mining and/or analysis, which again constitute entire disciplines with excellent literature,[102] can be deployed. Also, if the actual content is of interest for the research question, methods from natural language processing are useful.[103] These approaches, however, often introduce heuristics or approximations rather than definitive solutions, as language is messy, which can be mitigated or circumvented where legal documents deliberately contain a high degree of structure, permitting a more straightforward process of analysis. All these methods can be applied to make sense of legal texts and metadata from legal documents.

The choice of methods could also be approached from a broader perspective, leveraging the societal function of law. Whether it attributes goods or guides behaviour, the law seeks to regulate relationships between humans; thus, data pertaining to social relationships and networks captured in legal data sources is of particular interest. Methods from network science are especially useful to analyse social relationships, and can be applied to a great variety of legal problems.[104]

Finally, results must be communicated to a relevant audience. While the characteristics of that audience are laid out in the following section, the importance of data visualization to that communication cannot be overstated. While a topic in its own right, covered by outstanding

scholars,[105] data visualization is nonetheless often treated as a mere afterthought in scientific publication. In particular, scholars from natural science disciplines are mostly focused on the accuracy of their visualizations. Accuracy is mandatory, but clearly communicating the results, especially to readers from law, social sciences or other disciplines, who might be less experienced in interpreting visual data, is of near equal importance. In many cases, readers who lack, for example, the mathematical training to fully digest the content of a quantitative legal study will judge it by the figures included. It is therefore worthwhile to invest time in figure optimization, as clear figures substantially increase the persuasive power of a paper.

The tools from the fields mentioned above are chosen based on their suitability for the creation of and use with the datasets described. Ultimately, their selection for any given project will be driven largely by the questions to which they can be applied, namely those questions which at the time seem both particularly interesting and urgent. For most projects, choosing and configuring methods is a core part of the work; entirely different tools than those discussed in this chapter might be suitable if they fit the problem, for example, agent-based modelling[106] for the evaluation of regulatory concepts. As the approaches discussed in this section have produced robust results for quantitative legal studies in Germany, one may expect readers will find them useful as well.

Audiences

As research is always produced for a particular audience, this section briefly describes key audiences for quantitative legal research, and suggests which to prioritize.

As mentioned throughout, the scientific community for quantitative legal studies in Germany is nascent, currently comprising only a few dozen scholars, regularly publishing papers in specialized national or international journals. The international community is bigger, but likely still fits into a larger conference room. As this group is most likely to provide peer review and incorporate findings into their own individual work, it is of utmost importance to the researcher. Generally, this audience is well versed in the relevant fields, and will see the value of quantitative legal research without additional persuasion. As most of them are academics, however, their reach into practice is limited. There is little this group alone can do to ameliorate the arduous aspects of the legal framework or to apply valuable findings in practical contexts.

This situation is vastly different for the second audience – policymakers. While their backgrounds may differ, many of them, especially within administrative institutions such as ministries and agencies, have a legal background. Depending on their subject-matter expertise, some may actually have a natural science background. Among members of parliament, the final policymakers, lawyers are by far the single biggest profession represented, with currently 190 members, more than a quarter of all delegates.[107] Their understanding of the law and regulatory system is influenced by their traditional, doctrinal education. Yet, they are the ones who could most effectively apply recommendations from quantitative legal studies, and substantially advance the future of the field through legislation that facilitates this kind of work. To convince this audience, and to equip them to pass better and more efficient legislation, quantitative legal scholars must relate their findings to the problems of this group, and, in doing so, speak its language. One enormous strength of this field of research is its robustness and precision in real-world scenarios. To leverage it, this audience should be a key target.

An equally important and similarly trained audience are legal practitioners. While they cannot change the rules quite so directly, their day-to-day application of the law in courts, and inside legal departments and law firms, strongly influences the public perception of the law. They can also detect legal system deficiencies and can therefore point to promising areas for quantitative legal studies. Research must be relevant for practice, which motivates practitioners and scholars to engage, an interchange that is especially important for quantitative scholars, who depend on empirical data and have to understand the relationship of the data to the practice. Ideally, both sides can profit from the right type of research, and the inspiration and learning can work both ways.

With the fourth audience, we return to legal scholars, albeit more traditional, normative researchers than the first audience discussed above. So far, many of the approaches and best practices in this chapter concern technical questions, potentially creating the impression that quantitative legal studies are mainly driven by the quality of the data science, and less so by substantive legal considerations. This is fundamentally wrong. Both the success of this discipline and the impact of its results heavily depend on experts of doctrinal law. They are an integral part of every research team, as they can interpret results and locate them in the relevant discussion of a particular legal subject area. They are of immense value, as they ensure that the assumptions in the models are accurate from their perspective. Given their expertise, they are also most qualified and most capable to highlight the shortcomings of quantitative legal research where appropriate. Therefore, good quantitative research should do its utmost to be relevant in doctrinal discussions and convince the more sceptical scholars through its relevance to their doctrinal field. The fact that there are examples of such research featured in prestigious doctrinal publications[108] indicates a fair chance exists of swaying this audience, and underscores why it is worth trying.

SUMMARY

This chapter has presented the current state of quantitative legal studies in Germany, providing an overview of the scholarly traditions which form the basis for future development of this discipline. For those ready to start quantitative research projects, it has provided practical lessons learned, and an overview of the regulatory framework for access to legal data. It has suggested best practices for handling datasets and working as a community, and it has proposed analytical tools and possible audiences. Finally, it argues for a collaborative approach to working with normative legal scholars. Based on the above discussion of the history of data-focused legal research, we should avoid going down the path of legal sociology, which at present has become a fringe discipline whose findings are adopted far less into the general legal discussion than they merit. In contrast, the field of law and economics could provide a useful blueprint for the interaction with normative legal scholars; with this approach, quantitative legal studies can become an integral part of the academic legal tradition. Researchers can change the legal system – and, indeed, society – for the better by using its diverse tools and methods to unravel these complex adaptive systems. While the discipline is still developing in Germany, there is much opportunity to shape this field.

NOTES

1. A practice-oriented overview in English is presented by Maria C. Caldarola & Joachaim Schrey, Big Data und Recht (2020); an academic overview in German is presented by Wolfgang Hoffmann-Riem, Big Data – Regulative Herausforderungen (2018).
2. Jonathan S. Ward & Adam Barker, Undefined By Data: A Survey of Big Data Definitions (2013), https://arxiv.org/abs/1309.5821.
3. J. B. Ruhl, *Law's Complexity – A Primer*, 24 Ga. St. U. L. Rev. (2012).
4. J. B. Ruhl & Daniel M. Katz, *Measuring, Monitoring, and Managing Legal Complexity*, 101 Iowa L. Rev. 191 (2015).
5. Daniel M. Katz & Michael J. Bommarito, *Measuring the Complexity of the Law: The United States Code*, 22 Artificial Intell. & L. 337 (2014).
6. J. B. Ruhl et al., *Harnessing Legal Complexity*, 355 Science 1377 (2017).
7. Thomas Raiser, Grundlagen der Rechtssoziologie 15 (6th ed. 2013).
8. Deborah R. Hensler & Matthew A. Gasperetti, *The Role of Empirical Legal Studies in Legal Scholarship, Legal Education and Policy-Making: A U.S. Perspective*, in Rethinking Legal Scholarship: A Transatlantic Dialogue (Rob van Gestel, Hans-W. Micklitz & Edward L. Rubin, eds., 2017).
9. Raiser, *supra* note 7, at 13.
10. Klaus F. Röhl, Rechtssoziologie (2007), https://www.ruhr-uni-bochum.de/rsozinfo/pdf/Roehl-RS-10-Nachtrag.pdf.
11. The German Journal of Law and Society, https://www.degruyter.com/view/j/zfrs (last visited June 5, 2020).
12. German Association for Law and Society, https://www.rechtssoziologie.info/ (last visited June 5, 2020).
13. Michael Wrase, *Rechtssoziologie und Law and Society – Die Deutsche Rechtssoziologie Zwischen Krise und Neuaufbruch*, 27 Zeitschrift für Rechtssoziologie 286, 308 (2006).
14. Eva Kocher, *Rechtssoziologie: Das Recht der Gesellschaft und die Gesellschaft des Rechts*, 8 Rechtswissenschaft [RW] 153, 176 (2017).
15. Andreas M. Fleckner, *Review of Hanjo Hamann: Evidenzbasierte Jurisprudenz (Grundlagen der Rechtswissenschaft, Vol. 23)*, 82 Rabels Zeitschrift für ausländisches und internationales Privatrecht [RabelsZ] 471, 471 n.1.
16. Niels Petersen, *Braucht die Rechtswissenschaft eine Empirische Wende?*, 49 Der Staat 435, 455 (2010).
17. Ino Augsberg, *Von Einem Neuerdings Erhobenen Empiristischen Ton in der Rechtswissenschaft*, 51 Der Staat 117, 124–25.
18. Hanjo Hamann, Evidenzbasierte Jurisprudenz (2014), one chapter of which is available in English at https://lawcat.berkeley.edu/record/1126108?ln=en (last visited June 5, 2020).
19. Fleckner, *supra* note 15, at 471–78.
20. Empirical legal research, however, is included in the official designation of at least two professorships, *see* Prof. Dr. Niels Petersen, Chair for Public Law, International European Law and Empirical Legal Research, WWU Münster, https://www.jura.uni-muenster.de/de/institute/lehrstuhl-fuer-oeffentliches-recht-voelker-und-europarecht-sowie-empirische-rechtsforschung (last visited on June 5, 2020) and Prof. Dr. iur. Emanuel V. Towfigh, Chair of Public Law, Empirical Legal Research and Law & Economics, EBS Universität, https://www.ebs.edu/de/organ/lehrstuhl-fuer-oeffentliches-recht-empirische-rechtsforschung-und-rechtsoekonomik (last visited on June 5, 2020).
21. *See* Hanjo Hamann & Friedemann Vogel, *Evidence-Based Jurisprudence Meets Legal Linguistics – Unlikely Blends Made in Germany*, 2017 BYU L. Rev. 1473 (2018).
22. Friedemann Vogel, *Legal Linguistics in Germany. History, Working Groups, Concepts*, in Legal Linguistics Beyond Borders: Language and Law in a World of Media, Globalisation and Social Conflicts 103, 103–4 (Friedemann Vogel ed., 2019).
23. Friedemann Vogel, *Rechtslinguistik: Bestimmung einer Fachrichtung*, in Handbuch Sprache im Recht (Ekkehard Felder & Friedemann Vogel eds., 2017).

24. *See* Computer Assisted Legal Linguistics, https://www.cal2.eu/index.php (last visited June 5, 2020) for examples , *see* Friedemann Vogel et al., *Computer-Assisted Legal Linguistics: Corpus Analysis as a New Tool for Legal Studies*, 43 Law & Social Inquiry 1340 (2018) for an introduction.
25. Corinna Coupette & Andreas M. Fleckner, *Quantitative Rechtswissenschaft*, 73 Juristenzeitung [JZ] 379, 381 (2018).
26. Corinna Coupette, Juristische Netzwerkforschung – Modellierung, Quantifizierung und Visualisierung relationaler Daten im Recht (2019).
27. For an example *see* Eugen Ruppert et al., *LawStats – Large-Scale German Court Decision Evaluation Using Web Service Classifiers*, in Machine Learning and Knowledge Extraction (Andreas Holzinger et al. eds., 2018).
28. Examples include Peter Agstner, *Shareholder Conflicts in Close Corporations between Theory and Practice: Evidence from Italian Private Limited Liability Companies*, Eur. Bus. Org. L. Rev. (2019); Christian Arnold et al., *Scaling Lower Court Decisions*, https://www.sowi.uni-mannheim.de/media/Lehrstuehle/sowi/Gschwend/Articel/JELS_arnoldetal.pdf (last visited June 5, 2020); Masha Medvedeva et al., *Using Machine Learning to Predict Decisions of the European Court of Human Rights*, Artificial Intell. & L. (2019).
29. Herbert Fiedler et al., Untersuchungen zur Formalisierung im Recht als Beitrag zur Grundlagenforschung juristischer Datenverarbeitung (UFORED) (1984).
30. Thomas Hoeren & Michael Bohne, *Von der Mathematischen Strukturtheorie zur Integrationsdisziplin*, in Informatik in Recht und Verwaltung: Gestern - Heute - Morgen 23–36 (Roland Traunmüller & Maria A. Wimmer eds., 2009).
31. *See* https://www.iri.uni-hannover.de/de/ (last visited June 5, 2020).
32. *See Prof. Dr. Thomas Hoeren*, ITM, https://www.itm.nrw/organisation/prof-dr-thomas-hoeren/ (last visited June 5, 2020).
33. *See* Institut Für Rechtsinformatik Universitat Des Saarlandes, https://rechtsinformatik.saarland/en/ (last visited June 5, 2020).
34. *See* Universitat Hamburg, https://www.inf.uni-hamburg.de/en/inst/ab/lt/home.html (last visited June 5, 2020).
35. *See Prof. Dr. Florian Matthes*, Sebis Public Website (Dec. 18, 2019), https://wwwmatthes.in.tum.de/pages/88bkmvw6y7gx/Prof.-Dr.-Florian-Matthes (last visited June 5, 2020).
36. *See Prof. Dr. Michael Gertz*, Database Systems Research Group, https://dbs.ifi.uni-heidelberg.de/team/gertz/ (last visited June 5, 2020).
37. *See Doctoral Research Group "Digital Law"*, Universitat Heidelberg, https://www.jura.uni-heidelberg.de/digitales_recht/index_e.html (last visited June 5, 2020).
38. *See* DGRI, https://www.dgri.de/ (last visited June 5, 2020).
39. *See Fachgruppe Rechtsinformatik*, FB-RVI, https://fb-rvi.gi.de/fachgruppen/rechtsinformatik (last visited June 5, 2020).
40. *See IRIS - Internationales Rechtsinformatik Symposion*, Universitat Wien, https://rechtsinformatik.univie.ac.at/iris/ (last visited June 5, 2020); *Conferences*, Jurix, http://jurix.nl/conferences/ (last visited June 5, 2020).
41. Examples include Coupette & Fleckner, *supra* note 24; Benjamin G. Engst et al., *Zum Einfluss der Parteinähe auf das Abstimmungsverhalten der Bundesverfassungsrichter – eine quantitative Untersuchung*, 72 JuristenZeitung [JZ], 816 (2017); Hanjo Hamann & Leonard Hoeft, *Die Empirische Herangehensweise im Zivilrecht*, 217 Archiv für civilistische Praxis [AcP], 311 (2017); Alexander Morell, *Die Rolle von Tatsachen bei der Bestimmung von "Obliegenheiten" im Sinne von § 254 BGB am Beispiel des Fahrradhelms*, 214 Archiv für civilistische Praxis [AcP], 387 (2014); Niels Petersen & Konstantin Chatziathanasiou, *Empirische Verfassungsrechtswissenschaft: Zu Möglichkeiten und Grenzen quantitativer Verfassungsvergleichung und Richterforschung*, 144 Archiv des öffentlichen Rechts [AöR], 501 (2019).
42. Judicial data includes other documents types, too. An example would be court organizational charts, a third-party source for which would be https://richter-im-internet.de (last visited on June 5, 2020).
43. Bundesgerichtshof [BGH] [Federal Court of Justice] Nov. 21, 1991, 116 Entscheidungen des Bundesgerichtshofs in Zivilsachen [BGHZ] 136.
44. BGH, Sep. 28 2006, I ZR 261/03, http://juris.bundesgerichtshof.de/cgi-bin/rechtsprechung/document.py?Gericht=bgh&Art=en&nr=39461.

45. Bundesverfassungsgericht [BVerfG] [Federal Constitutional Court], Dec. 15, 1983, Entscheidungen des Bundesverfassungsgerichts [BVerfGE] 65, 1.
46. European Union Directive 2016/679, 2016 O.J. (L 119) 1.
47. BGH, June 20, 2018, Entscheidungen des Bundesgerichtshofs in Strafsachen [BGHSt] 63, 156.
48. BVerfG, 1 BvR 857/15, Sep. 14, 2015, https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/DE/2015/09/rk20150914_1bvr085715.html.
49. Federal Administrative Court [Bundesverwaltungsgericht], Feb. 26, 1997, Entscheidungen des Bundesverwaltungsgerichts [BVerwGE] 104, 105.
50. BGH, Apr. 5, 2017, 2017 Neue juristische Wochenschrift [NJW] 1819.
51. https://hilano.de (last visited June 5, 2020).
52. Coupette & Fleckner, *supra* note 24, at 381.
53. The most recent, partial study is from 2006: Wolfgang Kuntz, *Quantität Gerichtlicher Entscheidungen als Qualitätskriterium juristischer Datenbanken*, JurPC Web-Dok, 12/2006 (2006).
54. Janis Beckedorf et al., *Analyzing High Volumes of German Court Decisions in an Interdisciplinary Class of Law and Computer Science Students – LSIA*, in Computational Legal Studies: The Promise and Challenge of Data-Driven Legal Research (Ryan Whalen ed., 2020).
55. *See* Section 11 Landesjustizkostengesetz Hamburg [LJKG-HH] [Hamburg Legal Cost Act], http://www.landesrecht-hamburg.de/jportal/portal/page/bshaprod.psml;doc.id=jlr-JKostGHArahmen (last visited June 5, 2020).
56. *See* Section 22 Justizverwaltungskostengesetz [JVKostG] [Federal Legal Cost Act], https://www.gesetze-im-internet.de/jvkostg/BJNR265500013.html (last visited June 5, 2020).
57. GmbH is short for Gesellschaft mit beschränkter Haftung [German limited liability company].
58. Thomas Fuchs, *Die Weiterverwendung der Gemeinfreien Rechtsdatenbank "Juris"*, https://delegibus.com/2011,2.pdf (last visited June 5, 2020).
59. Verwaltungsgericht Köln [Cologne administrative trial court] Sep. 12, 2002, 6 K 4342/99, http://www.justiz.nrw.de/nrwe/ovgs/vg_koeln/j2002/6_K_4342_99urteil20020912.html (last visited June 5, 2020).
60. Jahresabschluss zum 31. Dezember 2018 und Lagebericht [annual report 2018], https://www.juris.de/jportal/cms/juris/media/pdf/unternehmen/juris_GmbH_Geschaeftsbericht_zum_31122018.pdf (last visited June 5, 2020).
61. Verwaltungsgerichtshof Baden-Würtemberg [VGH] [Higher Administrative Court] May 7, 2013, 10 S 281/12, http://lrbw.juris.de/cgi-bin/laender_rechtsprechung/document.py?Gericht=bw&nr=16959 (last visited June 5, 2020).
62. BVerwG, June 15, 2015, 7 C 13.13 (not reported).
63. https://rechtsprechung-im-internet.de (last visited June 5, 2020).
64. Since 2016 consolidated as Rechtsprechung im Internet, https://www.rechtsprechung-im-internet.de/ (last visited June 5, 2020); For state level *see Rechtsprechung*, Justizportal des Bundes und der Länder, https://justiz.de/onlinedienste/rechtsprechung/index.php (last visited June 5, 2020).
65. Coupette & Fleckner, *supra* note 24, at 38.
66. *See* Sections 147, 406e Strafgesetzbuch [StGB] [Code of Criminal Procedure], https://www.gesetze-im-internet.de/stgb/; Section 299 Zivilprozessordnung [ZPO] [Code of Civil Procedure], https://www.gesetze-im-internet.de/zpo/; Section 100 Verwaltungsgerichtsordnung [VwGO] [Rules of the Administrative Courts], https://www.gesetze-im-internet.de/vwgo/ (last visited June 5, 2020).
67. BVerwG, Oct. 9 1985, 1986 NJW 1277.
68. *See* § 476 German Code of Criminal Procedure [Strafprozessordnung].
69. *See* 63 BGHSt 156.
70. *See Willkommen in DIP*, DIP, http://dipbt.bundestag.de/dip21.web/bt (last visited June 5, 2020).
71. *See Parlamentsdokumente*, Bundesrat, https://www.bundesrat.de/DE/dokumente/dokumente-node.html (last visited June 5, 2020).
72. *See Links zu den Parlamentsdokumentationen*, Parlamentsspiegel, https://www.parlamentsspiegel.de (last visited June 5, 2020).

73. The materials for Saarland and Saxonia are available under *Dokumente*, Landtag Des Saarlandes, https://www.landtag-saar.de/dokumente (last visited June 5, 2020) and Säschsischer Landtag, http://edas.landtag.sachsen.de/ (last visited June 5, 2020).
74. The best starting point for research would be the joint database of the Federal Statistical Office and its state counterparts; *see Gemeindeverzeichnis-Informationssystem GV-ISys*, Statistisches Bundesamt (Destatis), https://www.destatis.de/DE/Themen/Laender-Regionen/Regionales/Gemeindeverzeichnis/_inhalt.html (last visited June 5, 2020).
75. *See* EUR-Lex.europa.eu, https://eur-lex.europa.eu/ (last visited June 5, 2020).
76. *See infra* section on datasets for details.
77. *See* Bundesgesetzblatt, https://www.bgbl.de/ (last visited June 5, 2020) for the official data in pdf containers. An alternative, though not official, is built under https://offenegesetze.de (last visited June 5, 2020).
78. *See* Gesetze im Internet, https://www.gesetze-im-internet.de/ (last visited June 5, 2020), for files in html, pdf, epub and xml format.
79. *See*, for example, privately run Dejure, https://dejure.org/ (last visited June 5, 2020), or Buzer, https://buzer.de (last visited June 5, 2020), which both provide federal laws and side-by-side diff views, and publicly supported Open Legal Data, https://de.openlegaldata.io (last visited June 5, 2020), which provides an API following the REST paradigm, https://lexetius.com (last visited June 5, 2020).
80. *See Normendokumentation*, Bundesamt für Justiz, https://www.bundesjustizamt.de/DE/Themen/Gerichte_Behoerden/Normendokumentation/Normendokumentation_node.html (last visited June 5, 2020).
81. *See Bundes-und Landesrecht*, Justizportal des Bundes und der Länder. https://justiz.de/onlinedienste/bundesundlandesrecht/index.php (last visited June 5, 2020).
82. Daniel M. Katz et al., *Complex Societies and the Growth of the Law* 10 Sci. Rep. 18737 (2020), available at https://doi.org/10.1038/s41598-020-73623-x.
83. BGH, Apr. 30, 2020, I ZR 139/15.
84. Those states are currently Bavaria, Lower Saxony and Saxony.
85. Github, https://github.com/okfde/ifg-statistik (last visited June 5, 2020).
86. *Informationsweiterverwendungsgesetz* [Federal Act on the Re-Use of Public Sector Information], https://www.gesetze-im-internet.de/iwg/ (last visited June 5, 2020).
87. *Gesetz zur Förderung der elektronischen Verwaltung* [EgovG] [Act to promote electronic government], https://www.gesetze-im-internet.de/englisch_egovg/index.html (last visited June 5, 2020).
88. Wissenschaftlicher Dienst des Bundestages [WD] [Research Services of the German Bundestag], June 28, 2019, WD 3 - 3000 - 134/19, https://www.bundestag.de/resource/blob/655082/32a17c3834d5c5c5d6f5a7232f0491c0/WD-3-134-19-pdf-data.pdf.
89. IT Planning Council, https://www.it-planungsrat.de/EN/home/home_node.html (last visited June 5, 2020).
90. GovData, https://www.govdata.de/ (last visited June 5, 2020).
91. *See* Jürgen Stember et al., *Studie zum E-Government-Gesetz*, http://egov.hs-harz.de/index.php/download/category/2-publikationen?download=32:studie-zum-e-government-gesetz (last visited Apr. 30, 2020).
92. E.g., Ulrich Zachert, *Der Arbeitsrechtsrechtsdiskurs und die Rechtsempirie – Ein schwieriges Verhältnis*, WSI Mitteilungen 421 (2007).
93. John W. Tukey, *We Need Both Exploratory and Confirmatory*, 34 Am. Statistician 23 (1980).
94. *See* Fair Principles, https://www.go-fair.org/fair-principles/ (last visited June 5, 2020).
95. E.g., Dryad, https://datadryad.org (last visited June 5, 2020); FigShare, https://figshare.com (last visited June 5, 2020); Zenodo, https://zenodo.org (last visited June 5, 2020) or intR2dok – a special purpose open access repository for publications by the German Research Foundation, https://intr2dok.vifa-recht.de (last visited June 5, 2020).
96. For an example of fallibility of even established packages see Jayanti Bhandari Neupane et al., *Characterization of Leptazolines A–D, Polar Oxazolines from the Cyanobacterium Leptolyngbya sp., Reveals a Glitch with the "Willoughby–Hoye" Scripts for Calculating NMR Chemical Shifts*, 21 Organic Letters 8449 (2019).
97. *See* XJustiz, https://xjustiz.justiz.de/downloads/index.php (last visited June 5, 2020).

98. *See* Rechtsprechung im Internet, https://www.rechtsprechung-im-internet.de (last visited June 5, 2020).
99. *See* Gesetze im Internet, https://www.gesetze-im-internet.de (last visited June 5, 2020).
100. Michael J. Bommarito II et al., *OpenEDGAR: Open Source Software for SEC EDGAR Analysis* (2018), https://arxiv.org/abs/1806.04973.
101. Charu C. Aggarwal, Data Mining The Textbook (2015).
102. Ronen Feldman & James Sanger, The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data (2006).
103. Ruslan Mitkov, The Oxford Handbook of Computational Linguistics (2d ed. 2014).
104. Coupette, *supra* note 25; Albert-László Barabási, Network Science (2016).
105. Edward R. Tufte, The Visual Display of Quantitative Information (2d ed. 2001).
106. Steven F. Railsback & Volker Grimm, Agent-Based and Individual-Based Modeling: A Practical Introduction (2d ed. 2019).
107. Deutscher Bundestag, https://www.bundestag.de/abgeordnete/biografien/mdb_zahlen_19 (last visited June 5, 2020).
108. Corinna Coupette & Andreas M. Fleckner, *Das Wertpapierhandelsgesetz (1994–2019) – Eine quantitative juristische Studie*, in Festschrift 25 Jahre WPHG 53–85 (Lars Klöhn & Sebastian Mock eds., 2019).

REFERENCES

63 BGHSt 156.

Aggarwal, Charu C. (2015), Data Mining the Textbook.

Agstner, Peter (2019), *Shareholder Conflicts in Close Corporations between Theory and Practice: Evidence from Italian Private Limited Liability Companies*, Eur. Bus. Org. L. Rev.

Arnold, Christian et al. (2019), *Scaling Lower Court Decisions*, available at https://www.sowi.uni-mannheim.de/media/Lehrstuehle/sowi/Gschwend/Articel/JELS_arnoldetal.pdf (last visited June 5, 2020).

Augsberg, Ino, *Von Einem Neuerdings Erhobenen Empiristischen Ton in der Rechtswissenschaft*, 51 Der Staat 117, 124–25.

Barabási, Albert-László (2016), Network Science.

Beckedorf, Janis et al., *Analyzing High Volumes of German Court Decisions in an Interdisciplinary Class of Law and Computer Science Students – LSIA*, in Computational Legal Studies: The Promise and Challenge of Data-Driven Legal Research (Ryan Whalen ed., 2020).

BGH, Apr. 30, 2020, I ZR 139/15.

BGH, Apr. 5, 2017, 2017 Neue juristische Wochenschrift [NJW] 1819.

BGH, June 20, 2018, Entscheidungen des Bundesgerichtshofs in Strafsachen [BGHSt] 63, 156.

BGH, Sep. 28, 2006, I ZR 261/03, http://juris.bundesgerichtshof.de/cgi-bin/rechtsprechung/document.py?Gericht=bgh&Art=en&nr=39461, (last visited June 5, 2020).

Bommarito, Michael J., II et al. (2018), *OpenEDGAR: Open Source Software for SEC EDGAR Analysis*, https://arxiv.org/abs/1806.04973.

Bundes-und Landesrecht, Justizportal des Bundes und der Länder.

Bundesgerichtshof [BGH] [Federal Court of Justice] Nov. 21, 1991, 116 Entscheidungen des Bundesgerichtshofs in Zivilsachen [BGHZ] 136.

Bundesgesetzblatt, https://www.bgbl.de/ (last visited June 5, 2020).

Bundesverfassungsgericht [BVerfG] [Federal Constitutional Court], Dec. 15, 1983, Entscheidungen des Bundesverfassungsgerichts [BVerfGE] 65, 1.

Buzer, https://buzer.de (last visited June 5, 2020).

BVerfG, 1 BvR 857/15, Sep. 14, 2015, https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/DE/2015/09/rk20150914_1bvr085715.html (last visited June 5, 2020).

BVerwG, June 15, 2015, 7 C 13.13 (not reported).

BVerwG, Oct. 9, 1985, 1986 NJW 1277.

Caldarola, Maria C. & Joachaim Schrey (2020), Big Data und Recht.

Computer Assisted Legal Linguistics, https://www.cal2.eu/index.php (last visited June 5, 2020).
Conferences, Jurix, http://jurix.nl/conferences/ (last visited June 5, 2020).
Coupette, Corinna & Andreas M. Fleckner, *Das Wertpapierhandelsgesetz (1994–2019) – Eine quantitative juristische Studie*, in Festschrift 25 Jahre WPHG 53–85 (Lars Klöhn & Sebastian Mock eds., 2019).
Coupette, Corinna & Andreas M. Fleckner (2018), *Quantitative Rechtswissenschaft*, 73 Juristenzeitung [JZ] 379.
Coupette, Corinna (2019), Juristische Netzwerkforschung – Modellierung, Quantifizierung und Visualisierung relationaler Daten im Recht.
Dejure, https://dejure.org/ (last visited June 5, 2020).
Deutscher Bundestag, https://www.bundestag.de/abgeordnete/biografien/mdb_zahlen_19 (last visited June 5, 2020).
DGRI, https://www.dgri.de/ (last visited June 5, 2020).
Doctoral Research Group "Digital Law", Universitat Heidelberg, https://www.jura.uni-heidelberg.de/digitales_recht/index_e.html (last visited June 5, 2020).
Dokumente, Landtag Des Saarlandes, https://www.landtag-saar.de/dokumente (last visited June 5, 2020).
Dryad, https://datadryad.org (last visited June 5, 2020).
Engst, Benjamin et al. (2017), *Zum Einfluss der Parteinähe auf das Abstimmungsverhalten der Bundesverfassungsrichter – eine quantitative Untersuchung,* 72 JuristenZeitung [JZ], 816.
EUR-Lex.europa.eu, https://eur-lex.europa.eu/ (last visited June 5, 2020).
European Union Directive 2016/679, 2016 O.J. (L 119) 1.
Fachgruppe Rechtsinformatik, FB-RVI, https://fb-rvi.gi.de/fachgruppen/rechtsinformatik (last visited June 5, 2020).
Fair Principles, https://www.go-fair.org/fair-principles/ (last visited June 5, 2020).
Federal Administrative Court [Bundesverwaltungsgericht], Feb. 26, 1997, Entscheidungen des Bundesverwaltungsgerichts [BVerwGE] 104, 105.
Feldman, Ronen & James Sanger (2006), The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data.
Fiedler, Herbert et al., Untersuchungen zur Formalisierung im Recht als Beitrag zur Grundlagenforschung juristischer Datenverarbeitung (UFORED) (1984).
FigShare, https://figshare.com (last visited June 5, 2020).
Fleckner, Andreas M., *Review of Hanjo Hamann: Evidenzbasierte Jurisprudenz (Grundlagen der Rechtswissenschaft, Vol. 23)*, 82 Rabels Zeitschrift für ausländisches und internationales Privatrecht [RabelsZ] 471, 471 n.1.
Fuchs, Thomas, *Die Weiterverwendung der Gemeinfreien Rechtsdatenbank "Juris"* (Apr. 30, 2020), https://delegibus.com/2011,2.pdf (last visited June 5, 2020).
Gemeindeverzeichnis-Informationssystem GV-ISys, Statistisches Bundesamt (Destatis), https://www.destatis.de/DE/Themen/Laender-Regionen/Regionales/Gemeindeverzeichnis/_inhalt.html (last visited June 5, 2020).
German Association for Law and Society, https://www.rechtssoziologie.info/ (last visited June 5, 2020).
Gesetz zur Förderung der elektronischen Verwaltung [EgovG] [Act to promote electronic government], https://www.gesetze-im-internet.de/englisch_egovg/index.html (last visited June 5, 2020).
Gesetze im Internet, https://www.gesetze-im-internet.de/ (last visited June 5, 2020).
Github, https://github.com/okfde/ifg-statistik (last visited June 5, 2020).
GovData, https://www.govdata.de/ (last visited June 5, 2020).
Hamann, Hanjo & Friedemann Vogel (2018), *Evidence-Based Jurisprudence Meets Legal Linguistics – Unlikely Blends Made in Germany*, 2017 BYU L. Rev. 1473.
Hamann, Hanjo & Leonard Hoeft (2017), *Die Empirische Herangehensweise im Zivilrecht,* 217 Archiv für civilistische Praxis [AcP], 311.
Hamann, Hanjo (2014), Evidenzbasierte Jurisprudenz.
Hensler, Deborah R. & Matthew A. Gasperetti, *The Role of Empirical Legal Studies in Legal Scholarship, Legal Education and Policy-Making: A U.S. Perspective*, in Rethinking Legal Scholarship: A Transatlantic Dialogue (Rob van Gestel, Hans-W. Micklitz & Edward L. Rubin, eds., 2017).

Hoeren, Thomas & Michael Bohne, *Von der Mathematischen Strukturtheorie zur Integrationsdisziplin*, *in* Informatik in Recht und Verwaltung: Gestern – Heute – Morgen 23–36 (Roland Traunmüller & Maria A. Wimmer eds., 2009).
Hoffmann-Riem, Wolfgang (2018), Big Data – Regulative Herausforderungen.
https://hilano.de (last visited June 5, 2020).
https://www.iri.uni-hannover.de/de/ (last visited June 5, 2020).
https://justiz.de/onlinedienste/bundesundlandesrecht/index.php (last visited June 5, 2020).
Informationsweiterverwendungsgesetz [Federal Act on the Re-Use of Public Sector Information], https://www.gesetze-im-internet.de/iwg/ (last visited June 5, 2020).
Institut Für Rechtsinformatik Universitat Des Saarlandes, https://rechtsinformatik.saarland/en/ (last visited June 5, 2020).
IRIS – Internationales Rechtsinformatik Symposion, Universitat Wien, https://rechtsinformatik.univie.ac.at/iris/ (last visited June 5, 2020).
IT Planning Council, https://www.it-planungsrat.de/EN/home/home_node.html (last visited June 5, 2020).
Jahresabschluss zum 31. Dezember 2018 und Lagebericht [annual report 2018], https://www.juris.de/jportal/cms/juris/media/pdf/unternehmen/juris_GmbH_Geschaeftsbericht_zum_31122018.pdf (last visited June 5, 2020).
Katz, Daniel M. & Michael J. Bommarito (2014), *Measuring the Complexity of the Law: The United States Code*, 22 Artificial Intell. & L. 337.
Katz, Daniel M. et al., (2020) *Complex Societies and the Growth of the Law* 10 Sci. Rep. 18737 (2020), available at https://doi.org/10.1038/s41598-020-73623-x.
Kocher, Eva (2017), *Rechtssoziologie: Das Recht der Gesellschaft und die Gesellschaft des Rechts*, 8 Rechtswissenschaft [RW] 153, 176.
Kuntz, Wolfgang (2006), *Quantität Gerichtlicher Entscheidungen als Qualitätskriterium juristischer Datenbanken*, JurPC Web-Dok, 12/2006.
Links zu den Parlamentsdokumentationen, Parlamentsspiegel, https://www.parlamentsspiegel.de (last visited June 5, 2020).
Medvedeva, Masha et al. (2019), *Using Machine Learning to Predict Decisions of the European Court of Human Rights*, Artificial Intell. & L.
Mitkov, Ruslan (2014), The Oxford Handbook of Computational Linguistics (2d ed. 2014).
Morell, Alexander (2014), *Die Rolle von Tatsachen bei der Bestimmung von "Obliegenheiten" im Sinne von § 254 BGB am Beispiel des Fahrradhelms*, 214 Archiv für civilistische Praxis [AcP], 387.
Neupane, Jayanti Bhandari et al. (2019), *Characterization of Leptazolines A–D, Polar Oxazolines from the Cyanobacterium Leptolyngbya sp., Reveals a Glitch with the "Willoughby–Hoye" Scripts for Calculating NMR Chemical Shifts*, 21 Organic Letters 8449.
Normendokumentation, Bundesamt für Justiz, https://www.bundesjustizamt.de/DE/Themen/Gerichte_Behoerden/Normendokumentation/Normendokumentation_node.html (last visited June 5, 2020).
Open Legal Data, https://de.openlegaldata.io (last visited June 5, 2020).
Parlamentsdokumente, Bundesrat, https://www.bundesrat.de/DE/dokumente/dokumente-node.html (last visited June 5, 2020).
Petersen, Niels (2010), *Braucht die Rechtswissenschaft eine Empirische Wende?*, 49 Der Staat 435, 455.
Petersen, Niels & Konstantin Chatziathanasiou (2019), *Empirische Verfassungsrechtswissenschaft: Zu Möglichkeiten und Grenzen quantitativer Verfassungsvergleichung und Richterforschung*, 144 Archiv des öffentlichen Rechts [AöR], 501.
Prof. Dr. iur. Emanuel V. Towfigh, Chair of Public Law, Empirical Legal Research and Law & Economics, EBS Universität, https://www.ebs.edu/de/organ/lehrstuhl-fuer-oeffentliches-recht-empirische-rechtsforschung-und-rechtsoekonomik (last visited on June 5, 2020).
Prof. Dr. Florian Matthes, Sebis Public Website (Dec. 18, 2019), https://wwwmatthes.in.tum.de/pages/88bkmvw6y7gx/Prof.-Dr.-Florian-Matthes (last visited June 5, 2020).
Prof. Dr. Michael Gertz, Database Systems Research Group, https://dbs.ifi.uni-heidelberg.de/team/gertz/ (last visited June 5, 2020).
Prof. Dr. Niels Petersen, Chair for Public Law, International European Law and Empirical Legal Research, WWU Münster, https://www.jura.uni-muenster.de/de/institute/lehrstuhl-fuer-oeffentliches-recht-voelker-und-europarecht-sowie-empirische-rechtsforschung (last visited on June 5, 2020).

Prof. Dr. Thomas Hoeren, ITM, https://www.itm.nrw/organisation/prof-dr-thomas-hoeren/ (last visited June 5, 2020).
Railsback, Steven F. & Volker Grimm (2019), Agent-Based and Individual-Based Modeling: A Practical Introduction (2d ed. 2019).
Raiser, Thomas (2013), Grundlagen der Rechtssoziologie 15 (6th ed. 2013).
Rechtsprechung, Justizportal des Bundes und der Länder, https://justiz.de/onlinedienste/rechtsprechung/index.php (last visited June 5, 2020).
Rechtsprechung im Internet, https://www.rechtsprechung-im-internet.de (last visited June 5, 2020).
Röhl, Klaus F. (2007), Rechtssoziologie, https://www.ruhr-uni-bochum.de/rsozinfo/pdf/Roehl-RS-10-Nachtrag.pdf (last visited June 5, 2020).
Ruhl, J.B. & Daniel M. Katz (2015), *Measuring, Monitoring, and Managing Legal Complexity*, 101 Iowa L. Rev. 191.
Ruhl, J.B. et al. (2017), *Harnessing Legal Complexity*, 355 Sci. 1377.
Ruhl, J.B. (2012), *Law's Complexity – A Primer*, 24 Ga. St. U. L. Rev.
Ruppert, Eugen et al., *LawStats – Large-Scale German Court Decision Evaluation Using Web Service Classifiers*, in Machine Learning and Knowledge Extraction (Andreas Holzinger et al. eds., 2018).
Säschsischer Landtag, http://edas.landtag.sachsen.de/ (last visited June 5, 2020).
Section 100 Verwaltungsgerichtsordnung [VwGO] [Rules of the Administrative Courts], https://www.gesetze-im-internet.de/vwgo/ (last visited June 5, 2020).
Section 11 Landesjustizkostengesetz Hamburg [LJKG-HH] [Hamburg Legal Cost Act], http://www.landesrecht-hamburg.de/jportal/portal/page/bshaprod.psml;doc.id=jlr-JKostGHArahmen (last visited June 5, 2020).
Section 22 Justizverwaltungskostengesetz [JVKostG] [Federal Legal Cost Act], https://www.gesetze-im-internet.de/jvkostg/BJNR265500013.html (last visited June 5, 2020).
Section 299 Zivilprozessordnung [ZPO] [Code of Civil Procedure], https://www.gesetze-im-internet.de/zpo/ (last visited June 5, 2020).
Section 476 German Code of Criminal Procedure [Strafprozessordnung].
Sections 147, 406e Strafgesetzbuch [StGB] [Code of Criminal Procedure], https://www.gesetze-im-internet.de/stgb/.
Stember, Jürgen et al., *Studie zum E-Government-Gesetz*, http://egov.hs-harz.de/index.php/download/category/2-publikationen?download=32:studie-zum-e-government-gesetz (last visited June 5, 2020).
The German Journal of Law and Society, https://www.degruyter.com/view/j/zfrs (last visited June 5, 2020).
Tufte, Edward R. (2001), The Visual Display of Quantitative Information (2d ed. 2001).
Tukey, John W. (1980), *We Need Both Exploratory and Confirmatory*, 34 Am. Statistician 23 (1980).
Universitat Hamburg, https://www.inf.uni-hamburg.de/en/inst/ab/lt/home.html (last visited June 5, 2020).
Verwaltungsgericht Köln [Cologne administrative trial court] Sep. 12, 2002, 6 K 4342/99, http://www.justiz.nrw.de/nrwe/ovgs/vg_koeln/j2002/6_K_4342_99urteil20020912.html (last visited June 5, 2020).
Verwaltungsgerichtshof Baden-Würtemberg [VGH] [Higher Administrative Court] May 7, 2013, 10 S 281/12, http://lrbw.juris.de/cgi-bin/laender_rechtsprechung/document.py?Gericht=bw&nr=16959 (last visited June 5, 2020).
Vogel, Friedmann et al. (2018), *Computer-Assisted Legal Linguistics: Corpus Analysis as a New Tool for Legal Studies,* 43 Law & Social Inquiry 1340.
Vogel, Friedemann (2019), *Legal Linguistics in Germany. History, Working Groups, Concepts*, in Legal Linguistics Beyond Borders: Language and Law in a World of Media, Globalisation and Social Conflicts 103, 103–4 (Friedemann Vogel ed., 2019).
Vogel, Friedemann (2017), *Rechtslinguistik: Bestimmung einer Fachrichtung*, *in* Handbuch Sprache im Recht (Ekkehard Felder & Friedemann Vogel eds., 2017).
Ward, Jonathan S. & Adam Barker (2013), Undefined By Data: A Survey of Big Data Definitions, https://arxiv.org/abs/1309.5821.
Willkommen in DIP, DIP, http://dipbt.bundestag.de/dip21.web/bt (last visited June 5, 2020).

Wissenschaftlicher Dienst des Bundestages [WD] [Research Services of the German Bundestag], June 28, 2019, WD 3 – 3000 – 134/19, https://www.bundestag.de/resource/blob/655082/32a17c3834d5c5c5d6f5a7232f0491c0/WD-3–134–19-pdf-data.pdf (last visited June 5, 2020).

Wrase, Michael (2006), *Rechtssoziologie und Law and Society – Die Deutsche Rechtssoziologie Zwischen Krise und Neuaufbruch,* 27 Zeitschrift für Rechtssoziologie 286, 308.

XJustiz, https://xjustiz.justiz.de/downloads/index.php (last visited June 5, 2020).

Zachert, Ulrich, *Der Arbeitsrechtsrechtsdiskurs und die Rechtsempirie – Ein schwieriges Verhältnis*, WSI Mitteilungen 421 (2007).

Zenodo, https://zenodo.org (last visited June 5, 2020).