Legal Knowledge and Information Systems
M. Palmirani (Ed.)

doi:10.3233/978-1-61499-935-5-71

E-Science and the Law. Three Experimental Platforms for Legal Analytics

Nicola LETTIERI [a,1], Alfonso GUARINO [b] and Delfina MALANDRINO [b]
[a] *National Institute for Public Policy Analysis (INAPP), 00198, Rome, Italy*
[b] *Department of Computer Science, University of Salerno, 84084, Fisciano, Italy*

Abstract. The paper presents three experimental platforms for legal analytics, online environments integrating heterogeneous computational heuristics, information processing, and visualization techniques to extract actionable knowledge from legal data. Our goal is to explore innovative approaches to issues spanning from information retrieval to the quantitative analysis of legal corpora or to the study of criminal organizations for research and investigative purposes. After a brief introduction to the e-science paradigm and to the role played in it by research platforms, we focus on visual analytics as a viable way to interact with legal data. We then present the tools, their main features and the results so far obtained. The paper ends up with some considerations about the computational turn of science and its role in promoting a much needed interdisciplinary and empirical evolution of legal research.

Keywords. e-science, analytical platforms, legal visual analytics, network analysis

1. Introduction

Last few years have been marked by a growing integration of traditional research methods and data-driven computational heuristics. The term *E-science* [1] today embraces in one definition the features of an emerging research paradigm in which every stage of the scientific endeavor, from the formulation of research questions to the distribution of findings, is somehow ”enhanced” by digital information-processing, computational technologies and distributed collaboration infrastructures. The paradigm is spreading not only in empirical research [2]- mainly through big data analytics and machine learning - but also in theory-making - mostly by means of computer simulation model building [3]. In this scenario, analytical platforms - software environments integrating different tools of the e-science pipeline - are becoming the cornerstone of a change that challenges established epistemologies in all the areas of science with a peculiar impact in the social sciences where computational heuristics have spread at a slower pace. Against this backdrop, a major challenge for legal scholars is to start drawing up new tools capable of exploiting the data streams and the computing power today available to answer existing

[1] Corresponding Author: Nicola Lettieri; E-mail: n.lettieri@inapp.org

research questions or come up with new ones. This work presents three experimental analytical platforms combining data mining, visualization and machine learning to extract knowledge from the analysis of legal materials. Projects presented explore new ways to address the needs of legal science and practice. In the background, the belief that the computational turn of science is providing precious food for thought for a much needed empirical, quantitative and interdisciplinary evolution of legal research.

2. Research platforms and Visual Analytics

The evolution of ICT has greatly pushed forward the computational science paradigm: today almost every area of science is shifting towards the model of a research powered by machines [4] and computing intruments. We are witnessing the spread of data and computation-driven research platforms supporting scientists in different ways (for an overview see: [5]): allowing to more easily explore the ever growing amount of scientific papers today available (e.g. *PubChase*); supporting the analysis and the management of large sets of data and programming code (e.g. *GitHub*); facilitating the collaboration between colleagues and the handling of online experiments (e.g. *LabGuru, Asana*); simplifying the publication of papers and the analysis of their impact(e.g. *eLife, GigaScience*).

Tools above mentioned are all somehow triggering innovation in the practice of science. Particularly interesting is, in this evolving scenario, the possibility of using research platforms to combine the insights offered by computational heuristics, with intuitive visualizations allowing to better make sense of the interaction with huge and often obscure amounts of data. This is even more true in those fields of social sciences that, like law, are less familiar with quantitative approaches and advanced computational methods.

A promising frontier in this regard is represented by the adoption of methods and technical solutions coming from *Visual Analytics*(VA) [6], a fledgling research field aiming to provide scientists with innovative ways to turn data into knowledge while also enabling them to act on their findings in real-time. As highlighted in [7], VA explores new ways to: i) synthesize information and derive insights from massive, dynamic, ambiguous, and often conflicting data; ii) detect the expected and discover the unexpected; iii) provide timely, defensible, and understandable assessments; iv) communicate these assessments effectively for action.

3. Visualization, analytics and law

As a matter of fact, the idea of exploiting visual metaphors to ease the management and the understanding of legal information has repeatedly made its appearance in the history of law. The use of charts and maps dates back to the Middle Ages when the so called *"arbor"* (*"arbor"* is the Latin word standing for *"tree"*) diagrams [8,9] were used to graphically exemplify legal concepts like the impediments to marriage, or to depict the stages of procedure in Roman Law. Legal metaphors appear again centuries later, when Henry Wigmore [10] proposed the use of diagrams - the *"Wigmore charts"* - to support the analysis of ambiguous evidences and facilitate reasoning required to confirm or rebut hypotheses presented in court. Over the years, the interest in graphical methods led not only to the implementation of Wigmore diagrams through computational tools, but also to other visual representations of legal matters like Bayesian networks that support probabilistic inference in forensic science [11,12,13,14,15,16]. In more recent times, the availability of increasingly powerful technologies (e.g. user-friendly visualization tools and data mining libraries) and of insightful computational heuristics has prompted a growing interest in the development of advanced tools for the analysis of legal information. This is witnessed by a number of experiences [17,18] at the boundaries between visualization, analytics and law, in an area we could define as *Visual Legal Analytics* (VLA).

Some examples can be useful to get an idea of trends emerging in the field. *Ravel Law* is a commercial web platform for computer-assisted legal research that integrates machine learning, natural language processing, and visualization to help lawyers and legal scholars in retrieving and analyzing US case law. While traditional legal search engines use textual interfaces offering a poorer user experience, *Ravel Law* exploits graph visualization not only to allow the access to case law full texts, but also to convey information about cases like the relevance of precedents or the connections between judgments. Another experience worthy of attention is *Lexmex*, an online system visually representing the relations between the French Civil Code and other pieces of legislation. The tool generates a graph transforming laws in nodes and citations in edges: the size of nodes depends on the number of connections between the nodes, while the colors allow to identify groups of highly connected norms. Interaction with data is enabled by essential navigation solutions such as zooming, node selection (to show contextual information) and search by keywords. A more recent and interesting work [19] presents an open source software for the analysis and the visualization of citation network of Dutch case law. The goal is to support legal research questions, including the identification of relevant judgments, the comparison of precedents with those identified in the literature, and the determination of clusters of related cases. In a similar direction, again, we can cite a project using visualization to depict and explore the history of Swiss Federal Law [20].

4. Three experimental platforms for legal analytics

In this section we present *KnowLex*, *EUCaseNet* and *CrimeMiner*, three experimental research platforms [21,22,23] exploiting e-science methods to meet in new ways the needs of legal science and practice. The description of the features of each tool is accompanied by a brief presentation of the results obtained so far.

4.1. Mapping and exploring norms' "neighborhood": KnowLex

KnowLex is a web application (https://bit.ly/2PfmqPu) designed for visualization, exploration, and analysis of legal documents coming from different Italian sources and connected to a given piece of legislation. Understanding the legal framework relating to a given issue often requires the analysis of complex legal corpora: when legal professionals or citizens try to understand how a given phenomenon is disciplined, their attention cannot be limited to a single source of law, but has to be directed on the bigger picture resulting from all the sources related to the theme under investigation. *KnowLex* exploits data visualization and quantitative analysis to support this activity by means of interactive maps making sense out of heterogeneous documents (norms, case law, legal literature, etc.) and their properties. The tool results from an analysis conducted with legal professionals and students and has already undergone a preliminary evaluation study aiming at evaluating the effectiveness of visualization (compared with that of textual interfaces of traditional database), the usability of the proposed system, and the overall user satisfaction [22].

Features. *KnowLex* is made up of several modules offering different insights and ways of interacting with available data. The "*Reference Norm Network*"(RNN) uses interactive graphs to represent the set of materials connected to a given piece of legislation. *KnowLex* gathers documents (amendments, Supreme Court judgments, constitutional judgments, preparatory works, legal literature) from different datasets and websites starting from a norm chosen by the user (the"Root" norm", Legge 22 Dicembre 2008, n. 203", in our example), and builds a map connecting all of them. The graph (see Fig.1) not only offers an overall view of documents properties and relations but also allows user to access text and information simply by interacting with the nodes. The *Semantic Doctrine Naviga-*

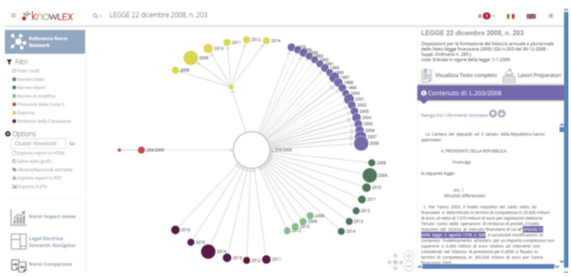

Figure 1. KnowLex: the Norm Graph Navigator.

tor module (Figure 2) shows a treemap depicting the tags used to classify the corpus of scientific articles published with reference to the *Root norm* exploiting the classification by subject used by the Italian bibliographic database DOGI. The map allows to: i) visually explore the law's impact on the different areas of the legal system (e.g., if 70% of the articles related to a certain law is tagged administrative law, it is likely that this is the field on which the law has had the most impact); ii) understand how the doctrine has evolved and what themes have drawn the attention over time (areas have a color that varies according to the date of the articles with a given tag); iii) retrieve papers abstracts and bibliographic data by clicking on the tiles. The *Norm Impact Meter* module (Figure 3) combines graphs relating to different categories of documents linked to the *Root Norm* (amendments, repeals, citations contained in other laws, judgments of different authorities that apply the norm, and reviews of constitutionality). The visualization allows to extract a coarse-grained quantitative image of the laws impact on the legal system. The *Norm Comparator* module (see Figure 4) exploits data relating to the papers tags from the *Semantic Doctrine Navigator* to compare two laws in semantic terms. Different views allow the user to understand: *i)* whether two given laws have dealt with the same topics (which is represented with an histogram), and *ii)* how much the two laws are similar to each other (which is done by calculating Euclidean distance and the Pearson

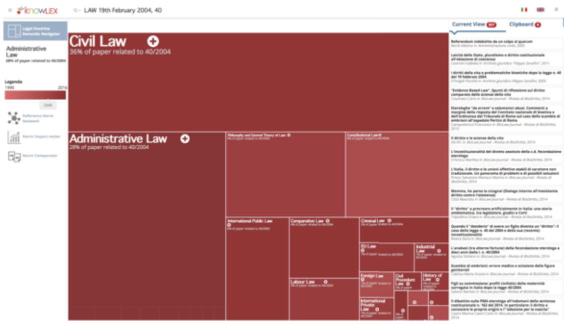

Figure 2. KnowLex: Semantic Doctrine Navigator.

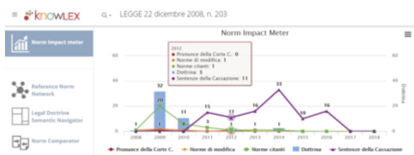

Figure 3. KnowLex: Norm Impact Meter.

correlation coefficient and is represented through gauge meters). The feature becomes interesting when used to compare two norms that have the same function like the Finance acts. A different "semantic fingerprint" of the two laws suggests that the legislator has focused his attention on different priorities (e.g., public education rather than health-care) in two budget years.

Technologies. From the technological point of view, KnowLex is built on a three-layer architecture: on the client-side, it has been developed using JavaScript open-source libraries (i.e., *Sigma.js*, *Linkurious.js* and *D3.js*). On the server-side, data are gathered through HTTP requests using cURL, while PHP wrappers parse different external sources and produce structured data in JSON format. Data are stored in a MySQL database.

4.2. Analyzing the features of EU case law: EUCaseNet

EUCaseNet is an online laboratory allowing legal scholars to explore the EU case law corpus in real time using computational heuristics and visualization techniques. The tool is freely accessible (https://bit.ly/2yVTFNr) and has already allowed [23] a series of interesting experiments about the advantages potentially deriving from network-based inferences in the discovery of features characterizing both the EU case law as a whole, and single judgments (e.g. the relevance of precedents)

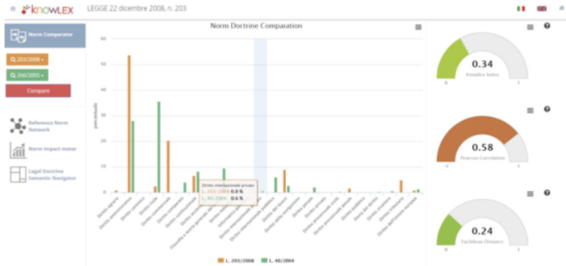

Figure 4. KnowLex: Norm Comparator.

Features.*EUCaseNet* offers basically three functionalities. *Citation Analysis*. EUCaseNet allows the interactive application of NA measures (centrality, Page Rank, Community detection etc.) to the network of the citations connecting all the ECJ judgments. The size and the color of the nodes can be varied based on the value of specific NA metric (e.g. betweenness centrality) allowing users to visually explore the potential correspondence between the results of computational analysis and the features of judicial decisions like judges' behavior (e.g., the propensity to cite other cases) or the nature and the features of the most cited precedents (Figure 5). *Topics and trends analysis*.A

Figure 5. EUCaseNet: citations network graph of the first 350 judgments in terms of in-degree and their citation network.

heatmap (Figure 6) allows to visually compare the topics most covered by the entire EU case law (a sort of trending topics dened by the system using ofcial ECJ data). The visualization allows to intuitively understand if a given ruling (e.g. a judgment recognized as particularly important by legal literature) deals with issues to which the EU case law has already reserved particular attention in the past. The measurement, moreover, suggests objective indices of the emergence of new trends, offering insights for further investigations. *Evolution of case law topics* A linegraph (Figure 7) depicts, in diachronic terms, some aspects of the evolution of the EU case law allowing to see how the number

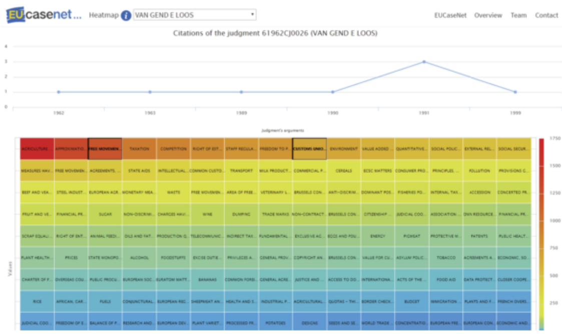

Figure 6. EUCaseNet: heatmap.

of judgments dealing with a given topic (e.g. free movement of goods in EU) evolved over time. The possibility of overlaying several lines related to different topics, allows legal scholars (e.g., historians of law) to make useful comparisons and to identify notable correlations in the changes of most frequent topics dealt with by ECJ.

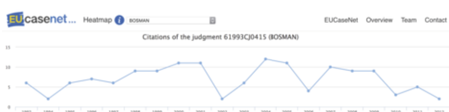

Figure 7. EUCaseNet: linegraph.

Technologies. *EUCaseNet* is built upon a three-tier architecture, implemented by following a typical Model-View-Controller layer architecture. The Data Persistence and Business layers are implemented server-side, through Java Servlet components, within Apache Tomcat. The User Interface Layer is implemented with commonly used JavaScript libraries (i.e., Sigma.js).

4.3. A social-legal analysis of organized crime: CrimeMiner

CrimeMiner is a platform aiming to explore how the combination of data mining, SNA, machine learning and data visualization can contribute to a deeper understanding of structural and functional features of criminal organizations starting from the analysis of even simple relational and investigative data. The tool, aimed at both scientific and investigative purposes, has been developed in collaboration with public prosecutors of the Italian Direzione Investigativa Antimafia (the Italian Investigative Directorate anti-mafia) and has been validated within a case study based on data coming from real criminal investigations [21]. Data currently handled by *CrimeMiner* consist in people records/charges and telephone/environmental tappings and they are visualized as a graph G=(V, E) where V=people and E=telephone/environmental tapping shaping the relationship among individuals that can be analyzed using SNA metrics.

Features. *CrimeMiner* offers visualizations that illuminate different aspects of the criminal group under investigation. The *Wiretaps graph* (Figure 8) offers an intuitive view of the communications/social interactions network within the organization. User can apply different SNA metrics (e.g. community detection algorithms, PageRank) to retrieve insights about the social role of single individuals (leader, broker etc.) or the features of subcommunities (e.g. specialization in given criminal activities).

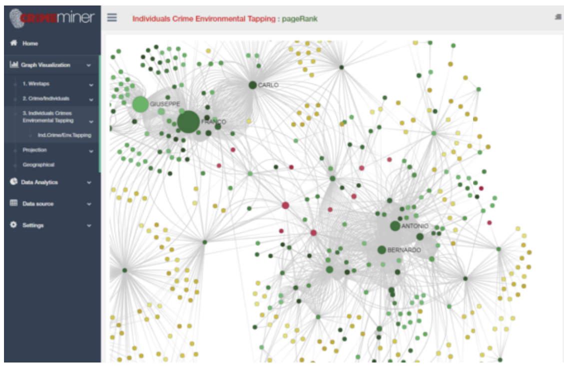

Figure 8. CrimeMiner: individual wiretaps graph

The *GIS map* displays on a map the geographical position (place of residence) of the people under investigation, preserving information about the properties (degree, betweenness, Page Rank values etc.) of single nodes. This allows to discover potentially invisible relationships between the features of the node and its geographical position.

Technologies. *CrimeMiner* is built upon the Java EE Spring Data *Neo4j* framework whose architecture is structured in four layers. In details: *i)* the *Storage layer* stores all data (including graphs) under examination, such as personal details of investigated people and tapping records; *ii)* the *Mapping layer* is responsible of the mapping of *Neo4j* relations and entities in Java classes; *iii)* the *Business layer* processes data mapped in the Mapping layer and provides developed services to the top layer (SNA metrics are defined in within this layer); *iv)* the *Presentation layer* includes the user interface allowing users to interact with *CrimeMiner* features (exploiting Linkurious.js and HighCharts.js).

5. Conclusions

The projects and the tools above described are still very preliminary attempts to explore potential intersections between the legal world and some of the emerging e-science methodologies, and will thus be subject of further developments. On the other hand, despite the publication of some works [24], the adoption of the e-science techniques in legal contexts is still in its infancy and in-depth reflections will be necessary to better exploit them. What is probably before us is in any case the opportunity of a significant methodological and scientific change, especially if we look at analytics and visualization not simply as ways to ease the access to legal materials, but as unprecedented solutions to delve into the complexity of legal world. Actually, collecting data and extracting knowledge from them is contributing to change the whole infrastructure of practices, technologies and perspectives in social sciences [25], and this is going to affect ever more legal research. From this viewpoint, our experience has resulted also into an opportunity to reflect in general terms on the future of the law both on a scientific and methodological standpoint. A first outcome of our reflection is a clearer idea of the gradual shift of law to what has been called the "machine science" paradigm [4], a research model that promises unprecedented scientific insights thanks to a wise combination of theories, data and code. In the perspective of an instrument-enabled future of law - similar arguments can be made for the evolution of legal enforcement towards techno-regulation [26,27,28] - legal scholars will ever more be involved in the challenge of designing tools, including platforms, with implications unfolding also on the theoretical level. It will be up to lawyers to gradually learn new skills and languages (from the computational to the technical ones), so to rethink their research questions, conceptual categories, methods of study and relationships with other sciences. A second worthy result of the work so far done, is the hunch of considering e-science methods as the entry point, in the legal field, of what has been called "computational empiricism" [29,30], the new perspective of an empirical research mainly rooted in the power of data and computational heuristics. After all, the project of what we call "legal computational empiricism", is consistent with some of the emerging trends in the debate on the future of legal science. As witnessed by the flourishing of fields like Empirical Legal Studies [31] or New Legal Realism [32], last few years have been marked by growing attention for the application to legal issues of empirical methods developed in the social sciences. Thanks to their capacity to integrate increasingly advanced ways of tracking and measuring reality, analytical platforms will probably be part of the empirical (experimental, quantitative) evolution of the legal science with an impact on the way scholars and practitioners think about their goals and methods. What is needed is not a sloppy juxtaposition of methods and scientific perspectives but a serious work on the theoretical and experimental level. To use the words of Franz Leeuw [33], "*the more empirical legal research is a "growth industry", the more important it is to understand and discuss epistemological problems of this field of study*" tackling with fundamental issues including "*how to operationalise legal concepts, where to find data (stored, but also Big Data)*" and, "*how to bring empirical evidence to the fore, in such a way that it can be understood and used by lawyers, legislators and regulators*". Research at the borders between e-science and law is not the only solution, but will be for sure part of the effort.

References

[1] T. Hey and A. Trefethen, "The data deluge: An e-science perspective," *Grid computing: Making the global infrastructure a reality*, 2003.
[2] R. Kitchin, "Big data, new epistemologies and paradigm shifts," *Big Data & Society*, 2014.
[3] E. Winsberg, *Science in the age of computer simulation.* University of Chicago Press, 2010.
[4] J. Evans and A. Rzhetsky, "Machine science," *Science*, 2010.

[5] N. Lettieri, A. Altamura, R. Giugno, A. Guarino, D. Malandrino, A. Pulvirenti, F. Vicidomini, and R. Zaccagnino, “Ex machina. Analytical platforms, Law and the challenges of Computational legal science,” *Future Internet*.

[6] J. Kohlhammer, D. Keim, M. Pohl, G. Santucci, and G. Andrienko, “Solving problems with visual analytics,” *Procedia Computer Science*, 2011.

[7] D. Keim, G. Andrienko, J.-D. Fekete, C. Görg, J. Kohlhammer, and G. Melançon, “Visual analytics: Definition, process, and challenges,” in *Information visualization*, Springer, 2008.

[8] A. Errera, *Arbor actionum: genere letterario e forma di classificazione delle azioni nella dottrina dei glossatori*. Monduzzi Ed., 1995.

[9] C. Radding and A. Ciaralli, *The Corpus iuris civilis in the Middle Ages: Manuscripts and transmission from the sixth century to the juristic revival*. Brill, 2006.

[10] J. H. Wigmore, “Problem of proof,” *Ill. LR*, 1913.

[11] P. Tillers, “Picturing factual inference in legal settings,” 2005.

[12] A. Biedermann and F. Taroni, “Bayesian networks and probabilistic reasoning about scientific evidence when there is a lack of data,” *Forensic science international*, 2006.

[13] F. Taroni, C. G. Aitken, P. Garbolino, and A. Biedermann, *Bayesian networks and probabilistic inference in forensic science*. Wiley Chichester, 2006.

[14] A. B. Hepler, A. P. Dawid, and V. Leucari, “Object-oriented graphical representations of complex patterns of evidence,” *Law, Probability & Risk*, 2007.

[15] T. F. Gordon, “Visualizing carneades argument graphs,” *Law, Probability and Risk*, 2007.

[16] B. Verheij, “Argumentation support software: boxes-and-arrows and beyond,” *Law, Probability and Risk*, 2007.

[17] R. Winkels, N. Lettieri, S. Faro, *et al.*, *Network analysis in law*. Napoli Edizioni Scientifiche Italiane9788849527698, 2014.

[18] R. Whalen, “Legal networks: The promises and challenges of legal network analysis,” *Mich. St. L. Rev.*, 2016.

[19] D. V. KUPPEVELT and G. V. DIJCK, “Answering legal research questions about dutch case law with network analysis and visualization,” in *Legal Knowledge and Information Systems: JURIX 2017: The Thirtieth Annual Conference*, IOS Press, 2017.

[20] S. N. André Ourednik, Peter Fleer, “A Visual Approach to the History of Swiss Federal Law,” in *DHd 2016: Modelling - Networking - Visualization*, 2016.

[21] N. Lettieri, D. Malandrino, and L. Vicidomini, “By investigation, i mean computation,” *Trends in Organized Crime*, 2017.

[22] N. Lettieri, A. Altamura, and D. Malandrino, “The legal macroscope: Experimenting with visual legal analytics,” *Information Visualization*, 2017.

[23] N. Lettieri, A. Altamura, A. Faggiano, and D. Malandrino, “A computational approach for the experimental study of EU case law: analysis and implementation,” *Social Netw. Analys. Mining*, 2016.

[24] E. Gomez-Nieto, W. Casaca, I. Hartmann, and L. G. Nonato, “Understanding large legal datasets through visual analytics,” in *Proceedings of 6th Workshop on Visual Analytics, Information Visualization and Scientific Visualization (WVIS) in SIBGRAPI*, 2015.

[25] J.-C. Plantin, C. Lagoze, and P. N. Edwards, “Re-integrating scholarly infrastructure: The ambiguous role of data sharing platforms,” *Big Data & Society*, 2018.

[26] R. Leenes, “Framing techno-regulation: An exploration of state and non-state regulation by technology,” *Legisprudence*, 2011.

[27] R. Brownsword, “In the year 2061: from law to technological management,” *Law, Innovation and Technology*, 2015.

[28] P. De Filippi and S. Hassan, “Blockchain technology as a regulatory technology: From code is law to law is code,” *arXiv preprint arXiv:1801.02507*, 2018.

[29] P. Humphreys, *Extending ourselves: Computational science, empiricism, and scientific method*. Oxford University Press, 2004.

[30] P. Humphreys, “Computational empiricism,” in *Topics in the Foundation of Statistics*, Springer, 1997.

[31] P. Cane and H. Kritzer, *The Oxford handbook of empirical legal research*. OUP Oxford, 2010.

[32] T. J. Miles and C. R. Sunstein, “The new legal realism,” *U. Chi. L. Rev.*, 2008.

[33] F. L. Leeuw, “Empirical legal research the gap between facts and values and legal academic training,” *Utrecht L. Rev.*, 2015.