Knowledge of the Law in the Big Data Age
G. Peruginelli and S. Faro (Eds.)

doi:10.3233/FAIA190004

Knowledge Machineries. Introducing the Instrument-Enabled Future of Legal Research and Practice

Nicola LETTIERI
Istituto Nazionale per l'Analisi delle Politiche Pubbliche - INAPP; Università degli Studi del Sannio - Dipartimento di Diritto, Economia, Management e Metodi Quantitativi (Italy)

Abstract. 30 years after the advent of the World Wide Web, information and communication technologies keep triggering deep changes in the way we access, produce and use knowledge. The convergence between data warehouse facilities and computational science heuristics is populating the Internet with cloud infrastructures designed to manage and process information in completely new ways. We are facing the emergence of a new generation of online platforms integrating knowledge management, data analytics, visualization and collaboration tools for purposes that gradually move from information retrieval to scientific research. This Chapter introduces the looming of the platform era in the legal world showing, also by means of concrete examples, how these tools can be used to make the most of the growing amount of legal information today accessible online. The analysis becomes an opportunity to dwell on how computational tools can turn into the emergence of new perspectives in legal research and practice.

Keywords. legal analytics, machine science, digital platforms

1. Introduction

In the era of planetary-scale computation, ICT are offering much more than the direct access to huge amounts of information. The gradual integration of data sharing facilities, online collaboration, computational heuristics, and visualization is bringing about the emergence of online infrastructures providing a wide range of innovative services that go far beyond information retrieval. Digital platforms are not only shaping sociality, economies, politics, and institutions [1] but also transforming the very way knowledge is produced, organized and shared in any field of human activity. This is real whether we think about the access to endless repositories of documents, or scientific research.

Even if in peculiar ways, this process involves law as well. Scholars and practitioners are getting to grips with the creation of platforms[1] as they appear to be a promis-

[1] A case in point – mainly focused on lawyers' perspective – is represented by a talk published online by the *LawLab* (an interdisciplinary teaching and research center focused on legal innovation and technology of the Illinois Tech - Chicago-Kent College of Law): Kennedy, D. (2017, December). *Agile Lawyering in the Platform Era* [Video file], https://vimeo.com/246985325.

ing way to enhance the knowledge of the law in the Big data era. The challenges are many: besides the still existing need to fine-tune technical standards allowing to manage legal documents, we have before us the opportunity of exploiting innovative heuristics to extract better knowledge from the increasing amount of legal information today available. A reference point, in this perspective, lies in the emerging data and computation-driven research paradigm in which every stage of the scientific endeavour, from questions definition to the sharing of results, is enhanced by digital information-processing, computational heuristics and distributed collaboration infrastructures [2]. Technological and methodological findings from the area of 'machine science' [3] can not only inspire new solutions for legal information retrieval or access to law but also open new scenarios in terms of research questions.

The Chapter sets out a critical introduction to the opportunities of a machine-driven evolution of law and legal research. The analysis, that brings together technical, methodological and epistemological issues, is structured as follows. Section 2 provides a brief overview of the emerging 'machine science' paradigm focusing on the role played by online research platforms as tools for data analysis and theory development. Section 3 and 4 introduce the appearance of online analytical platforms in legal field also by referring to some ongoing research projects. Section 5 focuses on a research project aiming to exploit machine learning and visualization to gain new knowledge from both legal information and administrative microdata. The last Section will sketch some considerations about the role of computational and data-driven tools in the evolution of legal science and practice. In the background, the idea that, together with other cultural and methodological factors, they can trigger much-needed changes in the scope, aims and methods of legal research: an empirical and interdisciplinary turn, and an opening towards a methodologically eclectic approach to scientific investigation.

2. Beyond Information Retrieval: Online Platforms and Machine Science

Science is facing a machine-driven future. Cloud computing, open/big data, artificial intelligence and a growing legion of computational heuristics, are pushing researchers toward an ever more symbiotic relationship with machines. Last 20 years have witnessed the development of a series of approaches to scientific research that, regardless of the name adopted – 'computational science' [2]; [3]; [4], 'data science' [5]; [6], 'e-science' [7]; [8] and, more recently, 'machine science' [3] – can be all traced back to a vision attaching to data and, above all, computational tools a prominent role for our understanding of reality.

In this perspective, technologies work as 'epistemic enhancers' [2]: they extend our natural observational and cognitive abilities reframing research practices and the very way we deal with information and knowledge [9]; [10]; [11]. Changes brought by the machine-driven approach to science are significant. If a first fundamental one has been the spread of computer simulations in new subject areas [12]; [13], a second profound change has been the data-driven turn of scientific inquiry. Indeed, once coupled with the power of computation, the volume and variety of information today available have triggered the rise of a further research paradigm [5]; [10] whereby scientific hypoteses are preceded – when not completely replaced – by the identification of hidden patterns in huge amounts of data, rather than stemming from observation made accordingly to an explicit theoretical model [11].

In a time of such profound changes, computational methodologies are spreading not only in empirical research – mainly through Big Data analytics and machine learning – but also in theory-making – mostly by means of computer simulation model building. Scientist's toolkit is steadily expanding, and research has a growing need for new tools seamlessly integrating the building blocks of the data and computation-driven scientific paradigm. In this scenario, digital platforms – infrastructures consisting of hardware, software, networking systems and data management components to perform computationally demanding tasks – are becoming the *sine qua non* of innovative practices in which technological infrastructures are used to support potentially all the stages of the research path from questions definition to interactive data exploration and visualization, experiment modeling, data analysis and sharing of the findings (Figure 1).

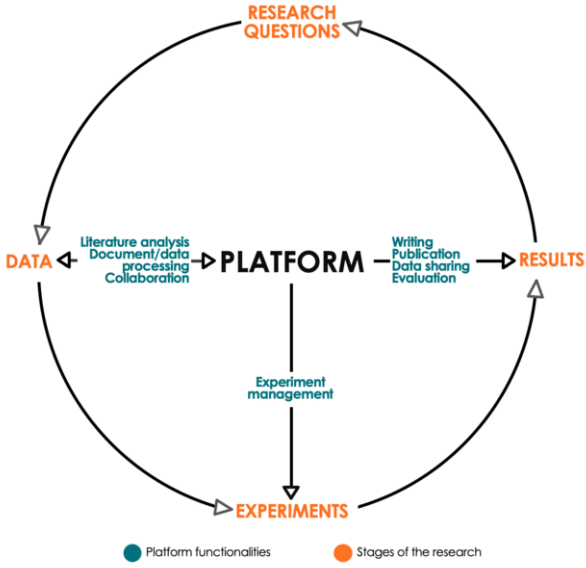

Figure 1. The machine-science research cycle

Actually, the machine-driven evolution of research is already a fact. Last ten years have been marked by the spread of a number of platforms aimed at supporting research in many different ways. The taxonomy proposed in [14], gives an idea of the variety of solutions emerged so far at the same time offering food for thoughts for further applications (Table 1).

Along with the computational science perspective, digital platforms are set to have a disruptive impact on social science and humanities that, for inherent features and historical reasons, are less familiar with quantitative and instrumental research practices than natural science. While growing amounts of information about substantially all social phenomena are stored in digital archives, they are pushing social sciences towards a paradigm in which the mainly qualitative analyses are increasingly backed up with data-driven and computational solutions. This way, the work of any social scientist is going to

Table 1. A taxonomy of digital platforms for science

Category	Description
Literature Analysis	*Goal*: *i*) help researchers in exploring the growing amount of scientific papers available online with ad hoc search engines; *ii*) ease the navigation within the materials integrating visualisation, bookmarking and publication-sharing system. *Examples: Bibsonomy, CiteUlike, Google Scholar, Mendeley, ReadCube, PubChase.*
Data and Code Sharing	*Goal*: *i*) support the management of large sets of data and programming code allowing to collect, share, cite and reuse materials in social and natural science. *Examples: Github, CodeOcean; Socialsci, GenBank; DelveHealth, BioLINCC.*
Collaboration	*Goal*: *i*) facilitate researchers in reaching out to other researchers and find expertise for scientific cooperation; *ii*) communicate research activities to the general public; *iii*) involve the general public in the research efforts according to the 'citizen science' paradigm (e.g. by sharing CPU time or classifying pictures). *Examples: Academia, ResearchGate, Loop; Kudos* and *AcaWiki.*
Experiments	*Goal*: *i*) help researchers in all the activities connected with scientific experiments: equipment and data management; scheduling; research protocols; coding and data analysis; generating and analysing data; visualising results. *Examples: Asana, LabGuru; Quartzy, Transcriptic, GitLab, Wolfram Alpha, Sweave*
Writing	*Goal*: *i*) support paper drafting helping researchers to write papers with other people while keeping track of modifications made by authors on the document; *ii*) allowing researchers to share with colleagues bibliographies, citations, and references. *Examples: Endnote, Zotero, Authorea, ShareLaTex, Citavi*.
Publication	*Goal*: *i*) ease the publication and discussion of papers accelerating scientific interactions and discovery; *ii*) allow authors to connect papers with additional materials like executable code. *Examples: GigaScience, Cureus, ArXiv, Exec&Share, RunMyCode.*
Research Evaluation	*Goal*: *i*) enhance research evaluation supporting in new ways paper review and analysis of the impact of publications. *Examples: PubPeer, Publons, Academic Karma, Altmetric, ImpactStory*.

include also the effort to imagine and experiment with innovative tools, 'places' where theories, data, computational heuristics can converge, be explored[2].

Basically the same applies to the legal world. To give just an example, as suggested by recent developments in Computational legal studies, the integration of legal data and sophisticated processing pipelines is becoming crucial to derive relevant knowledge from legal documents. Projects using machine learning to predict the behaviour of supreme courts [15] or exploiting computational heuristics to assess the structural and semantic complexity of US legal corpora [16] witness the interest in innovative solutions to analyze legal information. In front of this, it makes sense to state that a significant part of future efforts in legal research will have to be devoted to the creation of tools allowing to extract actionable knowledge from data. What is at stake goes further than the effi-

[2]As highlighted by Kitchin [11], the emerging field of computational social science provides an important opportunity to develop more sophisticated models of social life allowing scientists to shift "from data-scarce to data-rich studies of societies from static to dynamic unfoldings; from coarse aggregations to high resolutions; from relatively simple models to more complex, sophisticated simulations".

ciency in carrying out traditional research practices but involves, actually, the potential emergence of new ways to delve into the complexity of legal world [17].

3. The Rise of Legal Machineries: A Brief Background

The development of integrated tools for the legal research has already been a reality for some time. There are several examples of fully-fledged systems – some of which online – dealing with the analysis of legal data and documents. On the other hand, the need for new 'analytical machineries' is looming on today's debate about aims and methods of legal studies. As a matter of fact, two flourishing research areas, Empirical legal studies (ELS) [18]; [19]; [20] and Computational legal science [21]; [22]; [23], are somehow pushing forward debate on the instruments by which law can be explored and studied. The call for a closer integration of empirical analyses into legal scholarship that characterises ELS, for example, inevitably results into the quest for data-driven tools and practices enhancing our understanding of law as a fact [24]. Likewise, even if with different research goals, computational legal scholars are working hard to figure out new ways to make the most of digital tools and heuristics. Current scenario shows different experiences heading in this direction, some linked in a specific way with the study of legal texts, others more oriented to empirical analysis. Here below, in a non-exhaustive manner, a brief overview of some of these experiences giving a sense of the ongoing trends in the development of those we define as 'legal machineries'.

LexMex[3] is a simple but interesting online system exploiting visualization techniques to represent the relations between texts of law for purposes of information retrieval and study. The attention is focused on the French Civil Code and related legislation. The tool generates a graph transforming laws in nodes and citations in edges. The semantics of the visualization is simple: the size of node depends on the number of connections it has with other nodes. Colours correspond to the cluster detected by means of community-detection algorithm allowing to identify groups of highly connected norms. The tool implements essential navigation such as zooming, node selection to show contextual information, and search by keywords.

Ravel Law[4] is a legal analytics research tool, a new type of search engine that combines analytics (natural language processing, machine learning), legal research and graph visualization to help finding the way through a comprehensive case law database from the Harvard Law Library. Unlike traditional legal databases, presenting search results in textual form, using long lists that often hide important cases pages back in search results, *Ravel Law* visually depicts the most important cases on a particular topic as nodes of a network, with edges pointing to subsequent cases citing it. The size of the nodes is used to represent the relevance of the precedents estimated using proprietary algorithms. Besides bringing about a change in how legal research is presented online, *Ravel Law* has also contributed to an interesting and still ongoing debate about the use of 'next gen' research tools in legal education [25].

The use of large collections of digitized texts and citation network analysis to come up with new insightfull methodologies for legal studies is discussed in [26]. Authors have developed an open source software for the analysis and visualization of networks

[3] http://www.lexmex.fr/.
[4] https://home.ravellaw.com/.

of Dutch case law, aiming to support both legal scholars and non-technical researches in their investigations. The basic goal of the research is to answer in new ways legal research questions, including those of determining relevance of case law precedents, comparing the precedents with those identified in the literature, and determining clusters of related cases. In a similar direction, again, it is possible to cite a work presented in 2016 exploiting visual approaches to depict and explore the history of Swiss Federal Law [27]. Authors wonder whether the degree of norms complexity can be measured over time and how it can be represented. To answer this question, they have organized in a same pipeline OCR, parsing (to obtain structured XML from textual documents), data analysis (in particular complexity measures like *Shannon entropy of word use; depth of the hierarchical structure* and *density of external references*), and visualizations.

Turning to the more 'empirically oriented' tools mentioned above, we could cite a number of works. Many experimental software systems have been developed, for example, that combine empirical data from criminal investigations (wiretaps, online communications, environmental tappings), data mining and visualization to support criminal court judges, public prosecutors and law enforcement agencies in the fight against crime [28]; [29]; [30]; [31]. Due to space limitations, we confine ourselves to cite just one recent and illustrative example of this category of tools. *SIIMCO* [32] is a forensic investigation software suite for identifying the influential members of a criminal organization. The system exploits data provided by public prosecutors and police departments (in particular, crime incident reports and mobile communication data about the members of the organization) to create network diagrams representing the structure of criminal organizations. *SIIMCO* employs then network analysis measures to quantify the degree of influence/importance of each individual, to detect subgroups, discover interaction patterns between groups, and identify central members.

4. Fiddling with Legal Analytical Platforms

To make concrete our speculations about the role of analytical platforms in enhancing legal information, we briefly present in the following the rationale and results of an ongoing experimental activity carried out along with the Department of Computer Science of the University of Salerno[5]. The research, still ongoing, aims to gain a first-hand experience with the prospects opened up, in the legal world, by the platform paradigm. Over the last four years, the initiative has already turned into the development of a series of experimental online environments in various ways dealing with legal data, knowledge mining and visualization. The choice to develop from scratch *ad hoc* tools despite the availability of many existing solutions (see, for instance, the variety of research platforms listed in Table 1), makes sense if we turn our mind to a series of needs arising both in legal research and practice:

- *Customization*: tailored solutions (algorithms, workflows) can better fit both the nature (structure, characteristics, errors etc.) of data handled and the research goals.

[5]The projects have seen the involvement of people from law (scholars, lawyers, public prosecutors), computer science, visualization, computational biology.

- *Openness*: from scratch development allows to avoid proprietary software increasing intelligibility and comparability of algorithms and easing the analysis and sharing of the results.
- *Integration*: custom-designed tools make it easier to integrate in one place different functionalities (e.g different kinds of visualizations) and heuristics (e.g. network analysis, machine learning, agent-based modeling etc.).

Based on the consideration of these needs, we have been working on the development of a series of platforms that variously integrate legal data, computational heuristics and visualization to support activities spanning from legal knowledge mining to scholarly research. In more details, experiments headed in the following research directions: enhance legal information retrieval; extend and step up the methodological apparatus available to legal scholars interested in the empirical analyses; figure out new ways to identify and measure the computational correlates of legal concepts (e.g. relevance of case law precedents). Here below a brief description of the tool realized over the years.

- *KnowLex* [33] is a web application designed to allow a more intuitive exploration and analysis of documents coming from different legal sources. The goal of the tool is to exploit visual analytics techniques to support the understanding of the legal framework on a given issue when, as often happens, this requires the analysis of complex legal corpora.
- *EUCaseNet* [34] is an online analytical platform in various ways allowing legal scholars to explore the features of the entire body of European Court of Justice case law: i) by applying network analysis metrics (centrality measures, *Page Rank*, community detection algorithms) to its citation network so to study, for instance, the relevance of precedents; ii) by exploiting statistical visual analytics tools applied to case law metadata.
- *CrimeMiner* [35] is an experimental knowledge mining platform exploiting empirical and legal data from real criminal proceedings (crime incident reports, wiretaps, environmental tappings, criminal records etc.) to support legal practitioners (public prosecutors, judges, law enforcement agencies) and scholars (computational social scientists, criminologists) in investigating the features of criminal networks and of their members. To this end the tool integrates into an ever evolving pipeline information extraction, graph analytics, agent-based social simulation [36] and, in a recent experiment, machine learning.

In more recent times, drawing on the wealth of technical experiences and methodological findings stemming from the projects just above mentioned, we have started a new initiative exploiting legal information and administrative microdata to support both the access to public sector information and data-driven policy design. A more detailed description is offered in the following section.

5. Argos: Visualization and Machine-learning with Normative and Administrative Data

Due also to the spread of the open government paradigm [37], recent years have been marked by a growing effort of public administrations to start the extensive collection and sharing of data generated within their institutional activities. Large-scale administrative data today show high levels of quality in terms of temporal resolution, volume, and struc-

ture [38]. Their diffusion and exploitation raise a growing deal of attention for different reasons. According to EU strategies for the development of the information society[6], they must be made easily available not only to increase the transparency of government and administrative activities but also because, when integrated with other public sector data and analysed with appropriate heuristics, they form the backbone of any evidence-based policy making and agenda setting. In this scenario, it is still difficult fully exploit the potential of data also due to the lack of tool offering tylored analytics and advanced interaction-design solutions in open source format. Challenges for researchers wishing to take advantage of large dataset are different: gaining access to data, developing data management and programming capabilities needed to work with large-scale datasets, and finally thinking of creative approaches to summarize, describe, and analyze information.

Argos is a project that points in this direction by developing a modular online platform allowing the visualization and analysis of large amounts of administrative, legal and economic data. The goal of the project is twofold: i) facilitate the interaction with large-scale administrative data using infographics that make access, comprehension and re-elaboration of information by experts and citizens easier and more intuitive; ii) experiment machine-learning techniques to extract actionable knowledge from cross-cutting reading of heterogeneous (administrative and normative) data. Among the target users there are for sure legal scholars, especially those interested in supporting their studies (e.g. those needed for regulatory impact analysis) with reliable estimates about correlations between regulatory measures and social/economic impacts. The platform prototype includes two modules, both still under development, that are being tested using administrative microdata relating to the impact of the labour market reforms adopted in Italy from 2008 up to 2015[7].

(1) Visualization module

The module (already online, see Figure 2) provides two visualizations that allow to interactively explore the dataset offering insights about the evolutions in the structure of the Italian labour market. More in details, the visualizations can be described as follows: i) the *Zoomable treemap* offers an intuitive and interactive navigable representation of the proportions that bind the elements of groups selected and hierarchically ordered by the user so to answer questions like: How many of the employment contracts signed from 2008 to 2015 are open-ended (OE) or fixed-term (FT)? To what extent education levels are associated with each of the two types of work relationships? ii) the *Bubble Chart/GIS map* allows to intuitively explore the evolution of multidimensional phenomena in order to make evident the fluctuations, over time, of variables that are supposed to be somehow correlated. Our attention was focused on representing trends in OE and FT employment contracts in the period 2008-2015, at the same time offering other information useful to interpret the phenomenon both in an economic and in a legal perspective. To this end, we put in the same visualization a set of different information: trends of OE/FT contracts in individual Italian regions; trend of OE/FT contracts in the North/Center/South areas; enactment of the main labour market reforms; start and end dates of the legislatures. The

[6]The Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information represents the starting point a regulatory process that has developed through different regulatory measures. Among the others, it is worth mentioning the Communication of the EU Commission *Open Data. An Engine for Innovation, Growth and Transparent Governance* COM (2011) 882.

[7]Year of the so called 'JOBS Act', a set of regulatory measures that, among the other things, have introduced in the Italian legal system permanent hiring subsidy and new regulations lowering firing costs.

Bubble Chart is complemented with a GIS map offering a geo-referenced and animated representation of data on labour market trends, so to allow users to visually relate the normative, spatial and temporal dimensions of the changes underway.

(2) Machine learning module
A second module, currently under development, aims at exploiting machine learning – in particular supervised learning algorithms [39] – and other techniques known as 'feature ranking' and 'feature selection' [40] to enhance our analysis. Last few years have been marked by a growing attention to the use of machine learning [38]; [39] in the analysis of economic data. This represents indeed an optimal solution when dealing with enormous amounts of data, when the data are gathered without a carefully controlled experimental design or when dealing with phenomena characterized by complicated nonlinear interactions. Against this backdrop, we have been (and still are) working to extract relevant information from administrative data about workers and contracts aiming to answer questions like: "what are, among the many available, the characteristics that define the workers having more chances to succeed? Are there any patterns in the behaviour of firms and workers possibly linked to regulatory measures adopted over time? After some difficulties encountered in the management of the large quantities of records at our disposal the analysis has already led to interesting insights that will be soon made available online, always using interactive visualization solutions.

6. Conclusions: Four Critical Remarks

In light of the above, we can make some points dealing, in more general terms, with the potential impact of machine science heuristics and tools on the legal field. Experience gained throughout the aforementioned projects suggests that platform technology is not only the enabling factor of more efficient ways of extracting knowledge from legal information but also, potentially, the driver of more deep changes in legal research perspectives. The claim is worthy of a more in depth reflection.

A first remark deals with the very objects of legal research. Last decade has seen a growing attention towards the empirical study of law also by means of quantitative and data-driven approaches[8]. Legal scholars anyway have so far reserved little consideration to the possibility of drawing inspiration from new tools to reconsider more fundamental aspects of their research. We have the feeling that analytical platforms could open up the gates to a sort of 'computation-enhanced legal empiricism' [2], a kind of technologically-enhanced version of empirical legal research [18] exploiting computation not only to identify trends and correlations in case law but also to investigate new topics e.g. using social simulations to illuminate the intricate network of individual and social mechanisms through which law emerges, is applied, produces effects. Besides scientific spill-overs, the fostering of a more empirical stance towards legal world is beneficial also for practical reasons. Factual investigation of legal phenomena increasingly appears to be an indispensable condition for more effective legal solutions able to cope with the complexity of the real world. A great deal of work will obviously have to be

[8] In the overview of the Journal of Empirical Legal Studies it read "With the explosion in information technology, data sources on the legal system are improving in quality and accessibility. Compared with just a few years ago, researchers today can easily access original data sets. [...] The time is ripe for empirical studies of the legal system".

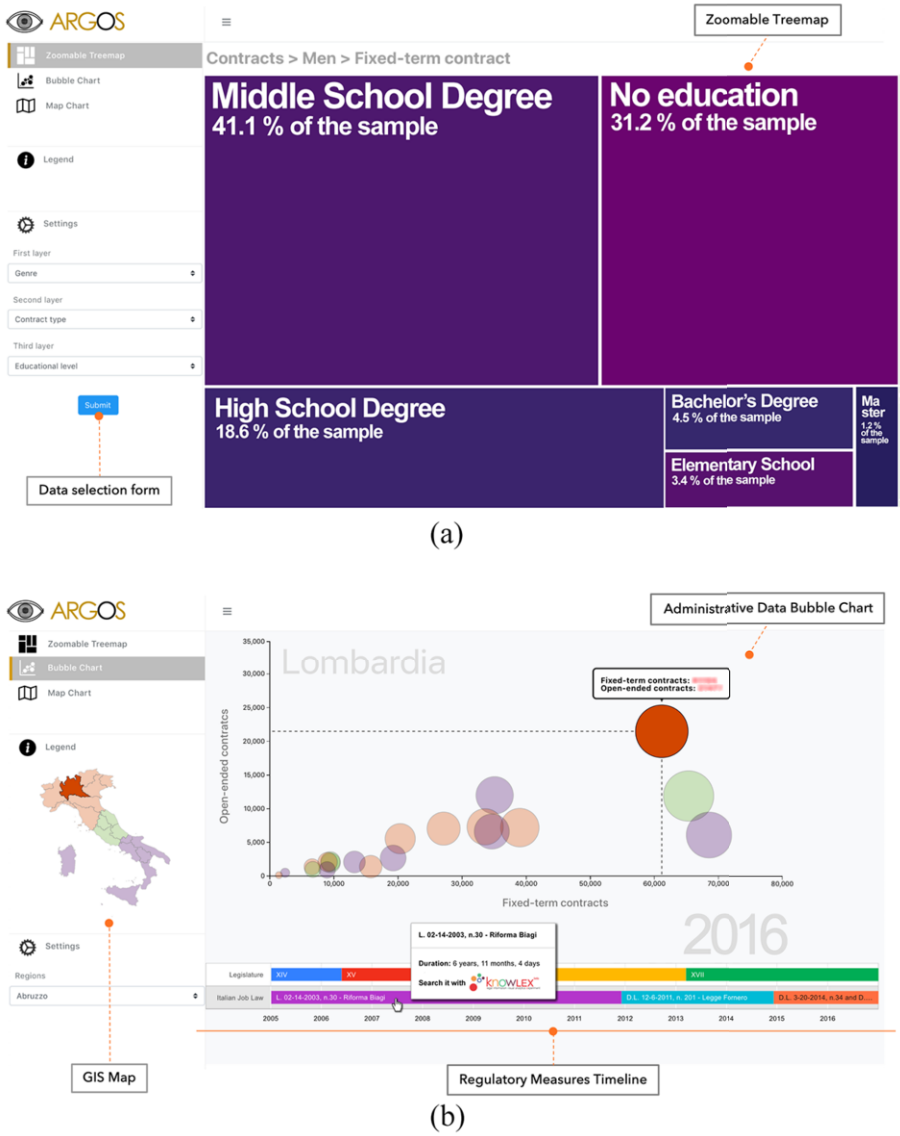

Figure 2. Argos: layout of the Treemap (a), and Bubblechart (b) modules

done to tackle with fundamental issues including how to operationalise legal concepts, where to find data and, above all, how to incorporate results from empirical studies into normative scholarship and practice [20].

A second point involves the relationship between technology and science. Our idea is that, as is happening in computational social science, also legal research is going to become more and more an 'instrument-enabled' practice [41]. Hence lawyers will soon be facing the design of new *ad hoc* 'machines' and heuristics if they want to advance their understanding of legal and social phenomena. The shift, it should also be stressed, is full of epistemological implications of which it is good to be aware. The way we design

and use research tool is a theory-laden process and this is true also for information technologies and analytical platforms. Decisions about data to be processed, functionalities, analytical methods to be implemented in the tool incorporate basic scientific options as well as, just to give an example, the choice to use agent-based simulation models underpins a generative and micro foundational approach to the study social phenomena. Scientific perspectives, research findings and methods coevolve with research instruments. Legal scholars engaged in the design of computational tools will have therefore to carefully dwell on the impact of the new heuristics on their research questions, conceptual categories, methods of study and work.

Our third consideration touches on methodological issues. In social science, the idea of overcoming what has been called the *war of paradigms* [42], has gradually led to the emergence of a pluralist perspective [43]; [44] according to which the integration of different research methods [45] is crucial to enhance the understanding of social phenomena. This is turning to be real not only in the more traditional areas of social research, but also in emerging fields like computational social science [41] and computational legal research where the merger of heterogeneous research methods spanning from data mining to social simulation or network analysis is ever more frequent. Thanks to the technical sophistication of the tools, and to the high levels of interoperability between applications achieved by web technologies, analytical platforms represent today the ideal place for the integration of different research methodologies.

Last remark, somehow extending previous points, deals with the potential role of analytical platforms in promoting the adoption of more interdisciplinary approaches to research, an issue that appears to be growingly topical also in the legal field (see, for an overview, [46]). In recent years [47] interdisciplinarity has been given increasing attention being seen not only as a scientific option, but also as an obligatory step to manage complex and pressing real world issues that "cannot be adequately addressed by people from just one discipline" [48]. The statement fits very well our case: giving an answer to many questions of legal science and practice – assess the impact of legal norms; understand the deep nature of legal systems; predict the evolution of law enforcement strategies – is a complex task involving scientific knowledge that transcend the boundaries of traditional legal scholarship. Our ability to integrate in new ways different knowledge and disciplines becomes therefore crucial and analytical platforms can play a relevant role to this end. Similarly to the computer-based artefacts conjured up by Parisi in [49], they provide scientist belonging to different research areas with powerful toolkits to develop integrated and non-disciplinary analyses of complex phenomena.

In an highly-cited paper dwelling on the lack of scientificity of legal scholarship [50], Richard Posner advocated a new approach to the study of law using the methods of scientific inquiry to more deeply understand legal systems. Taking as reference "the prestige and authority of scientific and other exact modes of inquiry in general" – among which he explicitly mentions those coming from computer science – Posner calls for a more prominent role of science in legal world and for a more interdisciplinary research attitude seen as an essential condition for the "understanding and improvement of the legal system". The change will probably take time and a gradual process of cross-fertilisation allowing to modify theoretical and methodological constructs entrenched over time. Online analytical platforms are certainly not the only means to trigger this transition, but will for sure play a role in finding new ways to generate legal knowledge in the Big data era. What it takes is the capacity to creatively merge different perspectives

and also the daring to concretely experiment new scientific and methodological solutions, as tough as that might be. The issue is not to believe in utopias, but to build prototypes.

Acknowledgements

Ideas proposed in this work largely stem from a longstanding interdisciplinary collaboration with the colleagues and friends of the Department of Computer Science of the University of Salerno. The Author wishes to thank Delfina Malandrino, Rocco Zaccagnino, Alfonso Guarino, Pietro Russo, Antonio Altamura, Armando Faggiano, Luca Vicidomini. Author is also truly grateful to Margherita Vestoso for the insightful comments on the content and the title of this work as well as for the support given in proofreading the paper.

References

[1] Bratton, B. H. (2016). *The Stack: On Software and Sovereignty*. MIT Press.
[2] Humphreys, P. (2004). *Extending Ourselves: Computational Science, Empiricism, and Scientific Method*. Oxford University Press.
[3] Evans, J. & Rzhetsky, A. (2010). Machine Science. *Science, 329*(5990), 399-400.
[4] Reed, D. A., Bajcsy, R., Fernandez, M. A. et al. (2005). *Computational Science: Ensuring America's Competitiveness*. President's Information Technology Advisory Committee Arlington VA.
[5] Hey, T., Tansley, S. & Tolle, K. M. (2009). *The Fourth Paradigm: Data-intensive Scientific Discovery*. Microsoft research, Vol. 1.
[6] Boulton, G., Campbell, P., Collins et al. (2012). Science As an Open Enterprise. *The Royal Society.*
[7] Hey, T. & Trefethen, A. (2003). The Data Deluge: An e-Science Perspective. In Berman, F., Fox, G. C. & Hey, A. J. G. (Eds.). *Grid Computing: Making the Global Infrastructure a Reality*. Wiley and Sons, 809-824.
[8] Hine, C. (Ed.) (2006). *New Infrastructures for Knowledge Production: Understanding e-Science*. IGI Global.
[9] Anderson, C. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired Magazine, 16*(7), http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory.
[10] Venkatraman, V. (2013). When All Science Becomes Data Science. *Science*, https://www.sciencemag.org/careers/2013/05/when-all-science-becomes-data-science.
[11] Kitchin, R. (2014). Big Data, New Epistemologies and Paradigm Shifts. *Big data & society, 1*(1).
[12] Humphreys, P. (2009). The Philosophical Novelty of Computer Simulation Methods. *Synthese, 169*(3), 615-626.
[13] Winsberg, E. (2010). *Science in the Age of Computer Simulation*. University of Chicago Press.
[14] Crouzier, T. (2017), *Digital Tools for Researchers*, http://connectedresearchers.com/online-tools-forresearchers/.
[15] Katz, D. M., Bommarito II, M. J. & Blackman, J. (2017). A General Approach for Predicting the Behavior of the Supreme Court of the United States. *PloS one, 12*(4).
[16] Katz, D. M. & Bommarito, M. J. (2014). Measuring the Complexity of the Law: The United States Code. *Artificial Intelligence and Law, 22*(4), 337-374.
[17] Ashley, K. D. (2017). *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*. Cambridge University Press.
[18] Cane, P. & Kritzer, H. M. (2012). *The Oxford Handbook of Empirical Legal Research.* Oxford University Press.
[19] Smits, J. M. (2012). *The Mind and Method of the Legal Academic*. Edward Elgar Publishing.
[20] Leeuw, F. L. & Schmeets, H. (2016). *Empirical Legal Research: A Guidance Book for Lawyers, Legislators and Regulators*. Edward Elgar Publishing.
[21] Faro, S. & Lettieri, N. (2013). *Law and Computational Social Science*. Edizioni Scientifiche Italiane.

[22] Ruhl, J., Katz, D. M. & Bommarito, M. J. (2017). Harnessing Legal Complexity. *Science, 355*(6332), 1377-1378.
[23] Russell, H. & Susskind, R. (2013). Tomorrow's Lawyers: An Introduction to Your Future. *Legal Information Management, 13*(4), 287-288.
[24] Berk, R. A., Sorenson, S. B. & Barnes, G. (2016). Forecasting Domestic Violence: A Machine Learning Approach To Help Inform Arraignment Decisions. *Journal of Empirical Legal Studies, 13*(1), 94-115.
[25] Lee, K. J., Azyndar, S. & Mattson, I. A. (2015). A New Era: Integrating Today's Next Gen Research Tools Ravel and Casetext in the Law School Classroom. *Rutgers Computer & Technology Law Journal, 41*, 31.
[26] Kuppevelt, D. & van Dijck, G. (2017). Answering Legal Research Questions About Dutch Case Law with Network Analysis and Visualization. *Legal Knowledge and Information Systems, 95*.
[27] Ourednik, A., Nellen, S. & Fleer, P. (2016). *A Visual Approach to the History of Swiss Federal Law*, http://www.dhd2016.de/abstracts/vortr%C3%A4ge-047.html
[28] Xu, J. J. & Chen, H. (2005). Crimenet Explorer: A Framework for Criminal Network Knowledge Discovery. *ACM Transactions on Information Systems, 23*(2), 201-226.
[29] Hutchins, C. E. & Benham-Hutchins, M. (2010). Hiding in Plain Sight: Criminal Network Analysis. *Computational and Mathematical Organization Theory, 16*(1), 89-111.
[30] Ferrara, E., Meo, P. D., Catanese, S. & Fiumara, G. (2014). Detecting Criminal Organizations in Mobile Phone Networks. *Expert Systems with Applications, 41*(13), 5733-5750.
[31] Didimo, W., Liotta, G. & Montecchiani, F. (2014). Network Visualization for Financial Crime Detection. *Journal of Visual Languages and Computing, 25*(4), 433-451.
[32] Taha, K. & Yoo, P. D. (2016). SIIMCO: A Forensic Investigation Tool for Identifying the Influential Members of a Criminal Organization. *IEEE Transactions on Information Forensics and Security, 11*(4), 811-822.
[33] Lettieri, N., Altamura, A. & Malandrino, D. (2017). The Legal Macroscope: Experimenting with Visual Legal Analytics. *Information Visualization, 16*(4), 332-345.
[34] Lettieri, N., Altamura, A., Faggiano, A. & Malandrino, D. (2016). A Computational Approach for the Experimental Study of EU Case Law: Analysis and Implementation. *Social Network Analysis and Mining, 6*(1), 56.
[35] Lettieri, N., Malandrino, D. & Vicidomini, L. (2017). By Investigation, I Mean Computation. *Trends in Organized Crime, 20*(1-2), 31-54.
[36] Lettieri, N., Altamura, A., Malandrino, D. & Punzo, V. (2017). Agents Shaping Networks Shaping Agents: Integrating Social Network Analysis and Agent-Based Modeling in Computational Crime Research. In Oliveira, E., Gama, J., Vale, Z., Cardoso, E. L. (Eds.). *Progress in Artificial Intelligence. Proceedings of the 18th EPIA Conference on Artificial Intelligence*. Springer, 15-27.
[37] Lathrop, D. & Ruma, L. (2010). *Open Government: Collaboration, Transparency, and Participation in Practice*. O'Reilly Media.
[38] Einav, L. & Levin, J. (2014). The Data Revolution and Economic Analysis. *Innovation Policy and the Economy, 14*(1), 1-24.
[39] Witten, I. H., Frank, E., Hall, M. A. & Pal, C. J. (2016). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann.
[40] Novaković, J. (2016). Toward Optimal Feature Selection Using Ranking Methods and Classification Algorithms. *Yugoslav Journal of Operations Research, 21*(1).
[41] Cioffi-Revilla, C. (2014). *Introduction to Computational Social Science*. Springer.
[42] Eckstein, H. (1998). Unfinished Business: Reflections on the Scope of Comparative Politics. *Comparative Political Studies, 31*, 505-534.
[43] Della Porta, D. & Keating, M. (2008). *Approaches and Methodologies in the Social Sciences: A Pluralist Perspective*. Cambridge University Press.
[44] Sil, R. & Katzenstein, P.J. (2010). *Beyond Paradigms: Analytic Eclecticism in the Study of World Politics*. Palgrave Macmillan.
[45] Teddlie, C. & Tashakkori, A. (2009). *Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences*. Sage.
[46] Siems, M. M. (2009). The Taxonomy of Interdisciplinary Legal Research: Finding the Way Out of the Desert. *Journal of Commonwealth Law and Legal Education, 7*(1), 5-17.
[47] Frodeman, R., Klein, J. T. & Pacheco, R. C. D. S. (2017). *The Oxford Handbook of Interdisciplinarity*. Oxford University Press.

[48] Nature (2015). *Mind Meld: Interdisciplinary Science Must Break Down Barriers Between Fields To Build Common Ground*, https://www.nature.com/news/mind-meld-1.18353.
[49] Parisi, D. (2014). *Future Robots: Towards a Robotic Science of Human Beings*. John Benjamins Publishing Company.
[50] Posner, R. A. (1986). The Decline of Law As an Autonomous Discipline: 1962-1987. *Harvard Law Review, 100*, 761.