

Article

The GDPR beyond Privacy: Data-Driven Challenges for Social Scientists, Legislators and Policy-Makers

Margherita Vestoso ^{1,2}

¹ ‘Scienza Nuova’ Research Centre, Suor Orsola Benincasa University, 80135 Naples, Italy; m.vestoso@gmail.com

² Department of Law, Economics, Management, Quantitative Methods, University of Sannio, 82100 Benevento BN, Italy

Received: 8 April 2018; Accepted: 4 July 2018; Published: 6 July 2018



Abstract: While securing personal data from privacy violations, the new General Data Protection Regulation (GDPR) explicitly challenges policymakers to exploit evidence from social data-mining in order to build better policies. Against this backdrop, two issues become relevant: the impact of Big Data on social research, and the potential intersection between social data mining, rulemaking and policy modelling. The work aims at contributing to the reflection on some of the implications of the ‘knowledge-based’ policy recommended by the GDPR. The paper is thus split into two parts: the first describes the data-driven evolution of social sciences, raising methodological and epistemological issues; the second focuses on the interplay between data-driven social research, rule-making and policy modelling, in the light of the policy model fostered by GDPR. Some theoretical reflections about the role of evidence in rule-making will be considered to introduce a discussion on the intersection between data-driven social research and policy modelling and to sketch hypotheses on its future evolutions.

Keywords: big data; General Data Protection Regulation (GDPR); law; data-driven social science; evidence-based decisions; rule-making; policy modelling

1. Big Data in the General Data Protection Regulation (GDPR): From Risk to Resource

Big Data is in the spotlight as the new General Data Protection Regulation (GDPR) [1] entered into force. The fast rise of the data in so many different contexts has been bringing about significant changes in science and society, paving the way to issues and opportunities which still need to be clearly identified [2]. However, the discussion on risks, benefits and application fields of Big Data is often biased by an ambiguous use of the expression, which fosters an overlap between aspects related to the collection, storing and processing of data [3]. Social research reflections on the topic, for example, are mostly focused on the benefits of data analysis, questioning the knowledge arising from the application of data-mining methods. From the legal perspective, instead, more attention is paid to the risks related to data gathering, with questions “about how we protect our privacy and other values in a world where data collection is increasingly ubiquitous” [4].

This misalignment makes a cross-fertilization between law and data-driven social sciences difficult, causing disorientation and limiting the opportunities to benefit from data, especially in the management of global social challenges [5]. Against this backdrop, the new GDPR seems to foster a reversal of course. While securing personal data from privacy violations, the regulation, in fact, challenges policymakers to exploit data-driven social research to provide knowledge-based policies. Actually, the opportunity to draw benefits from Big Data has been already taken into account by several international agendas, which suggest governments to rely on data analysis in order to improve economic growth and public well-being [6,7]. With the GDPR entering into effect, however,

the opportunity to exploit knowledge arising from Big Data turns into a rule for member states, with an explicit push to consider data as a resource for the design of public policies. A feature as interesting as it is little considered so far in regulation analyses, which are often focused on the protection of personal data.

Without a doubt, data protection remains the main goal of GDPR, which innovates the previous regulatory framework with a set of rigorous constraints aimed at protecting citizens from any kind of abusive storage and processing of personal data, including automated decision-making and profiling practices. In this vein, it can be read the Recital 71 which says: “The data subject should have the right not to be subject to a decision, which may include a measure, evaluating personal aspects relating to him or her which is based solely on automated processing and which produces legal effects concerning him or her or similarly significantly affects him or her”. The regulation aims at defining, to some extent, an ontology of the legal risks related to Big Data, which include not only privacy violations deriving from unauthorized collection of data, but also threats to civil liberties caused by arbitrary and discriminatory uses of data analysis tools—e.g., cases of data “secondary use” [8] or control creep [9].

On the other hand, the GDPR seems to enhance the opportunities arising from data-driven social sciences, calling for an interaction between open data, social research and policy decisions. Recital 157, in fact, underlines how massive data—specifically health-related data—can represent an ally in social research, providing information about long-term correlations in social conditions, such as unemployment and education with other life conditions. As it claims, “Research results obtained through registries provide solid, high-quality knowledge which can provide the basis for the formulation and implementation of knowledge-based policy, improve the quality of life for a number of people and improve the efficiency of social services” [1]. The claim should be coupled with Article 89 of the GDPR: on the one hand, the norm extends regulation safeguards to the scientific exploitation of data; on the other, it requires the member states to set out specific derogations to those rules, specifically with regard to the restriction of processing [10], when data is used for scientific purposes or in the public interest.

Without claiming to be complete, the work aims at contributing to the reflection on the implications of GDPR statements on evidence-based policy modelling, starting from a consideration on the tricky interaction between Big Data and social sciences. The regulation requires policymakers, indeed, not simply to exploit data, but to use data-driven social research as support for better policies. The paper is thus split into two parts: the first, Section 2, discusses the impact of Big Data on the social sciences, exploring methodological and epistemological issues; the second, Section 3, focuses on the impact of data-driven social research on evidence-based policy design, in the light of the ‘knowledge-based’ model fostered by GDPR. In particular, starting from some theoretical reflections on the role of empirical evidence in rule-making, the section analyses how data-driven research today challenges policy modelling. Thus, the concluding section will account for some possible evolutions the integration between data, social research, and policy-making which could contribute to better deal with the GDPR call for knowledge-based policy.

2. Social Science, Data and Their Tricky Interaction

The “Datafication”, which refers to the possibility to convert real-life information in computer data, has fostered in recent years a relevant shift in computer processing of social reality events, transforming almost any phenomenon “in a quantified format so it can be tabulated and analyzed” [11]. The continuous growth in computational power and the exploration of new markup languages, able to encode human-generated data in computer data, not only have heightened the range of available information but allowed new kinds of analysis to be performed. This has extended the research fields interested in using data mining and machine-learning techniques [12], with a significant impact on the methods and times of evolution of science. As pointed out by Jim Gray, one of the main Microsoft researchers [13,14], data science and computing are radically challenging the scientific method setting-up a “fourth paradigm” [14] in science, even beyond simulations.

Although the change involves all research areas, social sciences seem to be the most affected. Some of the reasons have been expressed in an influential paper [15] which has emphasized how the increasing availability of social data and analytical tools is steering social sciences toward a new computational dimension and to a different way of building the scientific knowledge. By shaping individual and social behaviours and keeping track of them, Big Data and related technologies significantly challenge traditional social research, fostering a more scientific approach to the study of behavioural patterns. As argued in [16], “the advent of the Big Data movement and the increasing convergence between data platforms in various domains of social life (e.g., the public, private, and social sectors) could allow sociologists to have fine-grained, large-scale data not only on individual choices but also on social network connections that were impossible even to contemplate before”.

We can find already in Bourdieu the idea of exploring correlations between some aspects of individuals’ lives, such as cultural conditions, and their social structures. His theory of ‘habitus’, indeed, endows individuals with “*an active residue or sediment of past experiences which functions within their present, shaping perception, thought and action and thereby shaping social practice in a regular way*” [17]. Even more interesting is Bourdieu’s idea to express such regularities in mathematical terms [18], with correspondence analysis (CA) and multiple correspondence analysis (MCA) used to obtain a geometric modelling of data [19]. The quantitative method adopted by Bourdieu not only is a criticism of most of the statistical methods used in social sciences but also provides a new frame-model for social reality representation, paving the way to many of the recent Big Data analysis methods.

As the development of new computer capacities has made social sciences closer to the scientific perspective above described, fostering a convergence between the computational paradigm and the social research [20], Big Data analysis has also led social scientists to undertake epistemological reflections, rethinking both purposes and methods of their research [21,22]. The discussion is laid out by the Big Data analysis effect calling into question the deductive method, changing the traditional testing role of the data and so the way in which social theories are typically produced. This emerges from conclusions expressed by Chris Anderson in his work “The End of Theory” [23], which describes the advent of Big Data as a turning point for science leading to the demolition of the traditional theory-making process. As he argues, “Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all”. A claim in line with the scenario imagined by Google’s research director, Peter Norvig, which considers all of science as usurped by exploratory, unsupervised clustering and data analysis [24].

According to these perspectives, advances in data analytics techniques suggest the need to review empirical research from the basis, with data no longer used to validate previously formulated hypotheses, but provided with a generative power in the theories’ formulation process. As highlighted by Steadman, indeed, “algorithms will spot patterns and generate theories, so there’s a decreasing need to worry about inventing a hypothesis first and then testing it with a sample of data” [25]. The result is that the causal analysis of social phenomena, driven by research questions embedded in an explanatory hypothesis, is converted into a pure data-mining activity capable of both identifying insightful patterns between behavioural data and forecasting how those behaviours could evolve in the future [9].

However, this idea is not free from criticisms, with researchers suggesting that explorations only based on data could not provide a scientific understanding of society [26]. Some critical considerations on this topic have been expressed, for example, by the data expert Rob Kitchin, Professor at Maynooth University Social Sciences Institute (MUSSI). In a work titled *Data Revolution* [9], he outlines four relevant statements about Big Data to be worth considering and questioning: (i) the possibility for Big Data to take into account all the features of a domain and to return in complete resolution; (ii) the idea that Big Data analytics can make useless the a priori activity of theorising, modelling and hypothesis; (iii) the belief that data can speak for itself without human bias and that any correlation identified is always relevant and useful; and (iv) the idea that data meaning could be apart from contextual knowledge.

According to Kitchin, the problem is to consider data as if they could speak for themselves [9]. This is an effect of the “deification of data” [27], namely the idea that correlations in the datasets could be inherently meaningful, without the need to embed them in a model or a theory that could explain their causal attitude [9], an assumption that can threaten the scientific result of research. Indeed, as underlined in [28], data is “always, inherently, speaking from a particular position”, so unexpected patterns could be the effect of random associations without any causal attitude [28]. In order to cope with the risk of overestimating knowledge arising from data analysis, Kitchin suggests including correlations into a wider validation process. In particular, he points out that “relationships should form the basis for hypotheses that are more widely tested, which in turn are used to build and refine theory that explains them” [9].

The need for coupling data with explanatory models reflects the research perspective fostered by scientists in the area of computational social science (hereafter CSS) [21,29]. Supported by a wide literature (see e.g., [15,20,21,29–34]), this scientific paradigm considers the interplay between social sciences and other research areas, like complexity sciences, physics, and computer science, as an opportunity for social research to grow in a quantitative dimension and to build theoretical and methodological solutions aiming to better understand the dynamics of global social phenomena. Considering these challenges, CSS researchers suggest combining different computational methods in order to increase the explanatory power and the scientific basis of social theories [35]. On the other hand, as highlighted by Conte et al. [36], the quantitative variant of CSS rooted in data statistical analysis is unable to explain mechanisms underlying social complexity and, thus, should not be considered as independent from a modelling need.

In this view, massive data can be seen as the basis for building theory-driven computational models, designed to have an explanatory role with regard to social life dynamics [24]. A recent work on the topic [29] in fact stresses how, “Computational social science can benefit from more broadly mining in the rich depths of social science’s methodological, theoretical, and conceptual imagination”, and thus “[...] from connecting reflectively to the diverse goals, metaphysics, theories, and topics that have enriched conventional social science to date”. A rational use of Big Data in social research thus requires combining it with other methods in a framework matching quantitative and qualitative approaches [37]. Some examples of integration presented in [38–42] allow in fact an understanding of how data’s possibility to enable sophisticated, finer-grained analyses of individual and collective behaviours is often strengthened by the link with theoretical models designed for enlightening the dynamics that lead the emergence of the social patterns.

The discussion so far sketched has implications also for the legal and policy fields, where Big Data has already started to play a role. If the benefits of data mining highlight the importance of drawing insights from evidence-based social research, the risks related to a ‘deification of data’ suggest reflecting on how to use data in support of policy design. In particular, it leads to think about the consequences of a naïve use of Big Data analysis in designing rules and public policies, especially when dealing with complex social challenges.

3. Data-Driven Challenges for Rule-Making and Policy Modelling

By displaying new ways for investigating individual and collective behaviours, data-driven evolution of social research is challenging not only social sciences, but also rule-making and policy design [22]. A number of organizations have suggested in recent years to exploit advanced Big Data related methods, such as deep learning, neural networks or sentiment analysis, to improve policy modelling, stressing the importance to endow public policies with a stronger evidence basis [43]. Governments (local, state, and national) have been incited both to rely on Big Data in policy decisions [6] and to increase the amount of open data to be processed for scientific purposes or in the public interest [44]. The GDPR, for its part, has contributed to make such a goal an official task for European policymakers, fostering in particular the integration between the knowledge arising from data-driven social research and the design of public policies and services.

The idea to formalize the use of social research in policy modelling is not new. Already in 1984, Ruttan [45] underlined that the institutionalization of research in the social sciences would have made it possible for policymakers to use social science knowledge and analytical skills in place of the expensive process of learning by trial and error. His statement is grounded on a long tradition of theoretical reflections suggesting the need to bridge law and rule-making to empirically ground social investigations. A brief overview of these reflections will be presented in the next subsection, sketching a link with the insights from the recent research area of computational social sciences. This can be seen as an introduction to Section 3.2, which discusses, in a more targeted way, how empirical evidence from data-driven social research can support policy-making, in the light of GDPR request for knowledge-based decisions.

3.1. The Role of Evidence in Rule-Making

The discussion on the relationship between empirical social research and rulemaking relies on historical and philosophical considerations that have been going on for centuries [46]. Already Hume, by sharing Vico's idea that "individual actions are shaped by their social environment which, in turn, they help alter" [47], outlined the need to explore the interplay between social behaviours and the evolution of legal systems with an approach inspired to the natural sciences [47]. A perspective that emerges also in Montesquieu's work [48] and in his conception of legal rules as "necessary relationships that derive from the nature of things" [48], namely from the social facts, that he considered able to affect legislator's decisions. This is what he describes as the 'Spirit of law', the idea that the effectiveness of norms depends on rule-maker's capability to make them suitable for the spirit of the country, the religious sentiment as well as the commercial practices in use, considering the habits and culture of the people that will have to comply with those norms [49,50].

One of the firsts attempts in this vein was made by Alexis de Tocqueville, who used the comparison between historical facts to empirically explore the social factors related to choice for a given constitutional model. The results of the study were reported in *De la démocratie en Amérique* [51], where information gathered on-the-field about the social, economic, and political conditions of American society were used in support of a discussion on the possible evolution of the American democratic model. It is thanks to Eugen Ehrlich, however, if the need for empirical analysis of social phenomena has received an explicit recognition by the legal science. Considered as one of the fathers of the sociology of law [52], along with Emile Durkheim and Max Weber, Ehrlich suggests indeed the development of a very sociological-juridical science [53] that frames society as the real reference point of law. In this vein, he defines the scientific study of law as an activity that necessarily requires the analysis of social relations and so is linked to social science empirical methods [53].

The idea of a necessary interaction between law, rulemaking, and empirical social research has flowed into the current projects connected to the law-and-society movement [54]. An empiricist ferment in fact is still underway, as witnessed by two relatively recent movements: the empirical legal studies (ELS) and the new legal realism (NLR) [55]. A canonical definition of their activity does not exist, but the two movements seem to share the idea of using the empirical social sciences to improve legal thinking and rule-making, while contributing to explaining to policymakers, and society as a whole how the legal system works [56]. In the area of ELS, this goes through applying rigorous empirical methods to law-related questions, while researchers in the NLR field stress the importance of relying on the methodological eclecticism, taking care of translations problems that derive from connecting theory and observation, as well as law and the social sciences.

Reflections so far considered suggest seeing the scientific investigation of social dynamics as a strategic knowledge in assessing rules and policies' effectiveness, encouraging legal science to move from a traditional dogmatic perspective to an empirical one [57] that includes an evidence-based analysis of society and its complex phenomena. The discussion is ongoing: the number of empirical studies on legal rules, institutions, and behavioural systems is growing, as suggested by the Annual Review of Law and Social Science, along with researchers from social, legal and political sciences

that stress the need for deepening the non-linear mechanisms of social systems in order to tackle the complex challenges of society [58–60].

On the other hand, as highlighted by a recent work [61], the design of suitable and effective regulatory solutions depends on a preliminary assessment of the dynamics shaping individuals' and groups' behaviour and preferences, as it allows formal institutions to identify the way for realizing collectively desirable outcomes. The use of data analysis and of other computational methods, such as agent-based simulations or social network analysis, can play a relevant role against this backdrop, providing a new slant to the empirical exploration of society. This allows, indeed, the identification of social problems otherwise impossible to catch while extending the chances for understanding collective and individual dynamics involved in the emergence and evolution of social phenomena. This can improve, in turn, the search for more effective policy solutions, able to shape social expectations according to the general well-being [22,31,59].

This perspective seems in line with GDPR's call for integrating data-driven social research into policy modelling. In order to verify how this can lead to a knowledge-based policy solutions, however, it seems interesting to link the methodological and theoretical considerations so far sketched with current experiences of integration between data and policy modelling, taking into account benefits and risks of such an integration. Understanding how data can enhance the empirical basis of rule-making and policy modelling processes requires, in fact, to consider the following question: to what extent data-driven social research can support the development of "knowledge-based" policies?

3.2. Data-Driven 'Knowledge-Based' Policies

A great deal has been written about the opportunity for data mining to support government decisions (see e.g., [6,11,62–64]). The idea often stressed is that Big Data could provide a sort of lens allowing phenomena otherwise impossible to detect or forecast by a human observer to surface [65]. The application of machine-learning techniques to large amounts of data, arising from car traffic, crime reports, school assessments, medical records and many other kinds of digital tracks, allow policy-makers to identify unexpected correlations, which can be useful in forecasting future scenarios. However, along with an interest in Big Data opportunities, a trend has also spread of enhancing the self-explanatory capacity of the data [27]. Policy-makers have been encouraged to rely on empirical evidence that turned out to be vague, as they overestimated the meaning of data as well as their ability to provide high-quality knowledge [66]. Forms of opaque data mining contributed, therefore, to discriminatory practices and put democratic principles at risk [66].

The discussion does not put into question the use of Big Data overall. It is likely to become unavoidable in the evolution of policy making to relying on data-driven models to design rules and public policies [64]. However, some of the issues reported above suggest not losing sight of the goals that lead to the use of Big Data in the public context, according to the collective outcomes to be achieved [5]. Relying on incorrect empirical premises can in fact foster the adoption of ineffective or even counterproductive policies. For instance, consider a real data study published in the *American Economic Review* in 2006 [67]. The work reports the effects of a policy that had increased taxes for smokers in order to lower cigarette consumption and to incentivize a healthier lifestyle. The researchers found out that the policy had produced, along with a minimal reduction in sales, a smokers' propensity to make their experience with cigarettes even more intense and dangerous, spurring them to opt for higher-tar brands or to use cigarettes without the filter [67].

Understanding how Big Data analysis can support policymakers thus requires a reflection on the kind of knowledge arising from it. The development of effective rules and policies entail exploring the dynamics of the phenomenon to be assessed and identifying the elements that can foster a collective behavioural change, by deciding whether a policy will work or not [61]. The analysis of massive social data represents an opportunity against this backdrop, but seems unable to provide such knowledge by itself. Policymakers should take into account this issue, considering whether evidence used to support

the policy modelling is able to provide a right knowledge with regard to the problem at hand, namely if it sheds light on the main factors involved in the phenomenon, and on how to interplay of them [68].

In a number of experiences, data-mining tools have been used with a naive approach to support policy decisions, without real understanding of the ways in which they process reality [66]. As pointed out by O’Neil [66], one of the problems with using Big Data in policy decisions is that code errors are not borne by those that design the algorithms, but by everyone else. On the other hand, even when data are available in large amounts and can be processed by means of innovative techniques, predictions led by data correlations risk missing the chance of explaining the causal dynamics of a phenomenon, increasing the possibility of making rough errors [69]. As argued by Conte and Paolucci [36], “[...] when the nature of behavior matters, accurate statistical analyses of social dynamics can maybe reach predictive power but cannot fully explain what is going on”.

These issues underline the need to rethink the current approach to Big Data in rule-making and policy modelling. There is little doubt about the ability of high-quality data mining to set out patterns in social behaviour which can steer policymakers toward new, real, existing problems. The exploration of real-time large data sets can provide a measure of the impact of a phenomenon [70], as well as allow drawing insights on its causal factors, or to make predictions on its evolution [71]. This can, in turn, be useful in assessing whether a policy should be introduced to achieve a collective outcome. However, the development of an effective policy depends not only on the definition of the expected outcome but also on understanding how the outcome could be achieved. This calls for exploring the reasons leading the emergence of behavioural trends, namely the causal factors that the policy decision has to focus on. This should be the inherent content of a public policy and, thus, the undeniable underpinning of a ‘knowledge-based’ policy model.

It is worthwhile remarking that the “high-quality knowledge” the GDPR suggests as a basis to develop better policies is represented not by Big Data, but rather by social studies grounded on data. In this sense, Big Data can represent a useful resource, especially if coupled with a theoretical framework and integrated with other computational approaches allowing for a better assessment of the complexity of social dynamics.

4. Conclusions

This work aimed to contribute to the reflection on some of the implications of the ‘knowledge-based’ policy model recommended by the GDPR, discussing the benefits and risks of using data in support of rule-making and policy-modelling. The main goal of the analysis was to shed light on the kind of knowledge provided by new Big Data analysis opportunities, specifically with regard to the possibility of understanding social dynamics. As suggested by the Recital 157 of the GDPR, indeed, data can provide useful, high-quality knowledge to improve policies. By means of it, both social scientists and governments are now able to gain information unattainable before and extend their ability to understand the world and fix its problems. In this vein, it is important for rule-makers to combine the protection of citizens’ privacy and their right to no discrimination with the need to make data open for scientific purposes. In particular, member states’ legislators should define as soon as possible the kind of derogations that Article 89 allows for data collected for research and historical purposes, as well as in the public interest.

On the other hand, the methodological issues discussed above, along with reflections on some current approaches to data-driven policy-modelling, lead one to believe that there is a risk of abusive appeals to data-mining techniques that occur by using “the practice of social mining (a) in a fragmentary, rather than integrated, way; (b) in a commercial/speculative-oriented rather than governance-oriented context; and (c) with a focus on short-term forecasting rather than on policy modeling” [31]. Several researchers in the CSS area suggest coping with this risk by combining data with different methods, deepening the range of possibilities offered by computational development. Results produced by sophisticated data analytics tools, for examples, can be embedded in large-scale simulation models, which aim to explain from the bottom-up the complex dynamics of a social

phenomenon and, eventually, to assess how such dynamics could react to a change, like the introduction of a new policy [72,73]. Similar integrated approaches can make new sense of data, allowing more robust knowledge to be drawn from it and also for it to be involved in a frame of long-term policy decisions [31,74,75].

Big Data can act as a glue between social research and the design of more informed policy decisions, but this depends on the choice of seeing data-mining as part of both a wider methodological framework and a theoretically-grounded research perspective. Some reliable predictions achieved through sophisticated data analysis techniques have been shown to be useful for some regulatory purposes, but they miss the chance to provide a resolution when complex social phenomena are involved. In such cases, they are unlikely to build by themselves scientific knowledge [31], while providing an inadequate response to regulation's call for a knowledge-based policy.

If "Science can now effectively be brought to bear on public policy-making" [76], as the political scientist Margaret Levi pointed out, a reflection is needed on how data can lead to scientific knowledge. Considerations sketched out so far suggest deepening the goals of the GDPR from this point of view. The need for knowledge-based decisions should encourage policy-makers to look also beyond Big Data, avoiding the option for a specific method prevailing over the need to understand the phenomenon to be regulated. Policy-makers should, therefore, strengthen the dialogue with social researchers, both to discuss problems with them and to make a more informed assessment of the evidence on which to base their decisions [43]. This is a challenge that seems to be fostered by the notion of 'knowledge-based policy' reported in the GDPR and that looks at Big Data as one phase of the more complex path leading to an improved quality and efficiency of public policies and social services.

Acknowledgments: The author is truly thankful to Nicola Lettieri for his careful assistance throughout the research path. His inspired discussions, occurring during the course of "Law and Computational Social Sciences" he held at the University of Sannio, have stimulated the reflection on the topics covered by this work while making him an incomparable guide.

Conflicts of Interest: The author declares no conflict of interest.

References

1. General Data Protection Regulation. Regulation of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC. *Off. Eur. Union* **2016**, *59*, 294.
2. Ward, J.S.; Barker, A. Undefined by data: A survey of big data definitions. *arXiv*, **2013**, arXiv:1309.5821.
3. Kitchin, R.; Dodge, M. *Code/Space: Software and Everyday Life*; MIT Press: Cambridge, MA, USA, 2011.
4. Podesta, J. Findings of the big data and privacy working group review. *The White House Blog*, 1 May 2014.
5. Sonal, S.; Russon Gilman, H. Making Better Use of Big Data. *Stanford Social Innovation Review*, 17 November 2015.
6. Kim, G.H.; Trimi, S.; Chung, J.H. Big-data applications in the government sector. *Commun. ACM* **2014**, *57*, 78–85. [CrossRef]
7. Habl, C.; Renner, A.T.; Bobek, J.; Laschkolnig, A. Study on Big Data in Public Health, Telemedicine and Healthcare. Final Report. 2016. Available online: <https://www.cabdirect.org/?target=%2fcabdirect%2fabstract%2f20173162483> (accessed on 10 June 2018).
8. Bollier, D.; Firestone, C.M. *The Promise and Peril of Big Data*; Aspen Institute, Communications and Society Program: Washington, DC, USA, 2010.
9. Kitchin, R. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*; Sage: Newcastle upon Tyne, UK, 2014.
10. Chassang, G. The impact of the EU general data protection regulation on scientific research. *Ecancermedicalscience* **2017**, *11*, 709. [CrossRef] [PubMed]
11. Cukier, K.; Mayer-Schoenberger, V. The rise of big data: How it's changing the way we think about the world. *Foreign Aff.* **2013**, *92*, 28.
12. Bello-Orgaz, G.; Jung, J.J.; Camacho, D. Social big data: Recent achievements and new challenges. *Inf. Fusion* **2016**, *28*, 45–59. [CrossRef]

13. Hey, T.; Tansley, S.; Tolle, K.M. *The Fourth Paradigm: Data-Intensive Scientific Discovery*; Microsoft Research: Redmond, WA, USA, 2009; Volume 1.
14. Gray, J. eScience: A Transformed Scientific Method. In *The Fourth Paradigm: Data-Intensive Scientific Discovery*; Hey, T., Tansley, S., Tolle, K.M., Eds.; Microsoft research: Redmond, WA, USA, 2009; Available online: http://research.microsoft.com/en-us/um/people/gray/talks/NRC-CSTB_eScience.ppt (accessed on 5 July 2018).
15. Lazer, D.; Pentland, A.S.; Adamic, L.; Aral, S.; Barabasi, A.L.; Brewer, D.; Christakis, N.; Contractor, N.; Fowler, J.; Gutmann, M.; et al. Life in the network: The coming age of computational social science. *Science* **2009**, *323*, 721. [[CrossRef](#)] [[PubMed](#)]
16. Bianchi, F.; Squazzoni, F. Agent-based models in sociology. *Wiley Interdiscip. Rev. Comput. Stat.* **2015**, *7*, 284–306. [[CrossRef](#)]
17. Crossley, N. The phenomenological habitus and its construction. *Theory Soc.* **2001**, *30*, 81–120. [[CrossRef](#)]
18. Bourdieu, P. *Distinction: A Social Critique of Taste*; Harvard University Press: Cambridge, MA, USA, 1984.
19. Lebaron, F. How Bourdieu “quantified” Bourdieu: The geometric modelling of data. In *Quantifying Theory: Pierre Bourdieu*; Springer: Dordrecht, The Netherlands, 2009; pp. 11–29.
20. Cioffi-Revilla, C. *Introduction to Computational Social Science*; Springer: London, UK, 2014.
21. Conte, R.; Gilbert, N.; Bonelli, G.; Cioffi-Revilla, C.; Deffuant, G.; Kertesz, J.; Loreto, V.; Moat, S.; Nadal, J.P.; Sanchez, A.; et al. Manifesto of computational social science. *Eur. Phys. Spec. Top.* **2012**, *214*, 325–346. [[CrossRef](#)]
22. Lettieri, N. Computational Social Science, the Evolution of Policy Design and Rule Making in Smart Societies. *Future Int.* **2016**, *8*, 19. [[CrossRef](#)]
23. Anderson, C. The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired Magazine*, 23 June 2008.
24. Murtagh, F. History of cluster analysis. In *The Visualization and Verbalization of Data*; CRC Press: Boca Raton, FL, USA, 2014; pp. 117–133.
25. Steadman, I. Big data and the death of the theorist. *Wired Online*, 25 January 2013.
26. Shah, D.V.; Cappella, J.N.; Neuman, W.R. Big data, digital media, and computational social science: Possibilities and perils. *Ann. Am. Acad. Political Soc. Sci.* **2015**, *659*, 6–13. [[CrossRef](#)]
27. Jenkins, T. Don’t count on big data for answers. *The Scotsman*, February 2013.
28. Crawford, K. The hidden biases in big data. *Harvard Business Review*, 1 April 2013.
29. Stuetzer, C.M.; Welker, M.; Egger, M. (Eds.) *Computational Social Science in the Age of Big Data. Concepts, Methodologies, Tools, and Applications*; Herbert von Halem-Verlag: Koln, Germany, 2018.
30. Gilbert, G.N. (Ed.) *Computational Social Science*; Sage: Newcastle upon Tyne, UK, 2010.
31. Conte, R. Computational Social and Behavioral Science. In *New Frontiers in the Study of Social Phenomena*; Cecconi, F., Ed.; Springer: Cham, Switzerland, 2016; pp. 1–7.
32. Alvarez, R.M. (Ed.) *Computational Social Science*; Cambridge University Press: Cambridge, UK, 2016.
33. Squazzoni, F. A (computational) social science perspective on societal transitions. *Comput. Math. Organ. Theory* **2008**, *14*, 266–282. [[CrossRef](#)]
34. Lettieri, N.; Faro, S. Computational social science and its potential impact upon law. *Eur. J. Law Technol.* **2012**, *3*, 3.
35. Squazzoni, F. *Agent-Based Computational Sociology*; John Wiley & Sons: Hoboken, NJ, USA, 2012.
36. Conte, R.; Paolucci, M. On agent-based modeling and computational social science. *Front. Psychol.* **2014**, *5*, 668. [[CrossRef](#)] [[PubMed](#)]
37. Vanderstoep, S.W.; Johnston, D.D. *Research Methods for Everyday Life: Blending Qualitative and Quantitative Approaches*; John Wiley & Sons: Hoboken, NJ, USA, 2008; Volume 32.
38. Scholz, R.W.; Tietje, O. *Embedded Case Study Methods: Integrating Quantitative and Qualitative Knowledge*; Sage: Newcastle upon Tyne, UK, 2002.
39. Karamshuk, D.; Shaw, F.; Brownlie, J.; Sastry, N. Bridging big data and qualitative methods in the social sciences: A case study of Twitter responses to high profile deaths by suicide. *Online Soc. Netw. Media* **2017**, *1*, 33–43. [[CrossRef](#)]
40. Halavais, A. Bigger sociological imaginations: Framing big social data theory and methods. *Inf. Commun. Soc.* **2015**, *18*, 583–594. [[CrossRef](#)]
41. Ford, H. Big Data and Small: Collaborations between ethnographers and data scientists. *Big Data Soc.* **2014**, *1*. [[CrossRef](#)]

42. Boero, R.; Squazzoni, F. Does empirical embeddedness matter? Methodological issues on agent-based models for analytical social science. *J. Artif. Soc. Soc. Simul.* **2005**, *8*, 1–6.
43. National Research Council. *Using Science as Evidence in Public Policy*; National Academies Press: Washington, DC, USA, 2012.
44. Attard, J.; Orlandi, F.; Scerri, S.; Auer, S. A systematic review of open government data initiatives. *Gov. Inf. Q.* **2015**, *32*, 399–418. [[CrossRef](#)]
45. Ruttan, V.W. Social science knowledge and institutional change. *Am. J. Agric. Econ.* **1984**, *66*, 549–559. [[CrossRef](#)]
46. Baier, M. (Ed.) *Social and Legal Norms: Towards a Socio-Legal Understanding of Normativity*; Routledge: Abingdon, UK, 2016.
47. Brown, R. *The Nature of Social Laws: Machiavelli to Mill*; Cambridge University Press: Cambridge, UK, 1986.
48. De Montesquieu, C. *Montesquieu: The Spirit of The Laws*; Cambridge University Press: Cambridge, UK, 1989.
49. Duverger, M. *The Study of Politics*; Springer Science & Business Media: Berlin, Germany, 2012.
50. Baum, J.A. *Montesquieu and Social Theory*; Elsevier: New York, NY, USA, 2013.
51. De Tocqueville, A. *Democracy in America*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 1889.
52. Bechtler, T.W. (Ed.) *Law in a Social Context: Liber Amicorum Honouring Professor Lon L. Fuller*; Kluwer: Alphen aan den Rijn, The Netherlands, 1978.
53. Ziegert, K.A. On Eugen Ehrlich, fundamental principles of the sociology of law. In *Classic Writings in Law and Society: Contemporary Comments and Criticisms*; Routledge: Abingdon, UK, 2017; p. 123.
54. Friedman, L.M. The law and society movement. *Stanf. Law Rev.* **1986**, *38*, 763–780. [[CrossRef](#)]
55. Suchman, M.C.; Mertz, E. Toward a new legal empiricism: Empirical legal studies and new legal realism. *Ann. Rev. Law Soc. Sci.* **2010**, *6*, 555–579. [[CrossRef](#)]
56. Eisenberg, T. Why do empirical legal scholarship. *San Diego Law Rev.* **2004**, *4*, 174.
57. Leeuw, F.L. Empirical Legal Research. The Gap between Facts and Values and Legal Academic Training. *Utrecht Law Rev.* **2015**, *11*, 19. [[CrossRef](#)]
58. Conte, R.; Andrighetto, G.; Campenni, M. (Eds.) *Minding Norms: Mechanisms and Dynamics of Social Order in Agent Societies*; Oxford University Press: Oxford, UK, 2013.
59. Squazzoni, F. A social science-inspired complexity policy: Beyond the mantra of incentivization. *Complexity* **2014**, *19*, 5–13. [[CrossRef](#)]
60. Lettieri, N.; Parisi, D. An evolutionary agent-based model of the interplay between punishment and damaging behaviours. *Artif. Intell. Law* **2013**, *21*, 425–453. [[CrossRef](#)]
61. Nyborg, K.; Anderies, J.M.; Dannenberg, A.; Lindahl, T.; Schill, C.; Schlüter, M. Social norms as solutions. *Science* **2016**, *354*, 42–43. [[CrossRef](#)] [[PubMed](#)]
62. Bertot, J.C.; Gorham, U.; Jaeger, P.T.; Sarin, L.C.; Choi, H. Big data, open government and e-government: Issues, policies and recommendations. *Inf. Political* **2014**, *19*, 5–16.
63. Raghupathi, W.; Raghupathi, V. Big data analytics in healthcare: Promise and potential. *Health Inf. Sci. Syst.* **2014**, *2*, 3. [[CrossRef](#)] [[PubMed](#)]
64. Höchtl, J.; Parycek, P.; Schöllhammer, R. Big data in the policy cycle: Policy decision making in the digital era. *J. Organ. Comput. Electron. Commer.* **2016**, *26*, 147–169. [[CrossRef](#)]
65. Aiden, E.; Michel, J.B. *Uncharted: Big Data as a Lens on Human Culture*; Penguin: London, UK, 2014.
66. O’Neil, C. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*; Broadway Books: New York, NY, USA, 2017.
67. Adda, J.; Cornaglia, F. Taxes, cigarette consumption, and smoking intensity. *Am. Econ. Rev.* **2006**, *96*, 1013–1028. [[CrossRef](#)] [[PubMed](#)]
68. Ostrom, E. A general framework for analyzing sustainability of social-ecological systems. *Science* **2009**, *325*, 419–422. [[CrossRef](#)] [[PubMed](#)]
69. Lazer, D.; Kennedy, R.; King, G.; Vespignani, A. The parable of Google Flu: Traps in big data analysis. *Science* **2014**, *343*, 1203–1205. [[CrossRef](#)] [[PubMed](#)]
70. Sasahara, K.; Hirata, Y.; Toyoda, M.; Kitsuregawa, M.; Aihara, K. Quantifying collective attention from tweet stream. *PLoS ONE* **2013**, *8*, e61823. [[CrossRef](#)]
71. Tsou, M.H.; Yang, J.A.; Lusher, D.; Han, S.; Spitzberg, B.; Gawron, J.M.; Gupta, D.; An, L. Mapping social activities and concepts with social media (Twitter) and web search engines (Yahoo and Bing): A case study in 2012 US Presidential Election. *Cartogr. Geogr. Inf. Sci.* **2013**, *40*, 337–348. [[CrossRef](#)]

72. Fosset, P.; Banos, A.; Beck, E.; Chardonnel, S.; Lang, C.; Marilleau, N.; Piombini, A.; Leysens, T.; Conesa, A.; Andre-Poyaud, I.; et al. Exploring intra-urban accessibility and impacts of pollution policies with an agent-based simulation platform: GaMiroD. *Systems* **2016**, *4*, 5. [[CrossRef](#)]
73. Qin, C.; Bressers, H.T.; Su, Z.B.; Jia, Y.; Wang, H. Assessing economic impacts of China's water pollution mitigation measures through a dynamic computable general equilibrium analysis. *Environ. Res. Lett.* **2011**, *6*, 044026. [[CrossRef](#)]
74. Darema, F. Dynamic data driven applications systems: A new paradigm for application simulations and measurements. In *International Conference on Computational Science*; Springer: Berlin/Heidelberg, Germany, 2004; pp. 662–669.
75. Kim, B.S.; Kang, B.G.; Choi, S.H.; Kim, T.G. Data modeling versus simulation modeling in the big data era: Case study of a greenhouse control system. *Simulation* **2017**, *93*, 579–594. [[CrossRef](#)]
76. Levi, M. Big Data and Better Government. In *Edge.org*; 2016. Available online: <https://www.edge.org/response-detail/26625> (accessed on 7 June 2018).



© 2018 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).