



Big data: big power shifts?

Lena Ulbricht

Berlin Social Science Center (WZB), Germany

Maximilian von Grafenstein

Humboldt Institute for Internet and Society (HIIG), Berlin, Germany

Published on 31 Mar 2016 | DOI: 10.14763/2016.1.406

Abstract: Facing general conceptions of the power effects of big data, this thematic edition is interested in studies that scrutinise big data and power in concrete fields of application. It brings together scholars from different disciplines who analyse the fields agriculture, education, border control and consumer policy. As will be made explicit in the following, each of the articles tells us something about firstly, what big data is and how it relates to power. They secondly also shed light on how we should shape “the big data society” and what research questions need to be answered to be able to do so.

Keywords: Big data, Power, Regulation of innovation, Big data education, Border control, Consumer protection, Agriculture

Article information

Received: 08 Mar 2016 **Reviewed:** 17 Mar 2016 **Published:** 31 Mar 2016

Licence: Creative Commons Attribution 3.0 Germany

Competing interests: The author has declared that no competing interests exist that have influenced the text.

URL: <http://policyreview.info/articles/analysis/big-data-big-power-shifts>

Citation: Ulbricht, L. & von Grafenstein, M. (2016). Big data: big power shifts?. *Internet Policy Review*, 5(1). DOI: 10.14763/2016.1.406

This Editorial is part of 'Big data: big power shifts?', a Thematic edition of the *Internet Policy Review* supported by the Vodafone Institute for Society and Communications.

PAPERS IN THIS THEMATIC EDITION

The ethics of big data in big agriculture

Isabelle M. Carbonell, *University of California, Santa Cruz*

Regulating “big data education” in Europe: lessons learned from the US

Yoni Har Carmel, *University of Haifa*

The borders, they are a-changin'! The emergence of socio-digital borders in the EU

Magdalena König, *Maastricht University*

Beyond consent: improving data protection through consumer protection lawMichiel Rhoen, *Leiden University***EDITORIAL: BIG DATA THROUGH THE POWER LENS:
MARKER FOR REGULATING INNOVATION**

Big data may be defined as the “cultural, technological, and scholarly phenomenon” made up of the interplay of algorithmic analysis of large datasets - in order to identify patterns and make economic, social, technical, and legal claims (boyd & Crawford, 2012, p. 663). Recently, big data has come under scrutiny: analyses of the rhetoric and representation of big data unveil that the term is commonly conceptualised as “a natural force to be controlled” or as “a resource to be consumed” (Puschmann & Burgess, 2014) and that it is valorised by an “aura of truth, objectivity, and accuracy” (boyd & Crawford, 2012, p. 663); such that critical scholars have identified an “insouciance” in the use of the term, considering its potential abuse by powerful interests (Portmess, Tower, 2015, p. 8).

Power should play an important role in how big data is conceptualised. Surveillance studies have, for instance, pointed out how individuals and social groups are dominated in their life chances by their “data doubles”, as constructed on the basis of their “activities, connections, performances, transactions, and movements” (Lyon, 2014, p. 6). Even those individuals who carefully protect their data are subject to control, manipulation and discrimination due to the data that others volunteer and that allows for inference on every individual (Barocas & Nissenbaum, 2014, p. 55). Used by the sovereign state, big data serves the purpose of knowing the population, the “body count on which it founds its authority” (Amoore, 2013, p. 36). But big data is not only used by state authorities; it serves a more general purpose, as Frank Pasquale (2015) recalls: “Knowledge is power. To scrutinize others while avoiding scrutiny oneself is one of the most important forms of power” (p. 3). The prominent role of large corporations in big data leads Shoshana Zuboff (2015) to assert that big data is the foundational component of surveillance capitalism (p. 75): The alliance between big data businesses and governments leads to “the sovereign power of a near future that annihilates the freedom achieved by the rule of law” (Zuboff, 2015, p. 81). It relies on the exploitation of “habitats inside and outside the human body [that] are saturated with data and produce radically distributed opportunities for observation, interpretation, communication, influence, prediction, and ultimately modification of the totality of action” (Zuboff, 2015, p. 82).

Facing these general conceptions of the power effects of big data, we were interested in studies that scrutinise big data and power in concrete fields of application. The present thematic edition brings together scholars from different disciplines who analyse the fields agriculture, education, border control and consumer policy. As will be made explicit in the following, each of the articles tells us something about firstly, what big data is and how it relates to power. They secondly also shed light on how we should shape “the big data society” and what research questions need to be answered to be able to do so.

FOUR VIEWS ON BIG DATA-INDUCED POWER SHIFTS

For *Isabelle Carbonell*, who scrutinises the role of big data for modern industrial farming, big data is a “a tool for revealing hidden patterns” and the core of a new “predictive business

model". The power dimension of big data can be qualified as a capacity to act, in the sense of taking the right decisions in farming and investment in agriculture: publicly available weather data, combined with the data collected by sensors mounted on tractors, allows for big data analyses that (potentially) guide all important farming decisions such as seeding, fertilisation, irrigation and allow for crop predictions. In Carbonell's analysis, which draws on the power concepts of French and Raven (1959 / Raven, 1965), this practice of "data-driven farming" or "smart farming" reinforces pre-existing power relations between big agribusiness and small farming to the detriment of small farmers. This is due to the uneven access to big agricultural data between big agribusinesses - such as Monsanto and John Deere, and the farmers who buy their seed. By signing Monsanto's technology use agreement, farmers are deprived from their data and informational rights (informational power); big data adds to existing tools of domination (coercive power). Small farmers not only lack the necessary data, they also seldomly dispose of the necessary expertise to use big data methods (expert power). In addition, Carbonell observes an unequal power balance between big agribusinesses and the public, due to the lack of transparency of how "big agricultural data" is shared and used. These power relations could nevertheless be partly overcome *with the help* of big data, if agricultural data became available to small farming and for open or publicly funded research.

Yoni Har Carmel's analysis lays out that the expansion of "learning analytics" and "educational data mining" that combine data extracted from digital learning resources (apps, virtual environments, platforms etc.) with conventional data generated in the education system for the purposes of making educational decisions, has given rise to a new powerful sector in the education system: the "edtech" industry. As schools rely increasingly on big data solutions of the for-profit sector for planning, designing and assessing learning processes, their autonomy decreases. The current practice of entrusting the edtech industry with school and student-related data analysis leads to an unequal power relation between, on the one hand the edtech industry, and on the other hand students and parents who perceive few opportunities to "opt out" of the use of their educational data once these practices have been institutionalised in schools. As the author shows, this has nourished many concerns by privacy advocates about possible stigmatisation and discrimination of students labelled "at risk". Thus, Har Carmel argues, big data in education not only challenges students' civil rights and liberties (this would threaten their empowerment), it also bares the potential of increasing students' educational opportunities, thereby offering new capacities to act.

Magdalena König scrutinises the use of big data in border control systems such as the Schengen Information System (SIS), the Visa Information System (VIS) and the EUROpean DACtylographic comparison system (EURODAC). The respective datasets are increasingly integrated, efficient and include more and more biometric data and network information, in addition to traditional border control information such as name, sex and date of birth. Big data thus arms state agencies with the capacity to make routine and systematic searches, not only for immigration control, but also for law enforcement, as Europol and Interpol have access to the systems. Through the use of social sorting, big data in border control becomes a tool of domination: individuals are classified under certain risk categories such as non-citizens, asylum seekers, irregular migrants or presumed terrorists - the concrete criteria for the categorisation remaining opaque to the public. The technically enhanced systems, König argues, "empower[...] the governmental entity managing the tools while disempowering people put in undesirable categories". This disempowerment results in the limitation of freedoms: the freedom of movement and the freedom from suspicion and reinforces social differences in the long term.

In *Michiel Rhoen's* study on how big data impacts consumers' power, big data is the massive

collection and use of data about consumer activities such as “personal communications, online behaviour, shopping, banking and public transportation” which, taken together, result in the permanent observation of consumers. That way, big data gives data controllers “the power to influence consumer behaviour through dynamic or discriminatory pricing, filter bubbles or subtly influencing individual decisions (nudging)”. The big data-induced power shift from consumers to data controllers is caused by unequal informational transaction costs and differential access to legal measures such as due process. Data protection law, which is actually meant to re-establish the power balance between consumers and data controllers, currently rather promotes than hinders this shift. Rhoen therefore discusses the refinement of the legal concept of protection against this power asymmetry by adding instruments of consumer protection to data protection instruments.

Summing up, the contributions in this thematic edition provide answers to the questions of what big data is and how it affects power. Refraining from exhaustive big data definitions such as the “Three Vs”¹ or the “13 Ps”² (Lupton, 2015), the authors define big data *for their specific field of analysis*. For a research field that is just emerging, it seems a legitimate strategy to explore the various phenomena under the label big data before striving for a general definition. With a growing body of conceptual and empirical findings it will then become possible to assess what is the essence of big data (if there is any). Yet looking at the bottom line of the big data definitions provided here, we observe that big data, in concrete applications, is a series of phenomena that encompass large datasets, fed by digital sources that are combined with conventional sources. These datasets, whose concrete content often remains opaque to data subjects and the public, are then processed by algorithms that are equally opaque and used for two main purposes: prediction (e.g., border control, harvest, educational success, consumer behaviour) and individualised treatment in “real-time” (e.g., farming, learning, marketing).

When it comes to the ways in which big data influences power, the articles of this edition illustrate the many dimensions of power: power can be the influence upon others; the capacity to act, to achieve something. Power can also take the form of the autonomous empowerment of an individual or a group, or, express itself in the self-discipline of a group under a common agreement, e.g. when a society binds itself to a constitution to protect certain individual rights (Göhler, 2009, pp. 34-35). When looking at big data, it becomes evident that these power dimensions interrelate: “big data power”, as *capacity to act* (making the right decision when it come to harvesting, learning, granting a visa or marketing) is distributed unevenly. It can therefore be used for seeking *influence upon others* (agribusinesses over farmers, edtech firms over students, authorities over migrants, data controllers over consumers). A common mechanism of that power shift is that data subjects (farmers, students, consumers) provide their data under terms of contract that are unilaterally defined by data controllers, thereby losing control over that data. A last dimension of “big data power” emerges when actors make an *agreement that binds* them, e.g. when students and parents consent to base their educational choices on algorithmic suggestions. A wider access to big data can finally foster the *empowerment* of formerly disadvantaged actors (e.g. researchers who try to prove that industrial agriculture has negative external effects).

Concluding on the question whether the way big data affects power brings about profound societal transformation or rather exacerbates inequalities, all authors observe that, without radically substituting the existing practices or systems, big data has brought enough newness to challenge much of the existing regulation.

SHAPING THE “BIG DATA SOCIETY”

Unethical use of big data can be controlled and unequal power balances can be recalibrated, as the contributions in this thematic edition assert: the advent of big data has led to conflicts and what we diagnose as the present state of things is contested and may well change. One way is to challenge the privileged position of data collectors, data brokers and data controllers by granting wider access to (some of) the data and data analysis: provide data subjects with participation rights and comprehensible information (Rhoen, Har Carmel), grant data access to a wider range of possible data users through open data initiatives (Carbonell), increase public transparency about governmental datasets (König), arm more possible beneficiaries with the tools to analyse that data – be it through use-friendly free software or with the help of publicly funded research (Carbonell). Where software is not freely available due to the protection of business models, code-audits can help to detect biases and prevent discrimination (Har Carmel). All these suggestions underline the importance of public transparency: intelligible information about datasets, data analysis and data use - including applied methodologies - are crucial for the necessary public debate about what uses of big data are welcome and legitimate in their various fields of application. The current opacity about the details of big data is thus a main obstacle for regulation.

A more general conclusion on the regulation of the “big data society” emerges from the contributions to this thematic edition: the power shifts caused by big data as well as the threats for individuals, companies and the society as a whole, occur in extremely diverse fields of social life. We cannot observe one phenomenon that we would call big data: there are many. Those seeking to provide protection against the threats caused by big data-driven innovation need to understand, first of all, what kinds of power shifts emerge, where, and which threats and unintended consequences they cause. Only on the basis of this knowledge, which should emerge from interdisciplinary research (see research questions below), will it be possible for civil society groups, companies, policy makers and other actors to choose the appropriate protection instruments against these threats. In addition, regulation of innovation often occurs in highly dynamic environments and produces a number of paradoxa (Kirby, 2008, pp. 373-381). For example, while regulatory inaction may allow producing and using technologies in a way that is regretted later on, precautionary instruments may turn into over-regulation. Compared to the speed of innovation, regulators may also react too slow. This can lead to the situation that regulatory instruments, once established after a parliamentary legislation process, turn ineffective because their target has already changed. The task of regulation will be, in light of the dynamics and unpredictability of innovation, a responsive learning process (cf. Black & Baldwin, 2010).

Facing big data, those looking for solutions hence have to deal with two challenges: first, they have to refine constantly the various objects of protection. This means taking into consideration that big data is not only a threat to privacy, but also to free educational choices, free movement, freedom to conduct business, free competition etc. Secondly, they have to adapt the concepts of protection, referring, solely or cumulatively, to data protection, consumer protection, antitrust law, protection against discrimination and other laws - some of which may still need to be defined. This is the way towards building trust in big data and striking a balance between the opportunities of big data-driven innovation and appropriate protection against its risks (Hoffmann-Riem, 2009).

PRESSING RESEARCH QUESTIONS

In a world where from agriculture to education, border control and e-commerce, everything becomes “smart”, scholars should be “wise” and scrutinise the power shifts, which include opportunities and threats, of the “big data society”. Considering the findings and blind spots of the articles of this thematic edition we regard the following research questions as particularly important:

We call for more empirical studies about the social consequences of big data: this thematic edition indicates that big data analyses might lead to unequal access to mobility and education. Yet, these assertions are conceptual. The articles do not provide any empirical evidence about the social consequences of big data. It is worthwhile to find out whether big data produces new marginalised groups or augments existing cleavages. More empirical studies should therefore enquire how big data analyses and the decisions based upon them affect individuals, when it comes to social status, gender, citizenship and other categories that traditionally account for social inequalities.

This leads us to the question of where to draw the line between use and abuse of big data. This question needs normative analysis that addresses questions of fairness, social equality and other principles for regulation. The question of use and abuse of big data also demands empirical studies that explore which big data practices are socially acceptable. The articles in this volume do not provide the respective insights. Philosophers, psychologists, legal scholars, political scientists, sociologists, economists, computer scientists (and other scholars) have to work jointly to find the many answers to this question.

Another important line of inquiry concerns what happens inside of the big data black box. Albeit not being a new question, there is still little evidence about who composes the respective algorithms, what mathematical models are used, what assumptions about the data, about causality, about the world are made. The articles in this volume do not fill this blind spot, but they point out that due to the opacity of datasets and data analysis, field access is a major challenge for this research. The seemingly abundant existence of data does in no way mean that it is available. Interdisciplinary work between mathematics/informatics and social sciences/humanities should help us understand what validity big data calculations have and how their results need to be contextualised.

Thank you to all the reviewers, Prof Jeanette Hofmann, Prof Wolfgang Schulz, Prof Ingrid Schneider, Dr Cornelius Puschmann and the many helping hands for their insightful feedback.

REFERENCES

- Amoore L. (2013). *The Politics of Possibility: Risk and Security Beyond Probability*. Durham, NC: Duke University Press Books.
- Barocas, S., & Nissenbaum H. (2014). Big Data's End Run around Anonymity and Consent. In J. Lane, V. Stodden, S. Bender, & H. Nissenbaum (Eds.), *Privacy, Big Data, and the Public Good: Frameworks for Engagement* (pp. 44-75). New York, NY: Cambridge University Press
- Black, J., & Baldwin, R. (2010). Really responsive risk-based regulation. *Law and Policy*, 32(2), 181-213. Retrieved from <http://eprints.lse.ac.uk/27632/>
- boyd, d., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, 15(5), 662-667. doi: 10.1080/1369118X.2012.678878
- French, J. R. P., & Raven, B. H. (1959). The bases of social power. In D. Cartwright (Ed.), *Studies in Social Power* (pp. 150–167). Ann Arbor, MI: Institute for Social Research.
- Göhler, G. (2009). Power to and Power over. In S. Clegg & M. Haugaard (Eds.), *The SAGE Handbook of Power* (pp. 25-39). London, England: SAGE Publications.
- Hoffmann-Riem, W. (2009). Responsibility for Innovation. In M. Eifert & W. Hoffmann-Riem (Eds.), *Innovation und Recht III: Innovationsverantwortung* (pp. 11-41). Berlin, Germany: Duncker & Humblot.
- Kirby, M. (2008). New Frontier: Regulating Technology by Law and 'Code'. In K. Brownsword & K. Yeung (Eds.), *Regulating Technologies: Legal Futures, Regulatory Frames, and Technological Fixes* (pp. 367-388). Oxford, England: Hart Publishing.
- Lupton, D. (2015, May 11). The thirteen Ps of big data [Blog post]. Retrieved from <https://simplysociology.wordpress.com/2015/05/11/the-thirteen-ps-of-big-...>
- Lyon, D. (2014). Surveillance, Snowden, and Big Data: Capacities, consequences, critique. *Big Data & Society*, 1(2), 1–13. doi:10.1177/2053951714541861
- Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.
- Portmess, L., & Tower, S. (2015). Data barns, ambient intelligence and cloud computing: The tacit epistemology and linguistic representation of Big Data. *Ethics and Information Technology*, 17(1), 1–9. doi:10.1007/s10676-014-9357-2
- Puschmann, C. & Burgess, J. (2014). Metaphors of Big Data. *International Journal of Communication*, 8(1), 1690–1709.
- Raven, B. H. (1965). Social influence and power. In I. D. Steiner & M. Fishbein (Eds.), *Current Studies in Social Psychology* (pp. 371–382). New York, NY: Holt, Rinehart, Winston.
- Zuboff, S. (2015). Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1), 75–89.

FOOTNOTES

1. These three Vs refer to a widely used and often criticised technical definition that states that

big data is defined by “volume”, “variety” and “velocity”. Volumes stands for the large scale of the data, variety for the many forms of data that emerge from the digitisation of societies. Velocity indicates the increasing speed of data gathering, processing and analysis that allows “real-time” analyses.

2. The 13 Ps were developed by Deborah Lupton, based on her readings of critical data studies and refer to the sociocultural dimensions of big data, such as “portentous”, “perverse” and “personal”.