Research Center for
Information Law
University of St.Gallen

Taming the Beast:
Big Data and the Role of Law

By:
Patrick Eggimann, Executive Manager, Research Center for Information Law (FIR-HSG)
Aurelia Tamò, Researcher, Research Center for Information Law (FIR-HSG)

1 Mapping the Problem

The term big data has become omnipresent – journalists, privacy scholars and politicians are becoming aware of its importance. The benefits as well as concerns that big data is linked to, are not fundamentally new to privacy advocates. The root and rationale behind big data had earlier been debated under the term data mining or data warehousing. Yet, big data goes beyond the known: with the increased velocity of data processing, the immense volume of generated data and potential to combine a variety of data sets, the so far undeveloped predictive element in our digital world has been released.[i]

Until now, «datafication», or the quantification of information about all things happening, has shifted the focus away from the search for causality. In order to reduce complexity, correlations within big data sets are analyzed. Based on these correlations, predictions are made.[ii] A future dimension of big data could well shift the focus of analytics back to causality once again. Overall, the high level of complexity for analyzing the "what or the why" requires complex, autonomous processes which are often opaque for users. Accordingly, the human capacity for understanding how data is being processed within the system, and on what grounds the outcomes are being justified, is seriously challenged.

The user's loss of control over and ignorance of how the data and information is handled and in what ways the resulting knowledge is used, leads to civil liberties concerns. Knowledge extracted from the information provided in the big data sets is in fact the "key to power in the information age".[iii] Even if the search for knowledge is in general to be considered a positive goal of the big data phenomenon, knowledge can turn out to be destructive depending on how it is used (a fact that Albert Einstein already recognized). From a social and democratic perspective, the concentration of knowledge as power in the hands of a few together with its potential misuse would represent such a destructive force. Moreover, an increased fear of being observed and analyzed could result in a threat not only to the freedom of speech or freedom of information, but more broadly to the individuals' willingness to participate in public and democratic debates, or even in social interactions on an individual level.[iv]

2 The Role of Data Protection Law in a Big Data Age

In Europe, the European Convention for Human Rights (ECHR) as well as the Charter for Fundamental Rights (EUCFR) protects the individual's private and family life (Art. 8 ECHR, Art. 7 EUCFR) as well as his or her personal data (Art. 8 EUCFR). These fundamental rights are incorporated into European data protection law (Directive 95/46/EC), which on the basis of the protection of the individual's right to personality, is the main approach when dealing with (big) data processing. In particular, the fundamental principles of consent, transparency, purpose

limitation, data minimization, security and proportionality are key to restricting the processing and evaluation of big (personal[v]) data sets.

When talking today about the limitations of data processing the focus lies primarily on private companies, such as Google, Facebook or Amazon. This fact is of special interest because the *ratio legis* behind the introduction of data protection law in Europe was the protection of the individual against the superiority of governmental bodies and the potential misuse of citizens' data and census databases rather than the threats from private entities.[vi] This scope is also reflected in the famous *Volkszählungsentscheid* of the German Supreme Court of 1983 which is seen as the fundament for the right of informational self-determination.[vii]

Even though, the data protection principles in Europe are applicable to both, governmental bodies and private parties that are processing data, the trend that private companies possess and handle a great deal of valuable information about individuals has shifted the balance of knowledge. The recent PRISM and Tempora affairs illustrate the fact that governments want to have what Silicon Valley has: vast amounts of private data and the most sophisticated technology to harvest it.[viii]

Distinguishing the actors that interplay in informational relationships is crucial, since the founding rationales governing the relationship are converse: When the government is processing the personal data of citizens, its actions must be democratically legitimized by legal norms, whereas the informational relationships between private entities and consumers are governed by the freedom of contract.

Against this backdrop and in light of big data processing, the principle of purpose limitation is of particular interest. This principle, also referred to in the US as purpose specification,[ix] stands in contrast to the mechanism of big data. A rigorous enforcement of purpose limitation would preclude big data since it lies in its logic to evaluate more data for purposes unknown at the moment of collection. The question remains therefore, whether this democratically legitimized principle stands above consent, i.e. the parties' agreements on data processing. Such an extensive application is suggested by the European Data Protection Authority, so-called Working Party 29.[x]

Restrictions among private parties were not conceived within the original purpose of data protection law in Europe. Even if it can be argued that the principle of consent is currently applied in a problematic way,[xi] there is no mandate for a state authority to narrow the scope of private consensus by restrictively applying data protection principles. Such an approach results in a hybrid understanding of data protection regulation, which collides with the underlying *ratio legis* of data protection law. By protecting the specified *raison d'être* of data processing, data protection authorities in Europe use a questionable paternalistic approach to overcome the information asymmetry between the data controller and the data subject. State interventions in

general, and legal provisions that are protecting the weaker party in particular, are by no means reprehensible and are usefully adopted in many areas of the law.[xii] Nevertheless, when it comes to data protection in a big data world such an approach reaches its limits.

3 Overcoming the Big Challenges

Different approaches toward overcoming the challenges arising out of big data have been discussed by legal scholars.[xiii] We argue that taking an approach based on consent when personal data is being processed by private entities is not totally amiss. In theory, contract law has the advantage of offering greater flexibility and respects considered, self-determined consumer choices.[xiv] In practice however, the downside remains the information asymmetry, which in our highly technologized world of big data is increasingly challenging. In addition, the option of negotiation as a vital element of a contract, is underdeveloped and in peril when agreement is already considered to be reached by the mere usage of a service.[xv] The question is how to overcome these practical obstacles by other means than strict regulatory intervention.

Overcoming information asymmetries (rather than the superiority of the state as rooted in data protection law outlined above) and creating options for successful negotiations are not singular problems of big data. However, big data accentuates asymmetries due to its complexity, unpredictability and individuals' lack of awareness that data is being processed. Contractual law has already established counter mechanisms to overcome these challenges, such as the principle of *culpa in contrahendo* regarding negotiations or the principle of good faith. Also the courts in civil law countries play an important role in concretizing such principles. In Switzerland for instance, a court ruling obliged banks to disclose relevant information to its clients in order for them to be able to contractually waive the profits out of retrocession payments by third parties.[xvi]

Solutions to enhance negotiation between private parties should be centered on improving the choices of the individuals. Here the option to choose the private entity that is processing the personal data is key. Already today, a variety of private entities lure users to their services by providing them with what they need without the exchange of personal information. The search engine duckduckgo, whose increasing user number was further boosted with the PRISM affair, or the software disconnect, as an example for a privacy by design solution provided by a third party, are two examples of how competition and innovation can lead to a more diverse field of options for consumers. Also mechanisms such as labeling could be implemented in an online world to counterbalance the information gap and facilitate more informed consumer choices.[xvii] Governments then have the responsibility to ensure market conditions that enhance such innovation through appropriate regulation.

As the argument laid out here shows, we are not claiming that governments should not play a role in the current debates on how to regulate our big data world. On the contrary, governments play a crucial role not only in the education of their citizens, but also in setting the underlying structures in which technology can and will flourish. Transparency and choice play an important role in this context: informed individuals should be put in the position to decide what they are willing to give up in order to gain new possibilities and make use of the latest technological advancements.

The willingness and ease with which people make use of new technologies is essentially determined by trust.[xviii] Trust is key when it comes to establishing a relationship since transparency is almost always only given to a certain degree. Nevertheless, transparency must be measured on its result, which ought to be clarity and not obfuscation. In this sense, the tools of big data are very likely to be not only the cause of the problem but also part of the solution. This can be seen in applications such as disconnect, which graphically captures the potential big data processors. In relation to the government, trust entails the expectation that the former will not fall short on its promise to enforce its laws.

Taking a step back, we believe it is important not to forget the social changes resulting out of the evolving consolidation of the digital and non-digital spheres. As a recent study of online-behavior on social networking sites by the Pew Research Center has shown, adolescents are adapting to the new privacy conditions online. This adaptation is in our opinion an important factor as it reflects an individual change of attitude[xix] that has not yet been integrated enough into debates between politicians, industry representatives and consumer protection organizations. We see here the potential for academia to provide further insights into the understanding of the relationship of society, law and technology.

---

[i] Paul C. Zikopoulos et al., IBM Understanding Big Data 15 (2012), https://www14.software.ibm.com/webapp/iwm/web/signup.do?source=sw-infomgt&S_PKG=500016891&S_CMP=is_bdebook1_bdmicrornav.

[ii] Mayer-Schönberger & Cukier, Big Data – A Revolution That Will Transform How We Live, Work, and Think 14, 79 et seqq. (2013).

[iii] Daniel J. Solove, The Digital Person - Technology and Privacy in the Information Age 74 (2004).

[iv] Fred H. Cate & Viktor Mayer-Schönberger, Notice and Consent in a World of Big Data, Microsoft Global Privacy Summit Summary Report and Outcomes 5 (2012), http://www.microsoft.com/en-us/download/details.aspx?id=35596; Lizette Alvarez, Spring Break Gets Tamer as World Wachtes Online, NYTimes (March 16, 2012), http://www.nytimes.com/2012/03/16/us/spring-break-gets-tamer-as-world-watches-online.html?_r=0.

[v] In Europe the Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data is applicable whenever personal data is being processed. The definition of the term "personal data" and the utopia of "anonymous" data have already been discussed in depth by legal scholars: Cf. Paul Ohm, Broken Promises of Privacy to the Surprising Failure of Anonymization, 57 UCLA L. Rev. 1701 (2010); Paul M. Schwartz & Daniel J. Solove, The PII Problem: Privacy and a New Concept of Personally Identifiable Information, 86 N.Y.U. L. Rev. 1814 (2011).

[vi] Colin Bennett, Regulating Privacy: Data Protection and Public Policy in Europe and the United States 29 et seqq. (1992).

[vii] Horst-Peter Götting, Christian Schertz & Walter Seitz, Handbuch des Persönlichkeitsrechts § 22 N 59 (2008).

[viii] Bits New York Times, Deepening Ties Between N.S.A. and Silicon Valley (June 20, 2013), http://bits.blogs.nytimes.com/2013/06/20/daily-report-the-deepening-ties-between-the-n-s-a-and-silicon-valley/?nl=technology&emc=edit_tu_20130620.

[ix] Daniel J. Solove, Understanding Privacy 130 (2008).

[x] Article 29 Data Protection Working Party, Opinion 03/2013 on purpose limitation 11 et seqq. (April 2, 2013), http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf.

[xi] Cate & Mayer-Schönberger, *supra* note 4, 3 seq.; Ira S. Rubinstein, Big Data: The End of Privacy or a New Beginning?, NYU Public Law Research Paper No. 12-56, 3, 5 (2013); cf. also Solon Barocas & Helen Nissenbaum, On Notice: The Trouble with Notice and Consent, Proceedings of the Engaging Data Forum: The First International Forum on the Application and Management of Personal Electronic Information (October 2009), http://www.nyu.edu/projects/nissenbaum/papers/ED_SII_On_Notice.pdf.

[xii] Cf. Jeremy A. Blumenthal, Emotional Paternalism, 35 Fla. St. U. L. Rev. 1 (2007).

[xiii] Mayer-Schönberger & Cukier, *supra* note 2, 173 argue for holding data controllers accountable for how they handle data and propose a tort law approach; Omer Tene & Jules Polonetsky, Big Data for All: Privacy and User Control in the Age of Analytics, 11 Nw. J. Tech. & Intell. Prop. 239, 263 seq. propose a "sharing the wealth" strategy with data controllers providing individuals access to their data in a usable format and allowing them to profit from data analysis with the help of tools provided; for more references see Ira S. Rubinstein, *supra* note 11, fn. 48.

[xiv] Cf. Robert Cooter & Thomas Ulen, Law & Economics 355 (6th ed. 2012).

[xv] Cf. e.g. Facebook Statement of Rights and Responsibilities as of December 11, 2012, 14, (3) states that "Your continued use of Facebook following changes to our terms constitutes your acceptance of our amended terms", https://www.facebook.com/legal/terms.

[xvi] Fabien Aepli et al., Landmark decision of the Swiss Federal Supreme Court (November 2, 2012), http://www.lexology.com/library/detail.aspx?g=61377f8b-bd6b-430f-b826-87fe0fed63f3.

[xvii] The European Interactive Digital Alliance has announced the approval of two technology platform providers to serve an Online Behavioural Advertising Icon, http://www.iabeurope.eu/news/edaa-names-truste-and-evidon-approved-oba-icon-providers.

[xviii] Cf. World Economic Forum in collaboration with the Boston Consulting Group, Rethinking Personal Data: Strengthening Trust (May 2012), http://www.weforum.org/reports/rethinking-personal-data-strengthening-trust; McKinsey Global Institute, Big data: The next frontier for innovation, competition, and productivity 116 (June 2011). http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation.

[xix] Pew Research Center, Teens, Social Media and Privacy, Berkman Center 8, 41-50, 63 (2013), http://www.pewinternet.org/~/media//Files/Reports/2013/PIP_TeensSocialMediaandPrivacy.pdf.