Effective Big Data Management and Opportunities for Implementation

Manoj Kumar Singh
Adama Science and Technology University, Ethiopia

Dileep Kumar G.
Adama Science and Technology University, Ethiopia

A volume in the Advances in Data Mining and Database Management (ADMDM) Book Series

Information Science REFERENCE
An Imprint of IGI Global

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA, USA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

Library of Congress Cataloging-in-Publication Data

Names: Singh, Manoj Kumar, editor. | Kumar G., Dileep, 1982- editor.
Title: Effective big data management and opportunities for implementation / Manoj Kumar Singh and Dileep Kumar G, editors.
Description: Hershey : Information Science Reference, 2016. | Includes bibliographical references and index.
Identifiers: LCCN 2016004454| ISBN 9781522501824 (hardcover) | ISBN 9781522501831 (ebook)
Subjects: LCSH: Big data. | Database management.
Classification: LCC QA76.9.D3 E335 2016 | DDC 005.7--dc23 LC record available at https://lccn.loc.gov/2016004454

This book is published in the IGI Global book series Advances in Data Mining and Database Management (ADMDM) (ISSN: 2327-1981; eISSN: 2327-199X)

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Chapter 9
Legal Responses to the Commodification of Personal Data in the Era of Big Data:
The Paradigm Shift from Data Protection towards Data Ownership

Emile Douilhet
Bournemouth University, UK

Argyro P. Karanasiou
Bournemouth University, UK

ABSTRACT

Big Data is a relatively recent phenomenon, but has already shown its potential to drastically alter the relationship between businesses, individuals, and governments. Many organisations now control vast amounts of raw data, and those industry players with the resources to mine that data to create new information have a significant advantage in the big data market. The aim of this chapter is to identify the legal grounds for the ownership of big data: who legally owns the petabytes and exabytes of information created daily? Does this belong to the users, the data analysts, or to the data brokers and various infomediaries? The chapter presents a succinct overview of the legal ownership of big data by examining the key players in control of the information at each stage of processing of big data. It then moves on to describe the current legislative framework with regard to data protection and concludes in additional techno-legal solutions offered to complement the law of big data in this respect.

INTRODUCTION

Big Data is a relatively recent phenomenon, but has already shown its potential to drastically alter the relationship between businesses, individuals, and governments. The issues surrounding privacy of the online users (Mayer-Shoenberger, Cukier 2013) and the overall ethical challenges involved (Schroeder, 2014) make big data a topical issue, especially in the aftermath of the Snowden revelations. Many organi-

DOI: 10.4018/978-1-5225-0182-4.ch009

sations now control vast amounts of raw data, and those industry players with the resources to mine that data to create new information have a significant advantage in the big data market. The use of predictive analytics in processing information tracked across different platforms to identify trends in the behaviour of individuals further adds value to big data (Fotopoulou, 2014) and makes it an important asset for any commercial entity. This rapid commodification of personal data has given rise to a new approach with regard to its legal protection in the era of big data: a shift from the traditional privacy protection regime to a wider protection under property law is considered by scholars as an appropriate legal response to the phenomenon of monetisation of personal data, once seen through the lens of big data (Victor, 2013).

The aim of this chapter is to identify the legal grounds for the ownership of big data: who legally owns the petabytes and exabytes of information created daily? Does this belong to the users, the data analysts, or to the data brokers and various infomediaries? The chapter presents a succinct overview of the legal ownership of big data by examining the key players in control of the information at each stage of the processing of big data. It then moves on to describe the current legislative framework with regard to data protection and concludes in additional techno-legal solutions offered to complement the law of big data in this respect, with a particular focus on the European context[1].

BACKGROUND

The transition from the traditional economic model of neoliberal markets in the post-industrial era to "informational capitalism" (Cohen 2016), based on a data-driven economy has challenged conventional legal thinking. Often referred to as the oil of the 21st century, data has become a valuable asset for the key stakeholders offering services in the digital era. At the same time, the law has been struggling to cope with this overbroad scope and definition of "data", as it does not purely address the user's privacy, being able to reveal one's identity but it can also be valorized and thus imply property entitlements for user generated data. The following section explores how data can be legally assessed during various stages of processing: in doing so, it is intended to demonstrate how big data appears to be an area not overly addressed by the current regulative framework, which focusses mostly on data protection and appears to bear little attention to how data can gain monetary value and thus allow for property based claims.

MAIN FOCUS OF THE ARTICLE

Issues, Controversies, Problems: The Four Stages in the Big Data Processing Cycle and Property Law

There are four main stages in the processing cycle of big data from its raw form to its use in predictive analytics:

1. Collection
2. Processing
3. Mining, and
4. Usage.

In the collection stage, raw data is collected through a number of means – either in a direct and voluntary manner by individuals themselves, or indirectly, inferred from the analysis of other data (Al-Khouri, 2012). In the processing stage, data is aggregated in databases and is formatted to be ready for analysis, either by a corporation or by a third party (a "data processor"). It should be noted that at this point the information is transformed from its original crude form at the collection stage and becomes part of one or more large datasets, put together by one or more separate corporations. Then, in the data mining stage, all gathered and processed data is analysed to create useful information. This new information created is essentially independent of the individual bits of information provided at the collection stage. Although it is the direct outcome of the analysis of segments of data from individual users, at this stage it also becomes the product of an analysis performed by entities completely separate from the data subjects, i.e. the users. Finally, in the usage stage, value is extracted from the information, through predicting analytics, data profiling, and any other number of methods able to exploit information for profit making.

Before however one is able to determine whether there are any legal grounds for data ownership in any of these stages, a preliminary question must be answered first: do property rights apply to data? The idea of propertisation of data, namely the protection of data under property or copyright law has been discussed extensively since the 1970s (Fromholz, 2000). A major difficulty in addressing data as property is its intangible nature added to the fact that it can be replicated many times without concrete evidence that its value is lost. On the other hand, copyright law reviewed in general within a digital environment increasingly shaped by big data, is greatly challenged: works are used 'in bulk' for purposes other than making their content available to the public, such as text mining and content mining (Borghi & Karapapa, 2013). Personal data in that sense –although treated as a tradable commodity online- has not yet received explicit protection under copyright law regime, falling mostly within the protective scope of privacy law. Moreover, the European approach to privacy maintains a narrow conceptual approach, regarding this as a human right, which cannot be traded away (Prins, 2004).

As such, there is no explicit legal right of ownership for individual pieces of information. Were we to apply the legal concept of property to big data in any of the three stages mentioned above, we would need to carefully consider the main legal features of the concept of property in general:

- *Usus* (the right to use),
- *Abusus* (to right to encumber or transfer) and
- *Fructus* (the right to enjoy the right) (Segal and Whinston, 2010).

In the absence of a formal right to ownership of big data, parties enjoying those rights should demonstrate these elements of ownership. Given the large amounts invested by the big data controllers, it would appear that the data collected and aggregated by corporations is under their ownership – they hold it in their databases, they process and aggregate it (*usus*), and they extract value from its analysis (*fructus*) and from selling it to other parties (*abusus*).

Nevertheless, data has a unique feature that complicates matters: the information is related to a person, gaining thereby an added aspect of privacy. Under the right to privacy, individuals enjoy a certain level of protection of their personal data, namely data able to identify them or to reveal private information about them without their consent. In this respect, the individual's right to data protection overrides the property right and economic interests of the data processors (Google Spain and Google Inc v Agencia Espanole de Proteccion de Datos of Mario Costeja, C 131-12, hereafter referred to as a the "Google Spain" case).

One of the most robust legislative frameworks dealing with data protection is the EU Data Protection Directive 95/46 (Levin & Nicholson, 2005). This provides us with a coherent legal regime, unlike the

US data privacy law, which is at large scattered (Gaff, Smedinghoff & Sor, 2012). For this reason, the focus here is mostly on the EU data protection laws. The main three distinctions used in the EU Data Protection Directive are "data subject", "controller", and "data processor". A data subject is an "identified or identifiable natural person [...] an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity.", while a "controller" means "the natural or legal person, public authority, agency or any other body which alone or jointly with others determines the purposes and means of the processing of personal data." Finally, a processor is "a natural or legal person, public authority, agency or any other body which processes personal data on behalf of the controller". It should be noted that the Directive uses words like "controller" and "processor" and avoids the appellation of "owner". Yet, although in none of the above definitions is the term ownership explicitly expressed, many provisions seem to suggest a property-based approach to data.

The Data Protection Directive establishes a number of rights for individuals relating to their data – such as the fact that collecting an individual's data requires their prior unambiguous consent (Article 7), that individuals should have access to their data (Article 12), or that they should be able to object to the processing of their data if they have compelling legitimate grounds to do so (Article 14). The Directive also notably includes restrictions on what the data's controller can do with the data, including restrictions on the transfer of that data (Article 25). Drafted in 1995, the Data Protection Directive is currently undergoing reform after the proposal of the General Data Protection Regulation (GDPR) in 2012. The GDPR increases the rights that individuals hold over their data, as well as the restrictions of data controllers (Article 29 Working Party, 2014). In particular, the restriction of what constitutes "consent" to the very high standard of "explicit consent" reinforces the idea that individuals have a property based right: the fact that ultimate control lies with the individual's consent is a clear indication of the data subject considered as data "owner". At the same time though, data can be processed without consent for a "legitimate interest pursued by a controller" (Article 6(1) (f) GDPR). Even though this provision is itself mitigated by the fact that "it shall not override the fundamental rights and interests of the data subject", the fact that the individual does not necessarily have a final say in what happens to their data tempers their power over the data.

Finally, an element of ownership can be found in the ability of the data subject to have an enforceable claim with regard to the online indexing of his personal data. The "right to be forgotten", which is gaining traction in the European context and has been enshrined in the recent Google Spain case, gives data subjects the ability to request that their data, held by data controllers, be de-listed from Google's search results. The right to be forgotten, as defined in the GDPR, is broadly defined (Rosen, 2012) and is important because of the fact that the data subject seems to have retained some rights over his data, even after willingly parting with it. This suggests a property-based approach to data.

SOLUTIONS AND RECOMMENDATIONS

Data Rights and Database Rights: How is Ownership Delineated?

The data protection provisions in the Data Protection Directive and the upcoming General Data Protection Regulation have clear indications that individuals are granted certain rights over their data that extend beyond the traditional framework of privacy. Thus, as it was earlier demonstrated, data controllers

demonstrate various elements of data ownership while processing big data; at the same time, under the right to be forgotten, it seems that at every stage, the data subject retains control over data being able to request erasure. This poses a legal conundrum: Can one be considered to own something in its entirety if at the same time someone else is further granted a right to command them erasure of indexed data?

Although puzzling, it seems that the issues becomes less complicated if one takes into account that a distinction needs to be drawn between segments of personal data and databases built on such data. So far, the chapter has explored the former; turning to explore now the latter, it appears that indeed there are strong indications in European law for specific provisions for a right to data ownership: the "database right".

The 1996 Database Right Directive 96/9 created a "sui generis" intellectual property right on data, the "database right" (Rendie, 2011), namely "the right to prevent extraction and/or reutilization of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of that database." Unlike other intellectual property rights, a database right does not require an original or technical achievement as a prerequisite for affording copyright protection. In fact, a person can have a database right provided that he a substantial investment is made in obtaining, verifying, or presenting data in the database (Reichman & Samuelson, 1997).

In this sense, database rights are automatically granted, and do not require to be registered or applied for. As such, the data constituting a database is not itself owned *per se* – it is rather the database in its entirety, having required time and effort to establish, that is protected. As a result, database rights enshrine in law the current practice of data controllers, granting them thus significant rights over the data they accumulate. Even though personal data is still protected under data protection regulation for the individual, the aggregation, processing and analysis of large amounts of data from a database, appear to have a separate existence separate. Outside the protective remit of privacy, databases are not protected as parts of the individual's identity and can thus be legally owned when one has a "substantial involvement" in it.

As a rule of thumb, EU Courts generally accept that database rights can be legally owned, unlike their components, namely the segments of personal data compiled for a database. The first main guidance for the database right came in 2004 - 8 years after the adoption of the Directive - from four judgments of the European Court of Justice (Aplin, 2005):

- *British Horseracing Board Ltd v William Hill Organization Ltd C-203/02 Judgment of the Court* (Grand Chamber 2004);
- *Fixtures Marketing Ltd v Organismos Prognostikon Agonon Podosfairou (OPAP) C-444/02 Judgment of the Court* (Grand Chamber 2004);
- *Fixtures Marketing Ltd v Svenska Spel AB C-338/02 Judgment of the Court* (Grand Chamber 2004);
- *Fixtures Marketing Ltd v OY Veikkaus Ab C-46/02 Judgment of the Court* (Grand Chamber 2004).

Most importantly, the database right in all these cases emphasise on the importance of organisation and assemblage of data, not on the original creation of the data in a database. Investment in the creation of data does not trigger a database right, assemblage and presentation do. That said, it is worth noting that the CJEU has been recently shifting away from the idea of database rights towards interpreting ToS as contractual obligations that have the potential to outline property over data. One such example is the

Ryanair v PR Aviation, where is was held that the use of data could be restricted by contract (through the online Terms of Use). This development however could be a double edged sword for the consumer, who is at most cases in a restricted bargaining position.

It could be argued that a form of database right also exists in the US (Xu, 2002): In a landmark US Supreme Court case, *Feist* (*Feist Publications, Inc. v. Rural Telephone Service Co., Inc., 499 U.S. 340, 1991*) copyright protection was extended to databases if there is originality in the "selection, coordination, or arrangement of contents for a database". Several cases since have built on this decision, and some state laws also provide from some protection for databases. The protection is however narrow (Gervais, 2007) because of the originality requirement, in contrast to the wider EU provision, which puts the focus on the investment in the database and the arrangement of data.

FUTURE CHALLENGES POSED BY WEARABLE TECH

Harvesting the Fruits of Self-Tracked Data and Controlling Mechanisms

So far it has been contended that although there is no explicit property right in data per se, there seems to be leeway for a property right to apply as far as databases are concerned. The rise of wearable tech, namely devices with sensors measuring the user's daily activities and habits has now posed a new legal challenge: how is legal ownership determined when a dataset is created and curated by the user himself? The growing tendency to self –track and quantify has taken off since its start in 2008 when two former Wired magazine editors, Gary Wolf and Kevin Kelly, co-founded the "Quantified Self" digital tracking group. The term is now used to describe the mainstream phenomenon of adults collecting data as means of recording and analyzing their lifestyle (Haddadi & Brown, 2014). It is estimated that 60% of US adults are currently tracking their weight, diet and exercise routine (Swan, 2014), actively collecting and analyzing their data in the context of their individual experiences (Nafus & Sherman, 2014). Although there is still a corporation acting as a data controller by providing tools for data analysis and storage on their servers, the user can also extract value from this data; this blurs the boundaries of the legal ownership of these commonly created datasets. This issue poses further legal questions, once online health repositories are considered: Microsoft Health Vault and Dossia are two examples of companies offering patients the chance to voluntarily store, collect and share health information with health providers and family members or other users (Steinbrook 2008).

Tim Berners Lee at the 2014 IP Expo Europe stressed the importance of data subjects owning their data instead of the corporations for the purposes of creating "rich" data, namely big data that if merged can be profitable for both the user and the corporations. Although the law has not yet offered a concrete answer to the issue of ownership of such "quantified self" datasets (Purtova, 2011) co-created by the users and the corporations, there is a growing tendency to allow the user for more control and ownership rights over his data with techno-legal solutions and alternative market models (Novotny & Spiekermann, 2013).

Personal Data Vaults (PDS) are currently one of the main technical solutions put forth in order to allow the user to gain control of his data back from the various corporations acting as info-mediaries in the big data market. The idea is to a privacy enhanced architecture enabling the user to access, control and trace their data once shared online (Mun et al, 2010). In this vein, there are many suggestions

employing technical means for the user to reclaim control over his data: Once such example is the MIT Open PDS app, which allows the user to see third-party requests for his data and make informed decisions (de Montjoye at al, 2014). An alternative means of user-controlled data comes from Cozy cloud, a French company that provides users with open sourced private clouds to store their personal data. Other examples include a rising number of start-ups, such as "Personal", "Reputation.com" and "Datacoup", whose aim is to help the user monetize and control own data. That said the law is still admittedly lagging behind in terms of providing user with more control over his data (Crawford, Miltner, Gray 2014; Boyd, Crawford, 2012).

Many countries have embraced user-controlled data as a promising economy boosting strategy: The Midata project, announced in 2011 in the UK, is a multi-stakeholder approach to boost consumer empowerment by giving "consumers increasing access to their personal data in a portable, electronic format" enabling them to "use this data to gain insights into their own behaviour, make more informed choices about products and services, and manage their lives more efficiently" (Department for Business & Innovation Skills, 2011). Similarly in the US, the Federal trade Commission (FTC) in its report entitled "Data Brokers: A Call for Transparency and Accountability" issued in May 2014, calls for tighter regulation of the data brokers, namely large companies trading the user's data without the user's knowledge or consent.

CONCLUSION

At present, data controllers have the most control over data under the database right protection and are thus the primary beneficiaries of the value extracted from big data. At the same time, there seems to be a slight shift towards empowering the user to control and perhaps "own" his data, although legally this is has not yet fully been established. Since 2012, when the European Commission first suggested vital changes to the legal framework on data protection in the EU, the issue of user generated data as one's own property has been frequently discussed, yet there is no legal provision supporting such claims.

In December 2015, the trilogue discussions between the European Commission, the Parliament and the Council were concluded and a final text for the GDPR was agreed, which is expected to come into force in the first half of 2018 and will have immediate effect. Although there is no direct reference to data as property (or even an entitlement to digital personhood, which might imply a certain level of control over one's data), the provisions included with regard to transnational data flows, indicate how the issue of trans-border data flows will be heavily discussed in the years to come. The GDPR's expanded territorial reach, offering protection to EU consumers against any data controller/processor targeting them, irrespective of where the latter is based combined with the obligation to implement technical and organizational measures to notify the controller of data breaches are both new additions to the Data Protection Directive that suggest a more nuanced approach towards personal data. Further to this, data portability and interoperable standards, although currently encouraged and not mandated, still present the user with the potential to re-use his own data across a number of platforms for personal purposes.

A right to own one's data is still far from being explicitly mentioned in the final text of the GDPR. It remains yet to be seen, how this nascent area of data prophetization will be further shaped by the CJEU rulings, especially in the light of the EU Digital Single Market.

REFERENCES

Al-Kouri, A. (2012). Data Ownership, Who Owns my Data? *International Journal of Management & Information, 2*(1).

Aplin, T. (2005). *The ECJ Elucidates the Database Right*. London: Intellectual Property Quarterly.

Article 29 Working Party. (2014). *Opinion 06/2014on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC, 9 April 2014,* 844/14/EN WP 217.

Borghi, M., & Karapapa, S. (2013). *Copyright & Mass Digitisation*. Oxford, UK: Oxford University Press.

Boyd, D., & Crawford, K. (2012). Critical Questions for Big Data: Provocations for a Cultural, Technological and Scholarly Phenomenon. *Information Communication and Society*, *15*(5), 662–679. doi: 10.1080/1369118X.2012.678878

Cohen, J. (2016). The Regulatory State in the Informational Age. *Theoretical Inquiries in Law, 17*(2).

Crawford, K., Miltner, K., & Gray, M. (2014). Critiquing Big Data: Politics, Ethics, Epistemology. *International Journal of Communication*, *8*, 1663.

de Monjoye, Y.-A., Shmueli, E., Wang, S. S., & Pentland, A. S. (2014). Open PDS: Protecting the Privacy of Metadata through SafeAnswers. *PLoS ONE*, *10*, 1371.

Department for Business & Innovation Skills. (2011). *The Midata Vision of Consumer Empowerment*. Retrieved from https://www.gov.uk/government/news/the-midata-vision-of-consumer-empowerment

Fotopoulou, A. (2014). *Tracking Biodata: Sharing and Ownership.* Report on Research Placement funded by the RCUK Digital Economy NEMODE Network.

Fromholz, J. (2000). The European Union Data Privacy Directive. *Berk. Tech. LJ*, *15*, 461.

Gaff, B. M., Smedinghoff, T. J., & Sor, S. (2012). Privacy and Data Security. *Computer*, (3), 8–10.

Gervais, D. J. (2007). The *Protection of Databases Kent L. Rev.*, *82*, 1109.

Haddadi, H., & Brown, I. (2014). *Quantified Self and the Privacy Challenge*. Technology Law Futures.

Levin, A., & Nicholson, M. J. (2005). Privacy Law in the United States, the EU and Canada: The Allure of the Middle Ground. *U of Ottawa Law & Technology Journal*, *2*(2), 362.

Mayer-Schoenberger, V., & Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work and Think*. J Murray.

Mun, M., Hao, S., Mishra, N., Shilton, K., Burke, J., & Govindan, R. et al. (2010). *Personal Data Vaults: A Locus of Control for Personal Data Streams*. ACM CoNext.

Nafus, D., & Sherman, J. (2014). This One Does Not Go Up to 11: The Quantified Self Movement as an Alternative Big Data Practice. *International Journal of Communication*, *8*, 1784–1794.

Novotny, A., & Spiekermann, S. (2013). Personal Information Markets AND Privacy: A New Model to Solve the Controversy.*11th International Conference on Wirtschaftsinformatik*, Leipzig, Germany.

Prins, J. E. (2004). The propertization of personal data and identities. *Electronic Journal of Comparative Law, 8*(3).

Purtova, N. (2011). *Property Rights in Personal Data: A European Perspective*. Kluwer.

Reichman, J., & Samuelson, P. (1997). Intellectual Property Rights in Data. *Vand. L. Rev., 50*.

Rendie, A. (2011). *Aggregation: Demystifying Database Rights*. Taylor Wessing.

Rosen, J. (2012). The Right to be Forgotten. *Stanford Law Review*, *64*, 88.

Schroeder, R. (2014). Big Data: Towards a More Scientific Social Science and Humanities. In M. Graham & W. H. Dutton (Eds.), *Society and the Internet: How Networks of Information are Changing our Lives*. Oxford, UK: Oxford University Press.

Segal, I., & Whinston, M. (2012). Property Rights: Handbook of Organizational Economics. Princeton University Press.

Steinbrook, R. (2008). Personally Controlled Online Health Data – The Next Big Thing In Medical Care?' (2008). *The New England Journal of Medicine*, *358*, 16.

Swan, M. (2013). *The Quantified Self: Fundamental Disruption in Big Data Science and Biological Discovery* (Vol. 1). Big Data.

Van Alstyne, M., Brynjolfsson, E., & Madnick, S. (1995). Why Not One Big Database? Principles for Data Ownership. *Decision Support Systems*, *15*(4), 267–284.

Victor, J. M. (2013). The EU General Data Protection Regulation: Toward a Property Regime for Protecting Data Privacy. *The Yale Law Journal*, 513.

Wu, X. (2002).. . *EC Data Base Directive. Berkeley Tech. LJ*, *17*, 571.

KEY TERMS AND DEFINITIONS

Infomediaries: Various entities that aggregate and link information on subjects or groups of subjects on behalf of commercial organizations and their potential customers.

Privacy by Design: A policy approach, which suggests a privacy friendly design as a techno-legal means of enforcing data protection legislation.

Privacy Enhancing Technologies: Technologies (both hardware and software) designed to protect the privacy of users in the online context.

Propertisation of Data: The tendency among legal scholars towards developing a theory that aims at conceptualizing data as property, bound by property and ownership rights.

Quantified Self: The use of self-collected personal data, mostly using biosensors built in wearable technologies, to improve one's health and well-being.

ENDNOTE

[1] A detailed version of this chapter has been presented at the 2016 IEEE International Conference on Cloud Engineering Workshops (IC2EW 2016). The authors are grateful to all reviewers for the comments and feedback provided. Any errors or omissions remain the sole responsibility of the authors.