

Association for Information Systems AIS Electronic Library (AISeL)

International Conference Information Systems 2016
Special Interest Group on Big Data Proceedings

Special Interest Group on Big Data Proceedings

2016

Ethical Framework of Big Data Application

Larisa Giber

National Research University Higher School of Economics, lbgiber@gmail.com

Follow this and additional works at: <http://aisel.aisnet.org/sigbd2016>

Recommended Citation

Giber, Larisa, "Ethical Framework of Big Data Application" (2016). *International Conference Information Systems 2016 Special Interest Group on Big Data Proceedings*. 11.
<http://aisel.aisnet.org/sigbd2016/11>

This material is brought to you by the Special Interest Group on Big Data Proceedings at AIS Electronic Library (AISeL). It has been accepted for inclusion in International Conference Information Systems 2016 Special Interest Group on Big Data Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Ethical Framework of Big Data Application

Larisa Giber

Faculty of Business and Management, School of Business Informatics,
Department of Modeling and Business Process Optimization,
National Research University — Higher School of Economics, Myasnitskaya
Street 20, 101000 Moscow, Russia
lbgiber@gmail.com

Abstract

In this we consider ethical issues of Big Data application by drawing the integrated ethical framework of Big Data implementation that unites ethical issues related to Big Data application and helps various companies to accommodate to local information privacy standards. This is a result of analyzing ethical risks, which vary in different countries depending on legal aspects and information privacy of individuals.

Introduction

Mayer-Schoenberger and Cukier (2013) claimed that Big Data may “undermine the idea of personal responsibility, particularly as one of the cornerstones of the modern worldview is the idea of free will”. Accidental but critical ethical consequences for individuals include leakage of private data, loss of individual’s identity and even creation of citizen’s identities black market that leads to possible legal actions against an enterprise from the state.

In this paper, we examine the implicit social aspects of big data development:

1. We conduct a survey on ethical issues and challenges of Big Data implementation and propose a way how to deal with the problem.
2. We suggest guidance for operating a large scope of information via framework that unites all ethical issues related to implementing of Big Data.

Empirical scientific survey

The sample of the survey consisted of specialists involved into Big Data analysis. We have collected over 300 responses, and after primary data processing, we ended up with 288 responses.

		Age groups			Mean value	N
		18-25	26- 35	36+		
Sex	Female	90	18	10	25	118
	Male	104	50	16	28	170

Table 3 Sample description: gender to age ratio 1

If we look closely to the results of the evaluation of the perceived importance of various factors of ethical framework (where 1 equals least important, 5 – most important) we see that all of them were indicated as those of primary (>4) to rather high (3,5) significance and those which should be thought of while applying Big Data Technologies.

We claim that no alterations in the framework are needed to be undertaken, while all theoretically outlined guidelines were identified as necessary for consideration while operating Big Data.

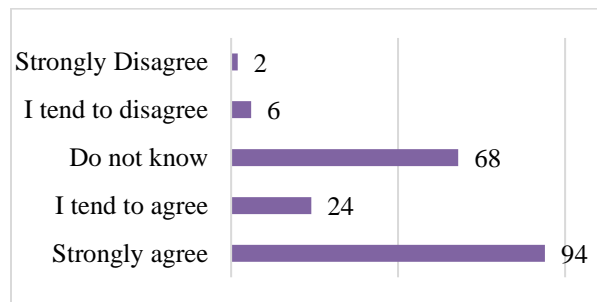
	Mean value of importance indicated by respondents (1-5)
Availability	4,1471
Long-term viability	3,966
Regulatory compliance	3,8495
Ownership	3,8349
Privileged user access	3,75
Data lifecycle	3,7427
Data confidentiality	3,7115
Identity	3,6683
Data segregation	3,5922
Privacy	3,5881
Data location	3,5049
Sensitivity of the data	3,3479
Reputation	3,3053
Recovery	3,301

Table 4 Evaluation of the perceived importance

The next survey question aimed on evaluating acceptable and unacceptable purposes of transferring data from one owner to another as perceived by respondents. We have supplemented the framework with new elements representing acceptable and unacceptable purposes of data transferring. The list of acceptable and unacceptable purposes of data transferring as perceived by absolute majority of respondents you may find in the table below.

If this is dictated by legislation	Acceptable
To investigate cases of alleged fraud linked to the website or violation of third party rights	Acceptable
For periodic information about products and services - implementation of mailing	Unacceptable
In order to offer services of third-party partners	Unacceptable
The transfer or acquisition of information about the browser history of client in order to respond to client's requests for various services	Unacceptable
Transfer of data to another owner for sale for profit	Unacceptable

Table 5 Acceptable and unacceptable purposes of transferring



Picture 1. All parties bear responsibility for the actions undertaken against data

This may seem obvious, still, degree of data openness is inversely proportional to what the individuals are willing to share. But what we have to contribute basing on the results of the survey is exactly the list of what is perceived and may be treated as public, freely shared and transferred, per contra, what seems to be private and unacceptable to transfer without the special permission.

Thus, sorting it by the order of decreasing degree of openness, the data can be distinguished between:

1. Public data
2. Freely disclosed data
3. Data which requires special permission for disclosure
4. Personal data
5. Confidential data

It was also shown, that the attitude towards undesirable and unpermitted transfer of personal data (such as contacts) is always negative and may cause worsening of perception of the company.

Results of data analysis signify that:

- respondents perceive acceptable only the personal data disclosure by an organization if it is prescribed by the law or if it may help to investigate cases of alleged fraud linked to the website or violation of third party rights;
- the unpermitted disclosure of personal data by an organization is perceived as an unacceptable practice if its purpose is profit, mailing or offering services of third parties;
- respondents believe that all parties bear responsibility for the actions undertaken against data. The unauthorized personal data disclosure may only result in negative reaction towards the organizations undertaking such actions from individuals.

Finally, we have analyzed the respondents' perception and definitions to such concepts as confidential data, personal data, data that requires special permission for disclosure, freely disclosed data and public data in an integrated framework for ethical Big Data Application.

Integrated Ethical Framework for Big Data Application

We have tested, estimated and discovered most vital issues, which should be considered while implementing Big Data. Thus, the integrated framework presented in this paper was thoroughly elaborated theoretically and carefully tested empirically.

All ethical concerns, which have to be discussed and taken into account while generating, managing, processing or analyzing big scopes of information, were allocated into two categories:

1. Social Security control related issues
2. Technological issues

Based on the theoretical and empirical surveys, we suggest allocation of twenty components of the ethical framework for the Big Data implementation, which is presented in the following table. Partially this

framework, or rather theoretical guidelines for it were presented in our paper “The ethics of Big data: analytical survey”¹.

Element	Operationalization of the element
	Social Security controls related
Identity	What is the relationship between one’s offline identity and one’s online identity?
Privacy	Who should control access to data? Transparency of the access
Ownership	Who owns data, has rights to transfer the data, and what are the obligations of people who generate and use that data?
Reputation	How can we determine what data is trustworthy? What are the possible ways to use data? How are we perceived and judged by using data
Data confidentiality	Is the information protected against unintended or unauthorized access? Is the data collected confidential, personal, freely disclose or public?
Availability	Is (or how often) information available for use by its intended users?
Data lifecycle	Right to be forgotten and to erasure
Sensitivity of the data	Is the data sensitive (medical patient records)?
Congruence with the law	All actions undertaken against data have to be undertaken in congruence with the state law
Transparency of the data politics	The companies’ politics about the data should be clear and transparent
	Technology related
Privileged user access	Who manage the data?
Regulatory compliance	Does one have permission to use?
Data location	Is the data stored in private storage?
Data segregation	Are the encryption schemes in place and tested?
Recovery	Will one have an ability to do complete restoration in case of disaster?
Long-term viability	Will the data be transferred in the event your provider ceases to exist, is acquired or contract ends?

¹ GIBER L. & KAZANTSEV N. (2015). “The ethics of Big data: analytical survey”. Cloud of science URL: <http://cyberleninka.ru/article/n/the-ethics-of-big-data-analytical-survey>

Data quality evaluation	The quality of the data should be thoroughly evaluated
Build-in privacy	The privacy and security technologies should be built into the products
Basic data analysis ethics	Conduct an accurate analysis of data, evaluate possible risks
Mutual benefit	Both organization and customers should benefit from the results of Big Data applications

Discussion

The topic of the ethical challenges of big data is usually discussed in the setting of the present concerns that call for urgent political and regulatory responses. We should stress that with the quickly growing scope and size of data that can be provided to businesses and researchers by Big Data technologies now, the concerns and dangers of its unethical usage are newly just being spoken and thought of. Hence, this paper, which specially aims at creating the ethical framework aimed at facilitating application of Big Data considering ethical codes in relation with the information security and privacy of personal data and technology related concerns, firstly addressed recent discussions of big data, by exploring the ethical matters rising from advancing knowledge production. The topics of privacy and security issues and data protection are discussed in the setting of big data implementation. Taking all that was discussed into consideration, advancing data driven knowledge technologies may be seen as a proof of a further dramatic transformation of the social order and more precisely an indication of the course of ongoing weakening of human autonomy.

Yet, it is tough to perceive ethical application of Big Data analysis at the aggregated level on which information is detached and anonymized, information on individuals is in its essence private and frequently sensitive, furthermore it may be claimed that there is direct impact of applied researches on individuals. For example, there is no doubt now that specialists in market researches currently more and more frequently resort to the assistance of “agents of influence”: specialists engaged in posting desirable feedback for managing the image of brand, what has its potential in changing consumer habits of individuals. Thus, according to the responses gathered, we claim that the respondents indicated none of the factors of the theoretical guidelines, which were revealed in the theoretical research, as non-significant, consequently all of the indicators will be included in the integrated ethical framework of Big Data implementation.

Next, the results of data analysis signify that respondents only perceive acceptable personal data disclosure by an organization if the law dictates it or if it may help to investigate cases of alleged fraud linked to the website or violation of third party rights.

Moreover, it was demonstrated that respondents believe that all parties bear responsibility for the actions undertaken against data and intended but unauthorized personal data disclosure may only result in negative reaction towards the organizations undertaking such actions from individuals.

Finally, we have analyzed the respondents’ perception and definitions to such concepts as confidential data, personal data, data that requires special permission for disclosure, freely disclosed data and public data. That gave us an opportunity to come to the conclusion that the degree of openness has a crucial impact on what people are willing to share, and more open the data is, the less individuals are willing to share.

Conclusion

As it was discussed, with the promptly growing volume of data provided to businesses, currently by implementing Big Data solutions, it may be of key significance to sustain an ethical framework for the Big

Data implementation. The paper puts forward an innovative framework for the Big Data implementation. This can serve as a stepping-stone from the theoretical perspective in the discussion of the ethical code of Big Data towards its practical implementation, what will prevent numerous risks and help to remain ethical while operation with big amounts of data.

For that motive, following research is expected to become vital for the experts in the fields of the data science and applied market studies, policymaking, and furthermore for a broader public, with the aim of alert about the settings of Big Data applications and their ethical consequences.

References

- Crawford, K. (2013). The hidden biases in big data. HBR Blog Network, 1.
- Crawford, K. (2014). The anxieties of big data. *The New Inquiry*, 30.
- Cukier, K., & Mayer-Schoenberger, V. (2013). "Rise of Big Data: How it's Changing the Way We Think about the World", *The Foreign Aff.*, 92, 28
- Davis, K. (2012). "Ethics of Big Data: Balancing risk and innovation". O'Reilly Media, Inc.
- Giber L. & Kazantsev N. (2015). "The ethics of Big data: analytical survey". *Cloud of science* URL: <http://cyberleninka.ru/article/n/the-ethics-of-big-data-analytical-survey> (accessed 22.05.2015)
- Peters, B. (2012). *The Age of Big Data*
- Puschmann, C., & Burgess, J. (2014). Big Data, Big Questions| Metaphors of Big Data. *International Journal of Communication*, 8, 20.
- Raicu, I. (2012). *Apps and Privacy: A Case Study*.
- Rayport, J. (2011). What big data needs: a code of ethical practices. 2013-8-29]. [http://www. technologyreview. com/news/424104/what-big-data-needs-a-code-of-ethical-practices](http://www.technologyreview.com/news/424104/what-big-data-needs-a-code-of-ethical-practices).
- Rayport, Jeffrey F. (2014) *What Big Data Needs: A Code of Ethical Practices* MIT Technology Review.
- Richards, N. M., & King, J. H. (2014). Big data ethics. *Wake Forest L. Rev.*, 49, 393.
- Rijmenam, Mark. (2013) *Big Data Ethics: 4 principles to follow by organisations*.
- Schneider, K. F., Lyle, D. S., & Murphy, F. X. (2015). *Framing the Big Data Ethics Debate for the Military*.
- Snijders, C. et al.(2012). "'Big Data': Big gaps of knowledge in the field of Internet". *International Journal of Internet Science* 7: 1–5.
- SO, K. (2011). "Cloud computing security issues and challenges". *International Journal of Computer Networks*, 11-14.
- Solove, D. J., Rotenberg, M., & Schwartz, P. M. (2003). "Information privacy law" , 268.
- Vayena, E., Salathé, M., Madoff, L. C., & Brownstein, J. S. (2015). Ethical challenges of big data in public health. *PLoS Comput Biol*, 11(2), e1003904.
- Velasquez M. et al. (1987) "What is ethics //Issues in Ethics"
- Volz D. (2014). *How Big Data Can be Used to Fight Unfair Hiring Practices*
- Wen, Howard.(2012) "Big Ethics for Big Data." *Data*. O'Reilly.