Towards Ethical Big Data Artifacts: A Conceptual Design

Olgerta Tona
Lund University
olgerta.tona@ics.lu.se

Ida Asadi Someh
University of Melbourne
ida.asadi@unimelb.edu.au

Kaveh Mohajeri
Lund University
kaveh.mohajeri@ics.lu.se

Graeme Shanks
University of Melbourne
gshanks@unimelb.edu.au

Michael Davern
University of Melbourne
m.davern@unimelb.edu.au

Sven Carlsson
Lund University
sven.carlsson@ics.lu.se

Miranda Kajtazi
Lund University
miranda.kajtazi@ics.lu.se

Abstract

Although Big Data generates many benefits for individuals, organizations and society, significant ethical issues are forcing governments to review their regulations so that citizens' rights are protected. Given these ethical issues and a gradual increase of awareness about them, individuals are in need of new technical solutions to engage with organizations that extract value from Big Data. Currently, available solutions do not adequately accommodate the conflicting interests of individuals and organizations. In this paper, we propose a conceptual design for an artifact that will raise awareness amongst individuals about Big Data ethical issues and help to restore the power balance between individuals and organizations. Furthermore, we set forward a design agenda outlining future activities towards building and evaluating our proposed artifact. Our work is grounded in discourse ethics and stakeholder theory and intertwined with the European General Data Protection Regulation (GDPR).

1. Introduction

Big Data as a phenomenon is attracting attention of both IS practitioners and IS scholars [1, 23]. Big Data is generated from a variety of sources including social media and other sensor-data sources [9]. When analyzed using advanced statistical techniques, big data generates significant value for organizations. Despite the many benefits of big data and analytics (e.g. improved national security, new approaches to medical research, better targeted services to consumers and more effective planning [15]) significant ethical issues are evident (e.g. breaches of personal privacy [5, 17],

discrimination, loss of freedom [20] and the move towards a surveillance society [28]).

Organizations are engaged in a balancing act between extracting value from big data and responding to regulatory and other stakeholder pressures to address ethical issues concerning the use of personal data. Google, for example, has changed the privacy policy for its services more than ten times since the start of 2013[10].

From a legal perspective, probably the most important change in data privacy regulation in the last two decades is the proposed European Union (EU) General Data Protection Regulation (GDPR) that will be enforced from 25 May 2018. GDPR aims to regulate and protect data privacy from data breaches for all EU citizens. Even organizations outside the EU that sell services to EU citizens will need to comply with this regulation. Non-compliance may result in harsh sanctions of up to 4% of annual global turnover of the organization, depending on the type of offence [8].

GDPR includes several requirements: individual consent must be concise and clearly written; the reasons for data collection and analysis must be provided to individuals within the consent process; individuals should be able to readily and easily withdraw their consent; individuals have the right to ask for an electronic copy of their personal data together with information regarding the processing and purpose of data collection and analysis from a controller (appointed by organizations); individuals can transfer their data from one controller [1] to another [8]. However, the vast majority of organizations that collect, store and analyse personal data of EU citizens,

---

[1] A data controller is "a person or entity which collect and process personal data" (http://ec.europa.eu/justice/data-protection/data-collection/index_en.htm)

Page 4523

HICSS

have yet to design a clear plan of what needs to be altered within the company's management of data privacy before the legislation is enforced, let alone developed solutions to effect the required changes. In fact, Gartner Inc. suggests that by 2018, more than 50% of organizations affected by GDPR will not be in full compliance with its requirements [7].

GDPR not only regulates organizational activities but also aims to empower individuals [8] by highlighting and increasing awareness about ethical issues that emerge during personal data collection and analysis. Although GDPR sets the legal parameters about data protection and individual rights, individuals need an understanding of who owns their data and who tracks their online activities before they can assert their legal rights.

Currently, there is a clear power imbalance between individual users and organizations as the key stakeholders involved with big data. Individuals frequently have little awareness of how their personal data will be used within the big data supply chain [15] and little control over what happens to their personal data [4]. Although there are some technological attempts to protect individuals from data harvesting (such as DuckDuckGo) we believe that these solutions typically do not offer a satisfactory solution for all stakeholders involved.

This paper reports on an ongoing research study whose goal at the current stage is to propose a tentative conceptual design of an artifact that enables a more ethical and fair treatment of all stakeholders involved in big data. We define clear objectives for the artifact and specify a design agenda as a means of building and evaluating the artifact. We contribute to knowledge by extending the research on ethical aspects of big data and move it towards the design of artifacts that meet the requirements of all stakeholders involved in big data.

We are guided by discourse ethics [18] and stakeholder theory [19] in specifying the objectives of our artifact and conceptualize them consistent with the GDPR. We argue that ethical discourse provides a useful theoretical base to inform the design of our artifact, as it provides a means for fair and rational discourse between individuals and organizations. We are also guided by stakeholder theory by recognizing the need to empower and raise awareness amongst individuals to ensure a fair and rational discourse. The artifact should encourage organizations to be more accountable for treating individuals fairly.

The paper is structured as follows. We first define big data and discuss the impact of big data on individuals, together with current technological solutions to ethical issues with big data. We then discuss relevant aspects of discourse ethics and stakeholder theory. We subsequently present the conceptual design of our proposed artifact based on discourse ethics theory, stakeholder theory and key aspects of the GDPR. A design science research agenda outlining future research activities involving building and evaluating the artifact is proposed and, finally, we conclude the paper by discussing implications of our design science research agenda for researchers and practitioners.

2. Background

Our aim is to propose a conceptual design for artifacts that enable a more ethical and fair treatment of all stakeholders involved in big data. To do so, we first discuss big data and its impact on individuals (currently the weakest stakeholder), highlighting three social processes associated with big data that impact individuals. We then discuss current technological solutions to these problems and highlight their strengths and weaknesses.

2.1. Big data and its impact on individuals

Big data is currently of great interest to both practitioners and researchers. Organizations are investing large amounts of money to acquire capabilities to store, manage and analyze big data. Big data potentially can create positive impacts for individuals, such as, enhanced search capabilities based on profiling and cures for terminal diseases. Although there are many benefits associated with big data, there are also many risks and costs. Examples include profiling of individuals that leads to discrimination, loss of privacy, and limiting individual's right to free choice and self-determination.

Big data is typically defined from a technological perspective, using Volume, Velocity, and Variety, the so called 3Vs [25]. Volume refers to the amount of data; velocity refers to the speed at which data is captured; and variety is the range of data types that are captured. Although the 3Vs of big data help to characterize the emerging technology, this perspective is silent on how the technology is used in different contexts and the consequences it creates for different stakeholders involved. To be able to understand the impact of big data on different stakeholders, we shift the focus from the technological characteristics of big data to the social processes that it enables. We argue that big data used by organizations gives rise to three main processes that affect individuals: (1) exploitation of individuals for data, (2) flow of individuals' data, and (3) targeting individuals with algorithmic decision-making.

First, big data originates from individuals. Individuals create the big data when they interact with online services, mobile devices, and sensory devices. Many big data applications harvest individuals for data, often collected from people without their knowledge or consent [28]. Terms and conditions for using big data services are often verbose, long, and not easily accessible or comprehended by the user. Individuals who do not agree with the terms and conditions are prevented from using the service. Although individuals may believe that using big data services (e.g. Facebook) are free of charge, they pay for using such services by providing their personal data. Facebook aims to increase individuals' time interacting with the platform in order to collect more data and monetize the data [28]. Such intentions are often hidden from the individual user.

Second, big data is not static, as it is not stored within the boundaries of one organizations. Rather big data is captured from individuals, shared and sold. From this perspective, big data has created a secondary market for many organizations, where individuals' data is extracted by tracking companies and shared with or sold to new intermediary entities such as data brokers, data aggregators and advertisers until all the value is exhausted [15]. Big data is essentially exchanged as a commodity. Many organizations have found this secondary market more lucrative than their primary business and push the boundaries to generate more value. This secondary market is largely hidden from individuals.

Third, individuals are not only exploited by organizations for data collection purposes, but also targeted by algorithmic decision-making. As data grows in size, algorithms dominate decision-making and human decision makers become unable to understand the inner logic leading to particular decisions. Algorithms can lead to discrimination when profiling individuals based on their sex, gender, race, and age. Algorithms, in most cases, are based on correlations and can lead to biases in the decision-making, particularly when working from the subjective and nuanced data generated by social media users. They not only predict the future but also regulate the future based on past behavior that in addition undermines individual's right to freedom of choice.

Overall, there are many benefits for individuals with big data, but there are also many risks and unintended consequences. In the domain of big data there is a major imbalance in power between organizations and individuals. As information privacy continues to be one of the most significant pressure points in the context of the digital world [26], individuals are giving away their valuable data for next to nothing and losing their privacy and freedom. They are in at risk of being taken advantage of by the big data services who dominate and control the market. Individuals need to become more aware of the risks and consequences of big data and to have more control over their data.

2.2. Current technological solutions

There are available technological solutions that attempt to address some of the ethical issues that arise from big data and its impact on individuals, such as, privacy controls in browsers and search tools like DuckDuckGo search engine. Browsers, such as Chrome, Firefox, and Safari have contributed to the increasing popularity of private browsing by incorporating privacy features in their browsers [2]. For example, the **Incognito mode** of Google Chrome represents a basic example of private browsing, which, when activated, does not log users' browsing history and cookies [3]. Despite the goodwill generated by protecting individual users from data harvesting, private browsing on Google Chrome suffers two main drawbacks. First, IP addresses are not totally invisible, allowing certain websites to identify the visitors [21]. Therefore, users are not fully protected from companies that may mine and use their data for different purposes. Second, users may experience lower efficiency during their browsing activities because the history, the sign-in information and the forms used are not saved for future use [6]. Alternatives like the Tor browser provide great protection but are even less efficient to use (see https://www.torproject.org).

DuckDuckGo (see https://duckduckgo.com) promises to protect users' data by offering an alternative search engine. Partly due to its inclusion as a search option in Safari [16] DuckDuckGo's usage has grown significantly with an average of 14 million searches per day and around 4 billion searches in total in 2016 [13]. Through DuckDuckGo, users can navigate the Internet without being identified by different websites. It means that users' data is neither collected nor shared and because of the lack of IP tracking, the advertisements are significantly reduced [27, 13]. However, we argue that this high security obstructs the effectiveness of customized services that for many people may be highly essential.

Anonymity is the key to current technological solutions that aim to protect users from data harvesting. Users' anonymity is preserved when data collection is prevented, e.g. in DuckDuckGo or Tor. However, a high degree of anonymity also leads to a lower service quality at the individual, organizational and societal level. Tackling ethical issues by positioning technologies at the extreme end of potential solutions

spectrum, we also lose the possibility to use users' data for the common good of the community (e.g. crime prevention), to customize services (e.g. recommendation systems), and to inform innovative ideas (e.g. potential start-ups).

These solutions amount at best to a patchwork of protections and certainly are a long way from creating a level playing field between individuals from who data is harvested, and organizations that profit from such data. What is actually missing in the present situation is an active communication between organizations and individual users regarding the collection and processing of personal data. On one hand, there are technical solutions that enable organizations to increase their profitability based on the extraction and use of personal data from individuals who are unaware of the big data process. On the other hand, other technological solutions that favor individuals by maintaining their anonymity but decrease the efficiency of an offered service and also hinder organizations to reach out to users as potential customers. To achieve a win-win situation, we argue that new technologies should encourage stakeholders to communicate with one another and should focus more on how to best accommodate their sometimes-conflicting needs and interests.

3. Theory

Two relevant bodies of theory for our study are discourse ethics and stakeholder theory. In this section, we discuss each and explain how they are relevant to our design agenda.

3.1. Discourse Ethics

It is clear from the preceding discussion that more attention needs to be given to ethical aspects of big data. There are major challenges in the design of big data IT artifacts that address ethical issues in big data in a meaningful and effective way. These ethical issues have been overlooked to a large extent previously.

From a design perspective, therefore, current big data artifacts suffer from 'low solution maturity' [11] when it comes to ethical concerns. To improve this situation, we argue that it is important to consider ethical theories and approaches that could support meaningfully engaging with ethical concerns when designing IT artifacts in the area of big data. In particular, we focus on 'discourse ethics' to inform the practice of ethics-driven big data artifact design. Discourse ethics is a specific approach to ethics that originated with the German sociologist and philosopher Jürgen Habermas. Discourse ethics is a

fairly recent, well-thought-out and coherent synthesis of some of the past most influential ethical theories and approaches. Mingers and Walsham [18] note that discourse ethics "is clearly Kantian in thrust, although with a very significant reorientation, but also sweeps in, to some extent, utilitarian and communitarian concerns" (p. 841). There are three specific reasons as to why discourse ethics is an appropriate choice to theoretically guide us in developing an ethics-driven big data artifact design agenda.

First, discourse ethics concerns the engagement of stakeholders in discourse about ethical issues. Ethical issues in the big data domain are heavily focused on the idea that many different types of interacting stakeholders are involved. As Mingers and Walsham [18] argue, "discourse ethics distinguishes moral issues that concern *everyone* involved in a particular situation" (p. 844). Mingers and Walsham [18] characterize such a quality in discourse ethics as *universalization*, and emphasize that discourse ethics therefore "pushes us to consider, and involve, as wide a range of stakeholders as possible in decisions and system designs" (p. 844). On this basis, discourse ethics also has a very close connection with the well-established body of theory known as 'stakeholder theory'. Whereas discourse ethics may provide the general philosophical basis and rationale for the idea of total stakeholder engagement, stakeholder theory informs us how to implement such an idea in the real world. Therefore, to develop a discourse ethics-driven artifact design agenda in the big data domain, we consider discourse ethics as the general ethical approach, and draw on stakeholder theory as a means of identifying and classifying different stakeholders involved in the big data artifact design [19]).

Second, the dialogic nature of discourse ethics enables the different values and conflicts of stakeholders in the big data domain to be discussed in rational and fair discourse. Ethical measures and norms cannot be imposed in a pre-established or pre-fixed form; rather, they must be continuously negotiated within rational and fair conversations amongst relevant stakeholders.

In contrast to many other (older) ethical theories and approaches, discourse ethics offers such a capacity in that "discourse ethics is … entirely procedural" ([18], p. 844). In other words, discourse ethics, due to advocating the notion of an *ideal speech situation*, is widely perceived as an ethical approach that "does not specify moral behaviors but only methods for agreeing upon them" ([18], p. 844).

Third, discourse ethics accommodates the practical issues that need to be considered within big data artifact design. In other words, such matters as efficacy, effectiveness and efficiency of big data

artifacts are also of high importance. One of the remarkable advantages of the discourse ethics approach is that it is not silent on practical issues. Indeed, discourse ethics goes beyond mere metaphysical and philosophical concerns, and recognizes the fact that there are always pragmatic issues and "questions that need to be settled through bargaining and even the exercise of strategic action" ([18], p. 844). This feature of discourse ethics therefore makes it very appropriate as a general ethical approach to support developing and pursuing a big data design agenda, which not only bears a social aspect, but also incorporates technological and practical aspects.

3.2. Stakeholder Theory

In our research, stakeholder theory provides a means to identify relevant stakeholders and define a typology of stakeholders based on their salience (importance) to a particular phenomenon [19]. In the context of big data, three relevant stakeholders are the individual, the organization and society. Three attributes of stakeholders that are relevant to the salience of the stakeholders are their *power* to influence the phenomenon, the *legitimacy* of their relationship to the phenomenon, and the *urgency* of their claim on the phenomenon.

We use stakeholder theory to argue that in the context of big data, organizations have high power, legitimacy and urgency and therefore high salience. Individuals and society generally have low power and urgency and therefore relatively low salience. However, in the case of the European Union, the introduction of the GDPR will substantially increase the power and therefore the salience of society. We use stakeholder theory to argue that we need to find a way to increase the power and urgency of individuals to increase their salience and thereby enable rational and fair conversations between organizations and individuals as required in discourse ethics.

4. A Conceptual Design

Based on discourse ethics and stakeholder theory, we build the conceptual foundations for an ethical big data artifact. Furthermore, we describe –at a conceptual level– our artifact as a social engagement platform and illustrate it through a use case scenario.

4.1. Foundations

Perhaps the most important aspect of the discourse ethics approach is that there must be a forum for different stakeholders involved in a situation to engage in a form of conversation or discourse, so that ethical norms and rules of behavior can be established on the basis of fair and rational argumentation and communication among all stakeholders. With big data, there are three major types of stakeholders that ought to engage in such a discourse. They are: (a) individuals, who are subject to big data collection and various impacts, (b) organizations that use, analyze, and monetize the big data collected from individuals, (3) societal actors who govern and regulate the developing interactions and technology.

It is clear that there is a power imbalance among individuals, organizations and society in the big data sphere, in terms of the extent to which they can exercise their rights and exert control. Drawing on stakeholder theory, individuals and society, while having a high level of legitimacy in big data, currently lack sufficient power and urgency to actually get engaged in big data processes. Considering Mitchell, et al. [19], one may classify individuals as a 'discretionary stakeholder' group which gets exploited by organizations and receives minimum support from societal actors. However, the current circumstance can change if individuals become aware of their own rights, so that the urgency of their needs and rights could be acknowledged. They could also become empowered and protected by societal actors, so that they could exercise their personal data rights more effectively. In other words, according to the stakeholder classification model of Mitchell, et al. [19], individuals, rather than being a 'discretionary stakeholder' group, could turn into a 'definitive stakeholder' group.

From a societal perspective, the EU has proposed a set of regulations to protect citizen's right, that is GDPR. We argue that some specific dimensions provided by GDPR can be effectively used to guide and structure a potential discourse among the stakeholders [7]. On the whole, the design of the artifact is motivated by giving individual stakeholders a voice in communicating with organizations (discourse ethics theory), creating a discourse structured and guided by GDPR, and consequently increasing the power and urgency of individuals in the discourse (stakeholder theory). The artifact thus provides a platform for fair and rational discourse between individuals, organizations and societal actors.

We expect individuals' perceptions to vary between individuals and across organizations even though GDPR, which sets the legal minimum requirements, applies in the same way across all organizations. Similar to Corporate Social Responsibility (CSR) – through which organizations moved beyond what was being dictated by environmental rules and regulations – GDPR represents not only a regulation to comply with

but also an opportunity to leverage key resources. Indeed, organizations may review their internal processes to further increase their data process transparency among their customers and clients as a way to gain both competitive advantage and reputation in regards to ethical behavior. This might affect in particular the direct communication with individuals when catering to their data privacy needs and requests.

4.2. A social engagement platform to create ethical discourse

We aim to design an artifact that facilitates the engagement and participation of stakeholders in pragmatic conversations. Social technologies could facilitate communication among individuals sharing their values, experiences and debating ideals, thus creating a public forum to address conflicting opinions.

The artifact we envisage is in the form of a social engagement platform that allows for creation of user-generated content and sets the stage for an open forum for ethical debate on the use and impact of big data [24]. Stakeholders (including individuals and organizations) can have explicit discussions about their ongoing daily concerns and together can define and agree on what is socially accepted.

The content generated by the stakeholders will be structured so that individual contributions relate to one organization (for example Google, Facebook, a bank etc.), along a number of dimensions. Organizations will have the opportunity to respond to individual contributions. This is similar to such systems as TripAdvisor, where each contribution from an individual relates to one hotel (for example) along a number of dimensions.

Figure 1 outlines the main concepts, functions and purpose of our proposed artifact. It represents a high-level conceptual design of a social engagement platform. Being a shared platform, individual users and other societal actors can interact regarding the ethical behavior of a specific organization. At the same time, besides the collective voice of individual users, we expect organizations to actively participate in the discourse, primarily to secure their reputation and also to be informed about individual concerns regarding the way the organization is handling their personal data. This in turn may influence organizational decisions about how to design future artifacts of their own entailing ethical considerations, such as, privacy by design.
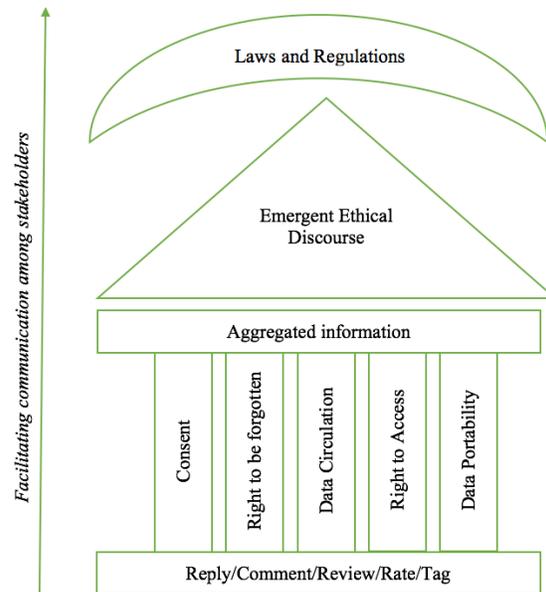

Figure 1. A conceptual design of a social engagement platform that facilities the discourse among stakeholders

The social engagement platform will afford certain features to activate the communication channel among stakeholders. The participants will interact by posting on a specific topic, commenting on other posts, replying to specific individuals or organizations, invite other actors in the discussion by tagging them, and rating the performance of an organization (e.g. on a Likert scale) based on five dimensions. The five dimensions, which derive from GDPR, are at the core of this conceptual design in order to lead and structure the main discourse among the stakeholders.

We now describe the dimensions from the GDPR [8] that help structure discourse on ethical issues in big data, with a particular focus on personal data. The five dimensions are consent, the right to be forgotten, the right to access, data portability and data circulation.

- Consent- information about consent within an organization, concerning the consent process, the ease of understanding and providing consent. Ratings can be provided for the clarity of consent information and the purpose of data collection and subsequent analysis and monetization. For organizations, it is important to recognize that implied consent is no longer sufficient.
- Right to be Forgotten – information about the process of deleting personal data based on individual request. In particular, the possibility of deletion and the ease with which personal data can be deleted. From an organizational perspective, this

indicates the availability and the role of a data controller to take charge of the process.

- Right to Access – information about the process of accessing personal data, obtaining confirmation that personal data is being processed and how it is being used. This dimension allows individuals to discuss how easy it is to contact organizations and request permission to access personal information, and to obtain an electronic copy of it.
- Data Portability –information about how personal data is transmitted between organizations. In particular, whether personal data is processed on the basis of consent, and the rights of individuals to withdraw the provided information their consent for their personal data to be transmitted between organizations.
- Data Circulation –information about how organizations transfer data to other countries that do not ensure an 'adequate level of protection'. Under the GDPR a Commission, which currently has approved very few countries, must assess the level adequacy of protection afforded by other countries. This information should be made available to individual users, thereby increasing individual's awareness and allowing organizations to increase their transparences.

Even though this is a non-exhaustive list, we believe that these crucial dimensions, through which a discourse should be structured, will keep the stakeholders' attention on the core ethical issues. When individuals provide content to the artifact, based on their expertise and experience, we argue that their power will increase. Rather than having a single individual who expresses concerns over the use of big data by organizations, the artifact will instead allow a larger group of individuals to participate and aggregate their experiences and concerns in relation to the ethical discourse.

Through one or many dimensions, stakeholders will comment on organizational actions/behavior based on their own interpretation. The content should purely be guided by their observation and judgment during the interaction with an organization and/or individual. For instance, individual users can share their experiences on how well- and intelligibly- written was a consent regarding data policy of a particular organization (Consent dimension) or how time-consuming it was to obtain a list of the companies to which personal data has been transmitted through that particular organization (Data Portability dimension). In other words, comments, reviews, replies and ratings will be centered around one of the dimensions, such as consent. On the whole, the discourse will be more focused, will cover a broad spectrum of ethical

concerns, and will be in line with the current regulations.

Considering the large amount of individual comments that a specific dimension can receive and the time it takes for an individual or organization to consume that information, the platform should have the capability of aggregating user-generated content at a higher level. Two main benefits are foreseen. First, by a quick scanning, more individuals will become aware of big data issues concerning a specific organization. Second, it gives organizations an incentive to actively participate in the discourse in order to protect their ethical behavioral image regarding big data.

We strongly believe that the communication among stakeholders will be facilitated and, potentially, based on discourse ethics, will lead to an 'ideal speech'. Consequently, the power between organizations and individuals, regarding the use personal data, may be balanced.

We intend, as future work, to translate the above dimensions and features to technical properties specific for our artifact. By designing a prototype, based on discourse ethics, stakeholder theory and the dimensions generated from the GDPR, we can demonstrate not only the advantages of our solution against others in tackling the ethical issues but also how ethics can be incorporated in a technology in line with laws and regulations. Further, we can show how a technology can reshape organizational activities, technologies and potentially support individual decision-making regarding a choice about a service from a diverse set of organizations. To elaborate on the design of such an artifact we present a use case in the next section.

4.3 A use case for the proposed artifact

Alpha Car Insurance has a good reputation for offering competitive premiums to its customers. This is largely due to Alpha's advanced analytical capabilities, which it employs to design predictive models for measuring customers' risk. These capabilities and models allow Alpha to price premiums that are attractive to the customer and profitable to Alpha. Customer data is collected from GPS-enabled tracking devices fitted to each customer's vehicle, and augmented with personal and behavioral data extracted from external sources, such as social media or third-party data brokers.

In the social engagement platform, as shown below, Alpha has relatively high ratings when it comes to protecting customers' data rights:

Consent: 4.3/5
Right to be forgotten: 3.8/5
Right to access: 4.1/5
Data Portability: 4/5
Data Circulation: 4.2/5

Alpha also monitors and responds when necessary to discussions on the social engagement platform. In a recent discussion, Alpha noticed that one individual has expressed considerable concerns about how Alpha acquires personal and behavioral data and, moreover, how that data is processed and used.

Smith: There have been rumors around that Alpha is profiling people based on religion to have a better prediction of alcohol consumption. According to me, this is unacceptable. I have been trying to reach the insurance company several times, through the contact information, but I always get an automatic email. (Smith also provides a very low rate to Right to Access: 1/5)

Eldin123: @Smith, you should wake up. All insurance companies use your personal data. That is why I have closed all my social media accounts.

Alpha: @Smith we are very sorry for any inconveniences. Your request sent via email is under review and we will be able to provide you with an electronic copy of your data as soon as possible. We do collect and process data in line with the consent that you have provided. We would really appreciate that once you have obtained the requested information, we can call you and discuss further your concerns.

After Smith got the electronic copy of his data, he decided to exert his right to be forgotten by deleting his data residing at Alpha. Losing one customer, and, moreover, risking the reputation of being an ethical company, forced Alpha to review its processes for a quicker response time to its customers.

5. A Design Science Research Approach and Future Steps

A design science research (DSR) approach is appropriate for this study as it provides the necessary framework and activities to design, build and evaluate an artifact [14, 12]. The artifact itself can take different forms such as constructs, models, methods and instantiations [14, 12, 11]. This paper represents an initial study towards a larger research project whose focus will be to further design, construct and evaluate an instantiation (i.e. a prototype) that addresses the ethical issues arising from the use of big data by individuals and organizations.

The *first activity* in the DSR approach is to identify the problem by understanding the current environment [12] because that provides the key for designing

relevant and effective artifacts [14]. We have already conducted this activity, in Section 2 and its subsections, through exploring the problem space from both academic and practical perspectives. We have argued that the use of big data has shifted the power balance towards organizations, and that individual users lack awareness and control of their personal data.

The *second activity* of the DSR approach is to define the solution objectives [11]. In this paper, we have reviewed some current technologies that aim to address the problem and identified their shortcomings. Furthermore, we have described the main objectives of a desirable artifact: to facilitate the communication among stakeholders towards an emergent ethical discourse. Drawing on discourse ethics, stakeholder theory and GDPR, we presented a tentative conceptual design of the artifact, in section 4.2.

As pointed out, we, in our research, have so far addressed the first and second DSR activities. In the following paragraphs, we outline the remaining DSR activities as an agenda for future research [22, 11, 14]:

The third activity of the DSR approach is to design and develop the artifact. In detailed design, we will translate the main GDPR concepts to artifact properties A prototype artifact will be developed using appropriate technology.

The fourth activity of the DSR approach is to evaluate the artifact by assessing its effectiveness and efficiency. There are different evaluation methods that can serve the purpose of understanding if our prototype is a solution to the identified problem space [12]. We plan to have multiple ways of evaluating. From an experimental approach, we will assess the artifact for its usability, functionality and accuracy. We will create use-case scenarios and assess the performance of the performs. Furthermore, we will use observational techniques and 'think-aloud' methods to investigate how potential users (representatives from organizations, regulators and individuals) will engage with it.

The fifth activity of the DSR approach is to communicate the findings of the study to both the academic and practitioner communities. Moreover, our findings will be communicated to the teams responsible for the GDPR at a country level to ensure that our results are in line with the new EU regulation.

6. Conclusion and Implications

Ethical collection, storage and analysis of personal data is a critical big data concern among individuals and governments aiming to protect citizens' rights. Unsurprisingly, individual users have turned their attention towards technologies that maintain their

anonymity. However, current technologies that aim to tackle the problem are lagging behind in accommodating conflicting interests among individuals and organizations. This paper presents the initial phases of designing an artifact that can meet the requirements and constraints of ethical issues in big data.

We propose an artifact that provides a shared communication platform where individuals and organizations can interact and engage. The discourse is structured according to five specified dimensions based on the GDPR. Facilitating communication will potentially lead to fair and rational discourse, or 'ideal speech' between stakeholders [18]. The artifact should make individuals more aware of ethical issues in big data and better balance power between organizations and individuals regarding the use of personal data.

There are several implications from the design of the artifact. First, it is an improvement to technological solutions addressing these concerns, as it provides a platform for fair and rational discourse between individuals and organizations. Second, it provides a solution that is in line with GDPR guidelines. Thus, we showcase how governmental regulations can be embedded in the design of new artifacts. Third, through this artifact, the ethical positions of both individuals and organizations will emerge and be shared among the stakeholders. Fourth, through the discourse that will take place organizations will get feedback, input and motivation for privacy by design efforts. They will take account of the discussion and potentially take actions to address individual's concerns highlighted in the discourse.

Acknowledgements

7. References

[1] A. Abbasi, S. Sarker and R. Chiang, "Big Data Research in Information Systems: Toward an Inclusive Research Agenda", Journal of the Association for Information Systems, 17 (2), 2016, pp. 3.

[2] G. Aggarwal, E. Bursztein, C. Jackson and D. Boneh, *An Analysis of Private Browsing Modes in Modern Browsers*, *USENIX Security Symposium*, 2010, pp. 79-94.

[3] Casey, "Browse in Private with Incognito Mode", 2017, Retrieved from https://support.google.com/chromebook/answer/95464?hl=en&ref_topic=2586065

[4] R. Clarke, "Big Data, Big Risks", Information Systems Journal, 26 (1), 2016, pp. 77-90.

[5] S. Furnell and N. Clarke, "Power to the People? The Evolving Recognition of Human Aspects of Security", computers & security, 31 (8), 2012, pp. 983-988.

[6] X. Gao, Y. Yang, H. Fu, J. Lindqvist and Y. Wang, *Private Browsing: An Inquiry on Usability and Privacy Protection*, *Proceedings of the 13th Workshop on Privacy in the Electronic Society*, ACM, 2014, pp. 97-106.

[7] Gartner, "Focus on Five High-Priority Changes to Tackle the Eu Gdpr. White Paper by Bart Willemsen at Gartner Inc.", 2016, Retrieved from https://www.gartner.com/doc/reprints?id=1-3V903Y9&ct=170309&st=sg

[8] GDPR, "Eu General Data Protection Regulation", 2016, Retrieved from http://www.eugdpr.org/eugdpr.org.html

[9] P. B. Goes, *Big Data and Is Research*, SOC INFORM MANAGE-MIS RES CENT UNIV MINNESOTA-SCH MANAGEMENT 271 19TH AVE SOUTH, MINNEAPOLIS, MN 55455 USA, 2014.

[10] Google Privacy and Terms, "Updates: Privacy Policy", 2017, Retrieved from https://www.google.com/intl/en/policies/privacy/archive/

[11] S. Gregor and A. R. Hevner, "Positioning and Presenting Design Science Research for Maximum Impact", MIS quarterly, 37 (2), 2013, pp. 337-355.

[12] A. R. Hevner, S. T. March, J. Park and S. Ram, "Design Science in Information Systems Research", MIS Quarterly, 28 (1), 2004, pp. 75-105.

[13] J. Keane, "Duckduckgo, the Privacy-Focused Search Engine, Hits 14 Million Searches in a Day", 2017, Retrieved from https://www.digitaltrends.com/web/duckduckgo-14-million-searches/

[14] S. T. March and G. F. Smith, "Design and Natural Science Research on Information Technology", Decision support systems, 15 (4), 1995, pp. 251-266.

[15] K. E. Martin, "Ethical Issues in the Big Data Industry", MIS Quarterly Executive,, 14 (2), 2015, pp. 67-85.

[16] C. Miller, "Duckduckgo Has Grown 600% since Apple Made It a Search Option (and Nsa Revelations)", 2015, Retrieved from https://9to5mac.com/2015/06/16/duckduckgo-growth-ios/

[17] C. L. Miltgen and H. J. Smith, "Exploring Information Privacy Regulation, Risks, Trust, and Behavior", Information & Management, 52 (6), 2015, pp. 741-759.

[18] J. Mingers and G. Walsham, "Toward Ethical Information Systems: The Contribution of Discourse Ethics", Mis Quarterly, 34 (4), 2010, pp. 833-854.

[19] R. K. Mitchell, B. R. Agle and D. J. Wood, "Toward a Theory of Stakeholder Identification and Salience: Defining the Principle of Who and What Really Counts", Academy of management review, 22 (4), 1997, pp. 853-886.

[20] S. Newell and M. Marabelli, "Strategic Opportunities (and Challenges) of Algorithmic Decision-Making: A Call for Action on the Long-Term Societal Effects of 'Datification'", The Journal of Strategic Information Systems, 24 (1), 2015, pp. 3-14.

[21] K. Pathak, "Is Chrome's Incognito Mode Really Private? 4 Things to Know About It", 2015, Retrieved from http://www.guidingtech.com/33828/chromes-incognito-mode-private/

[22] K. Peffers, T. Tuunanen, M. A. Rothenberger and S. Chatterjee, "A Design Science Research Methodology for Information Systems Research", Journal of management information systems, 24 (3), 2007, pp. 45-77.

[23] G. Phillips-Wren, L. S. Iyer, U. Kulkarni and T. Ariyachandra, "Business Analytics in the Context of Big Data: A Roadmap for Research", Communications of the Association for Information Systems, 37 (1), 2015, pp. 448-472.

[24] A. C. Powers, "Social Networking as Ethical Discourse: Blogging a Practical and Normative Library Ethic", Journal of Library Administration, 47 (3-4), 2008, pp. 191-209.

[25] P. Russom, "Big Data Analytics", TDWI best practices report, fourth quarter, 192011, pp. 40.

[26] H. J. Smith, S. J. Milberg and S. J. Burke, "Information Privacy: Measuring Individuals' Concerns About Organizational Practices", MIS quarterly1996, pp. 167-196.

[27] L. Whitney, "Search Engine Duckduckgo Blocked in China", 2014, Retrieved from https://www.cnet.com/news/search-engine-duckduckgo-now-blocked-in-china/

[28] S. Zuboff, "Big Other: Surveillance Capitalism and the Prospects of an Information Civilization", Journal of Information Technology, 30 (1), 2015, pp. 75-89.