Communications of the Association for Information Systems

Volume 44 Article 34

5-2019

Ethical Issues in Big Data Analytics: A Stakeholder Perspective

Ida Someh
The University of Queensland, i.asadi@business.uq.edu.au

Michael Davern
The University of Melbourne

Christoph F. Breidbach
The University of Melbourne

Graeme Shanks
The University of Melbourne

Recommended Citation

Someh, I., Davern, M., Breidbach, C. F., & Shanks, G. (2019). Ethical Issues in Big Data Analytics: A Stakeholder Perspective. Communications of the Association for Information Systems, 44, pp-pp. https://doi.org/10.17705/1CAIS.04434

Communications of the Association for Information Systems

Research Paper DOI: 10.17705/1CAIS.04434 ISSN: 1529-3181

Ethical Issues in Big Data Analytics: A Stakeholder Perspective

Ida Someh
UQ Business School,
The University of Queensland
i.asadi@business.uq.edu.au

Michael Davern
Department of Accounting,
Faculty of Business and Economics
The University of Melbourne

Christoph F. Breidbach
School of Computing and Information Systems
The University of Melbourne

Graeme Shanks
School of Computing and Information Systems
The University of Melbourne

Abstract:

Big data analytics is a fast-evolving phenomenon shaped by interactions among individuals, organizations, and society. However, its ethical implications for these stakeholders remain empirically underexplored and not well understood. We present empirical findings from a Delphi study that identified, defined, and examined the key concepts that underlie ethical issues in big data analytics. We then analyze those concepts using stakeholder theory and discourse ethics and suggest ways to balance interactions between individuals, organizations, and society in order to promote the ethical use of big data analytics. Our findings inform practitioners and policymakers concerned with ethically using big data analytics and provide a basis for future research.

Keywords: Analytics, Big Data, Ethics, Delphi Study.

This manuscript underwent peer review. It was received 11/17/2017 and was with the authors for 5 months for 2 revisions. Oliver Müller served as Associate Editor.

1 Introduction

Big data analytics uses algorithms to analyze large and complex data sets in order to uncover patterns, correlations, and other insights from data (Martin, 2015). While one can apply big data analytics in numerous ways to gain economic and social value such as improved healthcare (Boyd & Crawford, 2012; Murdoch & Detsky, 2013), public safety and security (Chen, Chiang, & Storey, 2012), and service innovations (Loebbecke & Picot, 2015), it has recently received criticism for having unethical consequences for various stakeholders (e.g., Zuboff 2015; Wigan & Clarke, 2013). Privacy breaches, extensive individual profiling, or discrimination against customers represent some concerns that have been raised publicly (Zwitter, 2014). These concerns show a conflict in stakeholder values whereby organizations' interests and incentives do not align with individuals' and society's interests and incentives (Markus & Topi, 2015). Consequently, we lack understanding about how one could evenly distribute big data's benefits and costs across stakeholder groups.

To address the value conflict inherent in big data analytics applications, we adopt a stakeholder perspective (Mitchell, Agle, & Wood, 1997) to analyze the inter-relationships among various stakeholders. We do so for two reasons. First, we follow Mingers and Walsham (2010), who have encouraged IS researchers to use discourse ethics to resolve ethical dilemmas in information systems. Discourse ethics proposes that morality emerges from equitable debates among stakeholders when an ideal-speech situation among stakeholders exists (i.e., stakeholders are on a "level playing-field"). Second, by focusing on stakeholders, we respond to Markus (2015) who has called for IS researchers to explore the consequences of big data analytics for different stakeholders and Newell and Marabelli (2015) who have suggested that understanding the impact of corporations' "non-responsible" (p. 9) use of big data on individuals and society represents a key research priority. While recent research has raised issues with ethics of big data analytics (Ananny, 2016; Boyd & Crawford, 2012; Crawford, Miltner, & Gray, 2014; Ekbia et al., 2015; Lyon, 2014; Markus, 2015; Martin, 2015; Richards & King, 2014; Wigan & Clarke, 2013; Yoo, 2015; Zuboff, 2015; Zwitter, 2014), it has not adopted a stakeholder perspective and explored the inter-relationships between big data analytics stakeholders.

In this paper, we first develop theoretical concepts that explain ethical issues for stakeholder groups involved in big data analytics (namely, individuals, organizations, and society). We collect data using a Delphi study that adopts an explicit stakeholder perspective and analyze the data using the Gioia method to develop "theoretical concepts" (Gioia, Corley, & Hamilton, 2012). Second, we help to satisfy the ideal-speech situation among big data analytics stakeholders and encourage communicative and moral action (Habermas, 1992). To do so, we use the theoretical concepts related to each stakeholder group along with stakeholder theory (Mitchell, Agle, & Wood, 1997) to explore the inter-relationships between the stakeholders. Thus, we explain the "salience" of particular stakeholders in the big data analytics context. Based on stakeholder salience, we suggest ways to balance interactions between the different stakeholder groups.

We contribute to both theory and practice. Theoretically, we provide a stakeholder perspective on big data analytics. In doing so, we develop definitions for theoretical concepts that underpin ethical issues for each stakeholder group. We draw on empirical data to identify the dimensions that aggregate to form these concepts and provide evidence for their salience and relative importance to different stakeholder groups. Practically, by considering the theoretical concepts and stakeholder analysis we develop, individuals can learn about how big data analytics influences their lives and how they can empower themselves against organizations' control, monitoring, and manipulation. Further, organizations can use our insights to better safeguard themselves against the ethical and reputation risks in big data analytics and encourage them to use data in a responsible manner. Our insights can also empower societal agents to better protect citizens and regulate the way in which organizations use big data analytics.

The paper proceeds as follows. In Section 2, we review the recent literature on big data analytics, introduce discourse ethics as our underlying ethical philosophy, and highlight the need for a stakeholder perspective on big data analytics. We then use stakeholder theory to identify relevant stakeholders (individuals, organizations, and society) and highlight that the current literature does not comprehensively identify and explain the ethical issues for each stakeholder group. In Section 3, we explain our Delphi study and concept-development approach and, in Section 4, discuss the most important theoretical concepts for each stakeholder group. In Section 5, we analyze the concepts using stakeholder theory to illuminate how stakeholder interactions might enhance how they ethically use big data analytics. Finally, in Section 6, we suggest future research directions and conclude the paper.

2 Research Background

2.1 Big Data Analytics

To date, big data analytics has predominantly been conceptualized using technological attributes such as volume, variety, and velocity (the 3 Vs) or the power of its underlying algorithms in generating insights (McAfee & Brynjolfsson, 2012; Russom, 2011). However, such a technological focus limits our understanding of big data analytics as a socio-technical phenomenon that affects different stakeholders (Ananny, 2016; Crawford et al., 2014). Based on analyzing the emerging literature (Constantiou, Kallinikos, & Kallinikos, 2015; Ekbia et al., 2015; Galliers, Newell, Shanks, & Topi, 2017; Marjanovic & Cecez-Kecmanovic, 2017; Markus, 2017; Zuboff, 2015), we complement the technological view by identifying three social processes that target and influence individuals: 1) data sourcing, 2) data sharing and 3) algorithmic decision making. First, many big data applications exploit individuals for data-collection purposes (Zuboff, 2015). Organizations and government agencies use a "catch-all-you-can" approach to collect maximum data from individuals (Yoo, 2015). This approach quantifies individuals' everyday life (Spiekermann & Korunovska, 2017), primarily to benefit the organization doing the analytics. Second, data harvested from individuals travels from one organization to another until they exhaust its value (Barocas & Nissenbaum, 2014; Martin, 2015). This logic has created a secondary market for organizations to sell or share customer data. Individuals cannot see the purpose behind data-driven services to extract and share customer data (Barocas & Nissenbaum, 2014). Third, organizations use algorithms to profile individuals—sometimes inadvertently based on their race, ethnic group, gender, and social and economic status—and restrict their options and choices (Ananny, 2016; Loebbecke & Picot, 2015; Madsen, 2015), which raises wider ethical questions about how markets operate in a fair and free manner and questions that pertain to freedom of choice for individuals (Crawford et al., 2014). Put differently, how can individuals be free if they are under the control and surveillance of algorithms that seek to influence their decisions (Zuboff, 2015)?

In this paper, we focus on big data analytics as organizations apply it to customers, particularly when organizations use big data analytics to offer services and products to individuals. We argue that ethical issues arise when organizations collect, analyze, share, and/or sell individuals' data without individuals' genuine consent or awareness (Barocas & Nissenbaum, 2014; Solove, 2013). In Sections 2.2 and 2.3, we draw on discourse ethics and stakeholder theory to address these challenges.

2.2 Discourse Ethics

Ethics concerns questions about how people should act and what constitutes truthful behavior (Lewis, 1985). While a number of traditional approaches to ethics exist (e.g., Utilitarian, Kantian, and Aristotelian), each has limitations and criticisms (Mingers & Walsham, 2010). For example, with the utilitarian perspective, one cannot easily identify the consequences that an action will have in today's world, and choosing the common good for the majority could discriminate against minorities in human societies. Here, we use discourse ethics as a guide to analyze the ethics of big data analytics. Mingers and Walsham (2010) introduced discourse ethics to the IS literature and advocated for its potential to address ethical issues that concern IS researchers and practitioners. Discourse ethics, which originated with Habermas (1992), represents a recent theory that synthesizes traditional—particularly Kantian and Utilitarian—ethics theories. Discourse ethics mainly focuses on morality and involves the notion of universalism (Mingers & Walsham, 2010). Universalism means that moral principles go beyond one context or community and concern actions that are equally good for everyone.

The discourse process, or the process of communicative action (Habermas, 1990), lies at the center of discourse ethics. Habermas argues that stakeholders can best achieve pragmatic, ethical, and moral norms through a debate process (Habermas, 1990). He argues: "Only those norms can claim to be valid that meet (or could meet) with the approval of all affected in their capacity as participants in a practical discourse" (Habermas, 1990, p. 66). In this way, ethical norms cannot pre-exist or be imposed; rather, relevant stakeholders must fairly debate them. The ethical discourse will emerge from the actual debate between the stakeholders who should continuously renegotiate it over time. This requires the existence of an ideal speech situation, which means all stakeholders should be able to participate equally in a discourse and freely question, claim, or express their attitudes (Mingers & Walsham, 2010).

We use discourse ethics as our overall ethical framework and argue that ethical big data analytics will emerge from stakeholder's engaging with and creating ethical discourse. However, whether stakeholders create such a discourse depends on equality among them and their satisfaction with the ideal speech situation. Since discourse ethics does not explain how to identify stakeholders and their salience, we draw on stakeholder theory (Mitchell et al., 1997) to identify, classify, and analyze the stakeholders and their inter-relationships.

2.3 Stakeholder Theory

We use stakeholder theory (Mitchell et al., 1997) to generate a typology of different stakeholders involved in big data analytics and to understand their salience. We take a broad view on stakeholders (Freeman 1984) and define them as any group or individual who can affect or be affected by big data analytics. Three attributes of stakeholders pertain to their salience: power, legitimacy, and urgency (Mitchell et al., 1997). Power refers to the extent to which a stakeholder can impose their will in a relationship, legitimacy to the extent to which a stakeholder's actions are desirable in a social system, and urgency to the extent to which stakeholder claims call for immediate action.

We identify three main stakeholder groups in big data analytics: individuals, organizations, and society (Markus & Topi, 2015). First, big data analytics originates from individuals as they interact with digital technologies that can track their behavior (Derikx, de Reuver, & Kroesen, 2016; Newell & Marabelli, 2015). Organizations use the resulting data to identify patterns and relationships for economic value (Newell & Marabelli, 2015). Second, organizations control big data (Crawford et al., 2014). They analyze big data to make decisions that impact individuals and broader society and share and sell data about individuals (Martin, 2015). Third, government agencies and societal authorities have the responsibility and oversight to control, govern, regulate, and shape big data analytics (Metcalf & Crawford, 2016). Thus, big data analytics constitutes a socio-technical phenomenon that the interactions between individuals, organizations, and society create, manage, and shape.

We view big data analytics as interactions among stakeholders (individuals, organizations, and society) (Zuboff, 2015). The various interactions between stakeholders may not equitably distribute big data analytics' costs and benefits (Markus & Topi, 2015). Organizations that develop and deploy the technology dominate the interactions, and individuals and society incur the costs and (both positive and negative) consequences from the interactions. The often non-reciprocal character of these interactions leads to ethical concerns or dilemmas for the different stakeholders.

While recent literature has raised concerns about the ethics of big data analytics, it does not comprehensively consider all its different stakeholders (Markus & Topi, 2015) and how they can engage with one another equitably. To perform a stakeholder analysis and understand stakeholder salience, we need to rigorously identify the concepts that underlie ethical issues for individuals, organizations, and society. In Section 3, we describe our Delphi study (Dalkey & Helmer, 1963) and the approach we used to develop theoretical concepts (Gioia et al., 2012). These concepts form the basis of our stakeholder analysis in which we work toward developing an appropriate ethical discourse between stakeholders.

3 Research Method

The ethics of big data analytics represents an emerging area of research in IS and, thus, researchers have conducted little empirical work on the topic. In this study, we combine two exploratory research techniques to develop theoretical concepts that explain ethical issues in big data analytics: 1) Delphi method to solicit opinions from an expert panel and 2) the Gioia method to analyze qualitative data and develop concepts. In analyzing and interpreting our data, we examined issues strictly from the relevant stakeholder's perspective.

3.1 Data Collection Using the Delphi Method

When one knows little about a phenomenon and its future implications, exploratory research using the Delphi method represents an appropriate research strategy (Paré, Cameron, Poba-Nzaou, & Templier, 2013). The Delphi method involves "systematically soliciting, organizing and structuring opinions on a particularly complex subject matter from a panel of anonymous experts until a consensus is reached on the topic or until it becomes evident that further convergence is not possible" (Anderson, Rungtusanatham, & Schroeder, 1994, p. 478). The Delphi study approach involves four important characteristics: 1) the researcher purposefully selects a panel of experts to provide their opinions, 2) the experts remain anonymous to each other to guard against biases and personal influences, 3) the expert panel uses moderated communication to manage feedback and develop consensus among the expert panel, and 4) the researcher uses multiple opinion-seeking rounds to iterate the decision-making process, which allows participants to shape their own opinions by reflecting on other members' opinions (Worrell, Di Gangi, & Bush, 2013). One needs to carefully design Delphi studies to ensure rigor. In Appendix A, we assess the rigor of our Delphi study using the checklist that Paré et al. (2013) provide.

3.2 Data Analysis Using the Gioia Method

The Gioia method combines data-driven concept formation (induction) with input from researchers and existing literature (abduction) to develop concepts that best explain a particular phenomenon (Gioia et al., 2012). We analyzed the data using a three-stage process following Gioia et al. (2012). In the first stage, we used "induction" as the logical data-analysis process; that is, we developed first-order categories from the data by adhering to participants' terms and wording. In the second stage, we used "abduction"; that is, we relied on our knowledge and existing literature to analyze and develop concepts that explained the data. We re-analyzed the data by acting as knowledgeable agents and using researcher-centric concepts. Focusing on the deep structure underlying the first-order categories and the similarities and differences between them, we reduced the first-order categories to more abstract second-order themes. In the third stage, we investigated the possibility of distilling the second-order themes into aggregate dimensions. We categorized the second-order themes and created aggregated dimensions for the issues related to each stakeholder. We created data structures (see Appendix B) that connected first-order concepts to second-order themes and aggregate dimensions. Using the data structures, we revalidated the final concepts back to the underlying data and established a clear connection between data, the emerging concepts, and the aggregate dimensions. By keeping the voices of both informants and researchers, we could rigorously develop detailed and accurate definitions of concepts from the data.

3.3 Delphi Panel

Our Delphi panel comprised 34 experts who had knowledge of big data analytics and its ethical implications. Including a variety of experts with complementary viewpoints represents a common best practice, and prior Delphi studies have advocated it (e.g., Kiel, Lee, & Deng, 2013; Schmiedel, vom Brocke, & Recker, 2013). We purposefully selected a variety of disciplinary perspectives and roles, including academics, professionals, and social activities. We required academics to have active engagement in research in the big data analytics area and its implications for the three stakeholders. We required practitioners to have a senior position or key role in organizational big data analytics initiatives, which required them to know about project risks and the regulatory environment. We required the social activists to have an active role in bringing about social change with respect to data. To build the expert panel, we initially identified 81 experts (academics, professionals, and social activists) and approached them for first round data collection. Of the initial 81, 40 experts agreed to participate in the study and completed the first-round survey from which we obtained 34 valid responses. Our panelists had 14 roles (see Table 1).

Table 1. Delphi Panel Participants

Role	Frequency	Role	Frequency
Business analytics and big data managers	8	Data-protection officers	1
Chief data and analytics officers	2	Technology ethics academics	7
Senior data scientists	1	Data regulators	1
Business analytics academics	6	Digital law practitioner	1
Information management professionals	1	Privacy law practitioner	1
Privacy commissioners	1	Data ethics consultant	2
Ethics committee member of a professional body (e.g., ACM)	1	IT/ethics editors	1

In selecting participants for our Delphi panel, we sought to ensure that we had appropriate breadth to yield meaningful input on ethical issues from all three (i.e., individuals, organizations, and society) stakeholder perspectives. We recognize that, while each panel member would be able to comment on ethical issues from all three perspectives, they would do so to varying degrees. To ensure that our resulting panel had sufficient coverage of the three perspectives, we employed a fuzzy-sets approach (Ragin, 2008b) to classify the extent to which they could comment on ethical issues from an individual, organizational, and societal perspective. As we describe in Appendix C, our fuzzy-set analysis provides evidence for a well-balanced panel with sufficient breadth of coverage across all three perspectives.

3.4 Delphi Study Procedure

Our Delphi study had three consecutive data-collection and concept-refinement rounds following Keil et al.'s (2013), Schmidt, Lyytinen, Keil, and Cule's (2001) and Schmiedel et al.'s (2013) recommendations. These rounds involved: 1) brainstorming, 2) concept refinement, and 3) validation. In each round, we collected data from our expert panel over a three-week period using Qualtrics and analyzed the data using Gioia et al.'s (2012) methodology for developing concepts from qualitative data. The total number of rounds emerged when the panelists reached sufficient consensus (Worrell et al., 2013). We looked at consensus from two perspectives: 1) the qualitative feedback they provided about the concepts' names and definitions and 2) the satisfaction ratings they gave to the concepts. A convergence in the comments from our expert panelists about the wording that the concept definitions adopted indicated consensus on their names and definitions. The participants measured satisfaction on a seven-point Likert scale (highly dissatisfied to highly satisfied). We determined consensus by a mean satisfaction score of at least 5.0, which follows recent Delphi studies that involve concept definitions (Schmiedel et al. 2013). We also ranked concepts in each stakeholder group in the third round by using the average score for relative importance (out of 100 in total) allocated to the concept[1].

3.4.1 Round 1: Brainstorming

The first Delphi round comprised a brainstorming exercise with our expert panel to elicit as many ethical implications associated with big data analytics as possible. We asked each panelist to submit at least five ethical issues for individuals, organizations, and society and to briefly describe each issue. We asked three separate questions on separate pages including:

- Please identify and define current issues and future challenges of using big data analytics for individual customers whose data organizations collect, analyze, or sell.
- Please identify and define current issues and future challenges of big data analytics for organizations using big data analytics to offer services and products to individuals.
- Please identify and define current issues and future challenges of using big data analytics for society.

We analyzed the data using the Gioia method's three stages. All four members of our research team individually coded at least one half of the data by retaining the panelists' terms and wording whenever possible and minimizing editing to preserve the intended meaning. Then, all four authors worked together on consolidating issues, removing duplicates, and refining definitions. We intensively discussed differences in the individual codifications until we reached consensus. At this stage, we developed seven concepts related to individuals (Data OWNERSHIP, awareness, data control, trust, privacy, self-determination, fear), 11 organizational concepts (data quality, data sourcing, data sharing, decision making, presentation, ethical capability, ethical culture, ethical data governance, ethical performance, reputation and competitive pressure), and seven societal concepts (power, dependence, social awareness, surveillance, principles and guidelines, authority and climate).

3.4.2 Round 2: Concept Refinement

In this round, we refined the theoretical concepts that we identified in the first round. Following recommendations from Keil et al. (2013), Schmidt et al. (2001), and Schmiedel et al. (2013), we provided all participants with a list of identified ethical concepts along with their definitions and asked them to 1) add, remove or change any item, 2) rate their satisfaction with the identified concepts, and 3) suggest improvements to the definitions (Schmiedel et al., 2013). Our approach to refining concepts again concurred with the Gioia method in maintaining both our own and our informants' voices. In this round, the average satisfaction rates for all concepts exceeded four in a scale one to five except for data ownership, self-determination, fear, decision making, presentation, ethical capability, reputation, power, dependence and climate (see Appendix D). Based on the satisfaction rates and qualitative feedback, we consolidated, renamed, and further refined the concept definitions. Thus, for individual issues, we removed data

[1] Unlike ranking-type Delphi studies (Paré et al. 2013), we focused on concept development. We collected data on the importance of the concepts only in the third round.

ownership and renamed self-determination to choice and fear to anxiety. For organizational issues, we removed presentation and ethical capability and renamed ethical performance to behavior. For societal concepts, we renamed dependence to coercion and climate to social mindset. Based on qualitative feedback, we also made several changes to how we defined all concepts.

3.4.3 Round 3: Concept Validation and Ranking

In the third round, we validated the concepts for each of the stakeholder groups and ensured that we had reached a sufficient level of consensus. Similar to the second round, we again asked the panelists to indicate their overall satisfaction with the codification and to suggest any improvements to items or their definitions. To rank the issues, we asked the experts to allocate a total of 100 points to the issues according to their level of importance for each stakeholder group. By doing so, we focused on engaging the panelists to critically reflect on the issues and, thereby, avoid complacency in responses (Schmiedel et al., 2013). We used the average importance score for each concept to inform the ranking of concepts for each stakeholder. In this round, we removed anxiety from individual issues and merged the privacy and data control concepts. For organizational issues, we merged data sourcing and data sharing into data trading, merged ethical culture and behavior became into ethical governance, merged competitive pressure into reputation, and renamed decision making to algorithmic decision making. After the third round, we had established consensus (minimum average satisfaction rate of 5 out of 7 for all the concepts (see Appendix D) and convergence of their comments about the wording of concept definitions based on the final responses from 23 expert panel members (10 academics and 13 practitioners). As such, we had a 28 percent dropout rate (common for Delphi studies) (Keil et al., 2013).

4 Findings

We structure our findings in terms of the three interrelated stakeholder groups: individuals, organizations, and society. Figure 1 summarizes the concepts by stakeholder groups and relative importance. For each stakeholder, we first provide a table that delineates and defines the highest ranked theoretical concepts we developed from the data following the steps that Gioia et al. (2012) outline. We do not include concepts that received less than 10 percent average ranking score from participants. We then elaborate on each concept in detail and explain how they give rise to ethical issues in big data analytics.

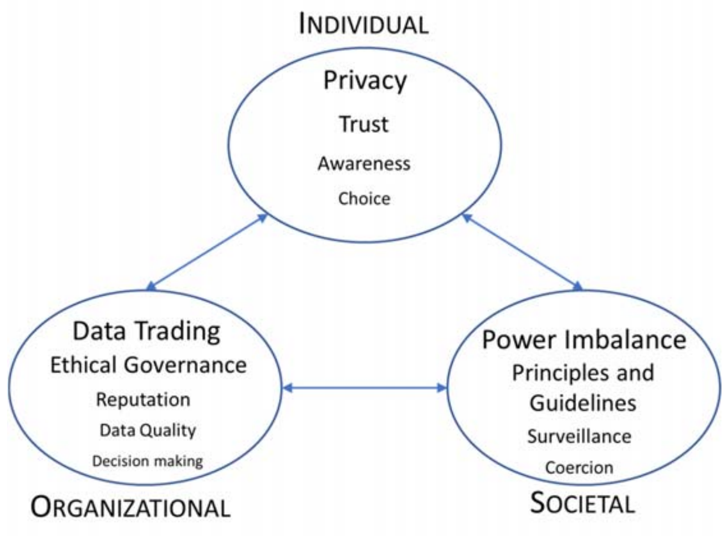

Figure 1. Concepts and Their Relative Importance to Stakeholders[2]

[2] Font size reflects the relative importance of the concepts as our panelists ranked them.

4.1 Implications of Big Data Analytics for Individuals

Table 2 presents the highest-ranked concepts that underlie ethical issues in big data analytics from individuals' perspective. We present the concepts in order of their relative importance as the participants ranked them in the third round of the Delphi Study.

Table 2. Individual Concepts

Rank (relative importance score)	**Concept**	**Definition**
1 (41.87)	Privacy	The extent to which an individual can restrict and control how organizations use and disclose their personal information.
2 (19.91)	Trust	The extent to which an individual can have confidence that the parties who have access to their data respect the individual's rights.
3 (15.36)	Awareness	The extent to which an individual mindfully consents to what data organizations collect about them and how they use it.
4 (14.95)	Choice	The extent to which an individual can freely make choices without being unfairly discriminated against or constrained by the use of big data analytics.

4.1.1 Privacy

Privacy refers to individuals' ability to restrict and to control how organizations use and disclose their personal information (Bélanger & Crossler, 2011). Three aspects of privacy that underlie ethical issues in this context include controlling how organizations can access, modify, and use personal data. First, even when individuals give consent for organizations to collect and share their data, individuals need to be able to control what data organizations collect and aggregate about them and who will have access to their aggregated data (Barocas & Nissenbaum, 2014; Tene & Polonetsky, 2013). While the separate databases that contain an individual's data might be anonymous, the aggregation process might re-identify the individual and make the data available to other parties without the individual's knowledge (Barocas & Nissenbaum, 2014). Second, individuals need to be able to modify data about themselves, which includes updating or deleting data to remedy incorrect, incomplete, or out-of-date data (Halavais, 2015). In particular, individuals require the ability to modify aggregated or shared data that might misrepresent them. Third, individuals need to influence how organizations use data about themselves. Although individuals may consent for the primary organization to use their data, they need to audit how and for what purposes other parties who have access to it will further exploit it (Tene & Polonetsky, 2013). Moreover, big data analytics leads to organizations' creating and sharing new knowledge about individuals. The new knowledge from aggregated data might reveal sensitive and unwanted information about individuals, create discomfort for them, and possibly have unintended consequences such as discrimination (Barocas & Nissenbaum, 2014; Wigan & Clarke, 2013).

4.1.2 Trust

Individuals need to be able to trust that organizations will fulfill their obligations, behave predictably, and not engage in inappropriate opportunistic behavior with their data. Aspects of trust that underlie ethical issues in big data analytics include unauthorized monitoring, unsolicited intrusions, and security of personal data. First, individuals need to have confidence that organizations collect their data only with their informed consent and that these organizations will use it only for clearly articulated purposes (Barocas & Nissenbaum, 2014). Individuals need to have confidence that they are not being observed and monitored in their everyday life. They may feel that an organization is exploiting them for data-extraction purposes and lose trust in it the organization if they believe it is manipulating their behavior (Martin, 2015; Richards & King, 2014). Second, individuals may receive unsolicited advertisements, emails, and promotional offers. It may be difficult to stop these intrusions since they may result from aggregated data that organizations have sold and widely distributed (Halavais, 2015; Zuboff, 2015). Individuals need to believe that organizations will not take advantage of big data analytics to unfairly profile them or use personal data in a manner that harms them (e.g., economically or socially) (Martin, 2015). Third, individuals need to have confidence that the organizations with which they engage will ensure their data's security, particularly in light of the degree of data sharing and cloud storage that commonly pervades the big data analytics context (Goes, 2014).

4.1.3 Awareness

Awareness concerns what individuals know and understand about big data analytics practices, such as how organizations analyze their data to offer products and services. Ethical issues arise when individuals lack awareness about why organizations use and the processes involved in big data analytics (Newell & Marabelli, 2015). Aspects of awareness that underlie ethical issues in big data analytics include understanding what big data analytics is, understanding rights regarding big data analytics, and understanding who holds the data and for what purpose. First, individuals need to learn about big data analytics, how it operates, and how it influences their choices and behavior (Crawford & Schultz, 2014). Individuals need to engage in public data literacy programs and recognize appropriate uses and consequences of big data analytics (Zuboff, 2015). By doing so, they can learn to better balance the personal costs and benefits of big data analytics (Newell & Marabelli, 2015). Second, individuals need to recognize policies, regulations, and laws that exist to protect them from the potential negative consequences of big data analytics (e.g., the European General Data Protection Regulation (GDPR)[3]). Subsequently, they can engage with governments and influence how regulations about big data analytics. Third, individuals need to know what data organizations collect about them, who owns and controls this data, and which third parties have access to it (Crawford & Schultz, 2014; Markus, 2015). Organizations often collect big data implicitly without clear informed consent, and they frequently hide secondary uses of the data from individuals (Barocas & Nissenbaum, 2014). Terms and conditions are obscure, and opting out from them can be difficult. Individuals need to recognize these practices since the analytics conducted on their data will ultimately influence their lives (Solove, 2013).

4.1.4 Choice

Big data analytics can restrict individuals' choices. As a result, it can discriminate against individuals and unfairly manipulate their behavior (Metcalf & Crawford, 2016; Zuboff, 2015; Zwitter, 2014). Aspects of choice that underlie ethical issues in big data analytics include the limiting of individuals' choices, incorrect analytics, and gamification. First, big data analytics may limit individuals' choices based on their past behavior, location, age, gender, and so on (Ananny, 2016; Newell & Marabelli, 2015). Organizations profile and categorize individuals according to their personal data and then send them services and products based on the resulting profile (Ananny, 2016). As a result, individuals may lose freedom of choice and face a less-than-free market (Richards & King, 2014). Second, the digital profiles that organizations create using aggregated data may not correctly represent individuals (Boyd & Crawford, 2012). This may arise due to low-quality aggregated data or inappropriate algorithms, which can yield an incorrect profile and predictions (Clarke, 2016; Wigan & Clarke, 2013). Further, organizations may still target and unfairly discriminate against individuals based on such data/algorithms (Clarke, 2016). Third, organizations may gamify individuals to further analyze and manipulate their behavior, such as by personalizing rewards until a certain customer shows certain behavior (Zuboff, 2015).

4.2 Implications of Big Data Analytics for Organizations

Table 3 presents (in rank order) the highest-ranked concepts that underlie ethical issues in big data analytics from organizations' perspective.

Table 3. Organizational Concepts

Rank (relative importance score)	**Concept**	**Definition**
1 (22.73)	Data trading	The extent to which organizations collect, buy, aggregate, share, and sell data from multiple sources in a manner that respects individuals' rights.
2 (21.09)	Ethical governance	The extent to which organizations have values, norms, and shared beliefs (informal governance) together with standards, decision rights, and responsibilities (formal governance) that promote ethical big data analytics practices.
3 (20.46)	Reputation	The extent to which relevant stakeholders, particularly customers, believe an organization will manage and use data about them ethically.
4 (14.59)	Data quality	The extent to which organizations ensure the quality of big data in a manner that respects individuals' rights.
5 (12.27)	Algorithmic decision making	The extent to which big data analytics and resulting organizational decisions respect of individuals' rights.

[3] More information on European General Data Protection Regulation can be accessed at http://www.eugdpr.org/.

4.2.1 Data Trading

Data has become a traded asset in the big data marketplace. Organizations source data from individuals either by directly interacting with them using their digital assets or by buying and aggregating data about them from multiple sources (Martin, 2015). They may subsequently share or sell this data to other organizations for monetary gain (Wixom & Ross, 2017). The mechanisms and processes by which organizations source, share, and trade data can raise ethical challenges. Three key aspects of data trading can create ethical challenges: explicit informed consent, transparency of data sourcing and sharing, and preserving anonymity and protecting data against unethical use along the big data value chain. First, data-sourcing practices harvest data from individuals often without their explicit informed consent or genuine voluntary participation (Abbasi, Sarker, & Chiang, 2016; Barocas & Nissenbaum, 2014). Organizations instrumentize their products and services with data-extraction features that track users' behavior (Davenport & Kudyba, 2016). These products and services quantify many aspects of individuals' lives without clearly communicating to those individuals what data they collect and for the purpose they collect it (Constantiou, Kallinikos, & Kallinikos, 2015). Terms and conditions are often vague and do not specify what exactly will happen to data after an organization collects it (Abbasi et al., 2016; Wigan & Clarke, 2013). Second, data-sharing practices can be opaque (Barocas & Nissenbaum, 2014; Martin, 2015). At best, many organizations may only inform individuals that they will share their data with third parties; however, they often provide no transparency about what data they will share, with which organizations they will share it, and how they or other partner organizations will use it. Third, as organizations share and aggregate data, protecting an individual's identity becomes more and more challenging (Zuboff, 2015). Data sharing can re-identify individuals who otherwise had anonymity at the data-collection point (Barocas & Nissenbaum, 2014). Moreover, the focal organization that collected the data might have limited control and influence over the access, quality, aggregation, and use of the data after organizations in the value chain share it multiple times (Martin, 2015). Clearly, while organizations have an opportunity to obtain benefits from sharing data, ethical challenges have significant potential to arise.

4.2.2 Ethical Governance

While formal governance concerns formal policies, standards, and accountabilities about data, informal governance concerns culture and is determined by what organizational actors believe and do based on their values, norms, and shared beliefs (Wixom & Markus 2017). Aspects of ethical governance that can create ethical issues include building ethical norms, establishing rules and procedures, and internalizing the costs. Unethical practices can become accepted and legitimized in an organization's culture (even despite formal policies) (Wixom & Markus, 2017). To prevent this problem, organizations need to be vigilant in data governance and governments must impose sanctions against unethical practices. Education and training is a means to build an appropriate set of shared norms, values, and beliefs and, ultimately, influence the actions of organizational actors with regard to data practices (Wixom & Markus, 2017). Organizations need to establish new rules and procedures to regulate and reinforce appropriate employee behavior. These rules can make data flows more transparent for customers and other organizations (Metcalf & Crawford, 2016). Where organizations share data, inappropriate use of the data by one organization in the chain can have negative effects for all the other organizations in the chain (Martin, 2015). These rules can also cover policies for data security where data is subject to conflicting laws and regulations in different locations. Organizations should balance the costs and benefits of big data analytics between individuals and organizations (Günther, Rezazade Mehrizi, Huysman, & Feldberg, 2017; Martin, 2015). Big data analytics can generate huge financial and market benefits for organizations, which might mean they extensively monitor and measure individuals' behaviors and manipulate their choices and behaviors. Currently, organizations mostly focus on generating value for themselves rather than internalizing some of the costs associated with big data analytics (Martin, 2015).

4.2.3 Reputation

Reputation concerns the extent to which relevant stakeholders, particularly customers, believe an organization will manage and use data about them ethically. An organization's reputation is determined by stakeholder perceptions of how the organization incorporates ethics into its big data analytics practices. Our concept of reputation here aligns with Scott and Walsham's (2005, p. 311) definition of reputation risk: "the potential that actions or events negatively associate an organization with consequences that affect aspects of what humans' value". Indeed, researchers have referred to reputation risk as the "risk of risks", and, to manage such risk, one needs to consider both the outcome and the underlying causes (Ross &

Lofthouse, 2005). Key aspects of reputation that underlie ethical issues in big data analytics include difficulty in developing an ethical culture, rogue employees' exploiting trust, and competitive pressure. First, organizations with a poor reputation for ethical big data analytics practices may struggle to develop an ethical culture internally. Second, rogue employees in organizations with good ethical reputations may be able to exploit individuals' trust, particularly as analytics capabilities evolve and new technologies and data sources become available. Finally, an organization's reputational concerns can counterbalance competitive pressures that may otherwise lead a firm to engage in unethical data practices in order to monetize data and outperform their competitors (Martin, 2015). Alternatively, if all competitors in a market have poor reputation, it may encourage organizations to remain unethical in their practices.

4.2.4 Data Quality

Big data quality underlies the correctness of decisions that organizations make using big data analytics. Although data quality is well defined for conventional data systems (Price & Shanks, 2005), it does not apply to the context of ethical big data analytics (Clarke, 2016). Specifically, aspects of data quality that can cause ethical issues in big data analytics include data quality criteria for big data, quality of aggregated data, and creation and maintenance of metadata. First, big data often comprises complex social data that organizations source in multiple formats and usually has an unstructured form (Clarke, 2016; Wigan & Clarke, 2013). Although structured data has quality dimensions including accuracy, timeliness, and completeness, it does not apply in the context of big data, so that adequate managerial guidelines are not available (Clarke, 2016). Second, organizations combine and aggregate data about individuals from multiple sources. Aggregated data might reveal information about individuals who otherwise had anonymity at the initial data-collection point (called the mosaic effect) (Barocas & Nissenbaum, 2014). We know little as to whether or not the aggregated data about individuals accurately represents them (Crawford et al., 2014). Third, when organizations source big data, they may not establish data definitions and not capture and maintain metadata information. Individuals might have contributed their data in different contexts for different reasons, particularly in the case of social media (Boyd & Crawford, 2012). The data definition problem is exacerbated by the fact that the data's meaning might change over time through sharing and aggregation processes, which can create even greater potential for unethical data use (Martin, 2015).

4.2.5 Algorithmic Decision Making

Decision making concerns the processes and outcomes of decisions that organizations make using big data analytics. Decisions made using big data analytics typically rely on complex statistical and computational methods. Aspects of decision making that underlie ethical issues in big data analytics include reliability of algorithms, lack of human involvement, and accountability of decisions. First, as the data increases in size, speed, and complexity, algorithms become more important in making sense of data, generating insights, and predicting the future (Newell & Marabelli, 2015). Organizations use algorithms to predict the future based on historical and subjective data, and, in most cases, they predict based on correlations only (as opposed to establishing a causal effect) (Ananny, 2016; Boyd & Crawford, 2012; Halavais, 2015). Organizations have no means to ensure that they have made an ethically appropriate decision about an individual. For example, an algorithm may inadvertently lead to racial profiling or some other act of discrimination (Ananny, 2016) because no theory exists to explain the relationships in the data since big data analytics relies mainly on inducing insights (Günther et al., 2017; Halavais, 2015). Second, decisions made using big data analytics are either automated, have no human involvement, or are visualized for the human decision maker. Visualizations typically convey a particular message or story and hide underlying assumptions, limitations, biases, and data-quality issues (Ekbia et al., 2015). The human decision maker has no means to understand how, or against what criteria, the decision has been made, which can limit the decision maker's ability to properly interpret the results. Third, when individuals make decisions using complex algorithms that humans find difficult or impossible to understand, the responsibility for decision outcomes becomes blurred (Ekbia et al., 2015). Such a situation could be problematic if, for example, poor-quality data or an unsuitable algorithm led to an unethical and discriminatory decision.

4.3 Implications of Big Data Analytics for Society

Table 4 presents (in rank order) the highest-ranked concepts that underlie ethical issues in big data analytics from society's perspective.

Table 4. Societal Concepts

Rank (relative importance score)	**Concept**	**Definition**
1 (35.14)	Power Imbalance	The extent to which a dominant group, organization, or government uses big data analytics in a way that imbalances power in society.
2 (28.96)	Principles and guidelines	The extent to which effective principles and guidelines exist and governments enforce them through policies, regulations, and laws to protect the rights of individuals impacted by big data analytics.
3 (16.41)	Surveillance	The extent to which organizations observe, monitor, measure, and profile individuals' lives in a society.
4 (10.64)	Coercion	The extent to which participation and functioning in society depends on contributing one's own data to a collection for analysis.

4.3.1 Power Imbalance

Power imbalance concerns the power, control, and influence relationships that arise from using big data analytics in society. This imbalance mainly arises because only a few entities dominate access to big data. From an ethical perspective, power imbalance undermines individuals' equality and their rights to freedom of choice (Boyd & Crawford, 2012; Crawford et al., 2014). Two aspects of power imbalance underlie ethical issues in big data analytics: limited access to big data analytics and knowledge asymmetries. First, groups, organizations, government agencies, and countries that have access to big data analytics; monitor, quantify, and aggregate data about many aspects of individuals' lives; and create detailed profiles about them. These entities get to know individuals better than they know themselves (Zuboff, 2015). Second, by contrast,, individuals using big data analytics services know little about organizational data practices, which results in knowledge asymmetries (Solove, 2013; Tene & Polonetsky, 2013) where a small number of organizations with a relatively small number of employees gain power and control over the rest of the population. The minority group can use this power to influence and modify individuals' behavior to generate economic or political value (Solove, 2013). The majority of the population has no choice or negotiation power, particularly when they remain oblivious about what happens with their data.

4.3.2 Principles and Guidelines

Principles and guidelines that regulatory bodies enforce through policies, regulations, and laws offer a means to protect individuals from the harms that big data analytics can cause. Currently, principles, guidelines, policies, regulations, and laws for protecting individuals from the consequences of big data analytics lag behind technological developments (Metcalf & Crawford, 2016; Zuboff, 2015). This lag creates an ethical challenge since actors in a society may be able to use big data analytics in legal or unregulated—but potentially unethical—ways (Metcalf & Crawford, 2016). Aspects of principles and guidelines that underlie ethical issues in big data analytics include the establishment of regulatory authorities that espouse principles and guidelines that balance the costs and benefits among stakeholders. First, governments, universities, big data associations, non-government organizations, and other not-for-profit organizations need to oversee how organizations use big data analytics and develop principles and guidelines to help them do so appropriately (Markus & Topi, 2015). Second, regulatory authorities need to be introduced to supervise, audit, or control organizations' practices and impose boundaries to ensure positive outcomes for the general public. Authorities can help to prevent organizations from musing data by investigating their data practices and penalizing deviation from norms (Metcalf & Crawford, 2016). Third, principles and guidelines, together with enforcing the laws, should help to better distribute the benefits of big data analytics between various stakeholders in society (Markus & Topi, 2015). Such distribution can empower individuals in society to practice their rights to privacy and self-determination and help them to develop public confidence in big data analytics services and, thus, shape big data analytics as a positive social phenomenon.

4.3.3 Surveillance

Surveillance refers to the extent to which organizations observe, monitor, measure, and profile individuals' lives. Aspects of surveillance that underlie ethical issues in big data analytics include loss of privacy and regulated behavior. First, big data analytics converts the everydayness of individuals into monetary value for organizations (Constantiou et al., 2015). The advent of the cloud, social technologies, and the Internet has dramatically increased this potential since it has created the technological foundation for a surveillance society (Lyon, 2014). The surveillance society functions by monitoring and collecting data about individuals' behavior, which limits their right to privacy and their freedom of choice (Crawford et al., 2014). Second, a surveillance society regulates rather than frees individuals' behavior. In contrast to unpredictable traditional markets, in a surveillance economy, organizations that have access to big data analytics know virtually everything (Lyon, 2014; Zuboff, 2015). These organizations analyze individuals' past behavior and predict or regulate what their future actions might or should be; thus, they manipulate individuals' behavior for their own benefit (Lyon, 2014; Richards & King, 2014; Zwitter, 2014).

4.3.4 Coercion

Coercion refers to the extent to which individuals' participation and functioning in society depends on contributing data to big data analytics services. Coercion arises mainly because organizations increasingly encourage or force individuals to use apps, social networks, and sensory devices to participate in many social and political activities (Newell & Marabelli, 2015). That individuals' contribute their data represents the sole condition for participating. Two aspects of coercion underlie ethical issues in big data analytics: individuals' lack of free will and their dependence on big data analytics services. First, individuals often do not have a free and unencumbered choice when contributing their data (Lyon, 2014). If they choose not to contribute their data, organizations will sanction or exclude them from participating in many personal, social, and political activities (Zuboff, 2015). Second, organizations coerce individuals into depending on big data analytics services without any knowledge of their real purpose (e.g., they become dependent on using a service without knowing that it takes and sells their data) (Galliers et al., 2017). Data-driven services have become ubiquitous, and, increasingly, individuals find them difficult to avoid.

5 Discussion

We now analyze the concepts that underlie ethical issues in big data analytics using stakeholder theory and discourse ethics to illuminate how stakeholder interactions might ensure that organizations more ethically use big data analytics.

5.1 Stakeholder Salience in Big Data Analytics

Three stakeholder attributes relevant to salience include their power to influence big data analytics, the legitimacy of their relationship to big data analytics, and the urgency of their claims on big data analytics (Mitchell et al., 1997) (see Figure 2). Power refers to the extent to which a stakeholder can impose its will in a relationship; legitimacy to the extent to which a stakeholder's actions are desirable, proper, and appropriate in a social system; and urgency to the extent to which stakeholder claims call for immediate action.

In big data analytics context, organizations have high power and urgency but varying degrees of legitimacy. Organizations have high power as they have the technology, data, and expertise necessary to engage in big data analytics activities that impact individuals and society. They have high urgency because they complete much of their data collection, algorithmic decision making, and subsequent actions in short time frames in part due to competitive pressures (e.g., high-frequency trading on stock markets) (Currie & Seddon, 2017). Organizations have varying degrees of legitimacy depending on the extent to which their actions oppose the individuals' and societies' values. Thus, organizations have high salience and urgency in the big data analytics context, and one may consider them dangerous stakeholders when their actions have low legitimacy (Mitchell et al., 1997).

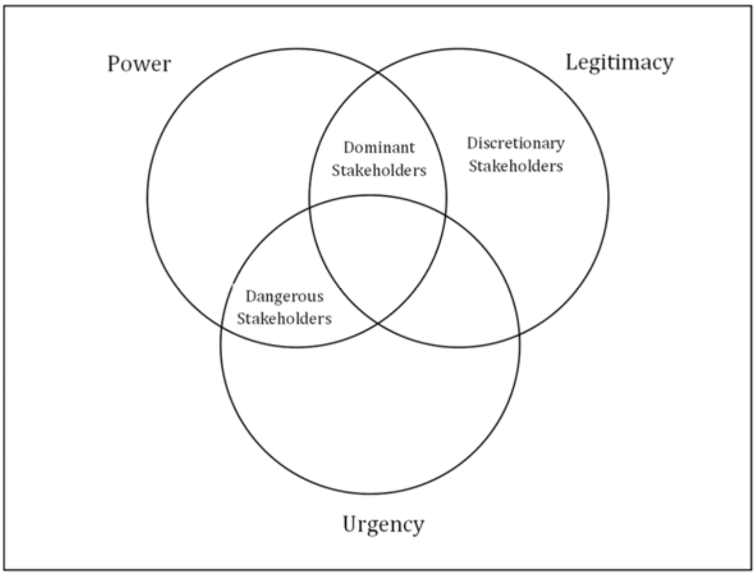

Figure 2. Stakeholder Salience in Big Data Analytics Context (Adapted from Mitchell et al. 1997)

Individuals engaged as stakeholders in big data analytics have low power and urgency but high legitimacy. Individuals have low power because they can rarely impose their will on other stakeholders involved in big data analytics and often do not know how organizations collect and use their personal data. They have low urgency because they participate in big data analytics rather passively and have relatively less need for immediate action. Individuals have high legitimacy because their actions are mostly desirable, proper, and appropriate in society. Inappropriate actions can lead to sanctions and legal consequences. Hence, individuals have low salience in the big data analytics context, and one may consider them discretionary stakeholders whom other stakeholders frequently ignore (Mitchell et al., 1997).

Society as a stakeholder in big data analytics has varying degrees of power, low urgency, and high legitimacy. Generally, societies impose their will through laws, regulations, guidelines, and sanctions. However, they cannot easily develop and implement such things in a context that features rapidly evolving technology and the need for consensus. Privacy laws, such as in particular the European Union's GDPR, exemplify societies' power. Societies have low urgency because developing and implementing policies, guidelines, and laws concerning big data analytics takes a long time, while big data analytics technology and its use by organizations evolves rapidly. Societies have high legitimacy because their actions are generally desirable, proper, and appropriate for their citizens. Therefore, society has low salience in the big data analytics context, and one may consider it as a discretionary stakeholder whom other stakeholders frequently ignore (Mitchell et al., 1997).

5.1.1 Dyadic Interactions of Stakeholders in Big Data Analytics: Towards an Ethical Discourse

We argue that, in big data analytics context, organizations have high salience and individuals and society have low salience, which implies that, currently, organizations that use big data analytics dominate interactions with individuals and society. They frequently take individuals' data as a free or cheap resource and use it for their own benefit (Zuboff, 2015). Societies implement few laws, regulations, guidelines, and sanctions that constrain and guide how organizations use big data analytics specifically (Metcalf & Crawford, 2016). Although big data analytics capabilities continue to evolve rapidly, future acceptance and use of big data requires ethical and transparent practices. Organizations that use big data will need to address these issues efficiently and effectively to ensure consumer and regulatory acceptance. Indeed, failure to do so could drive a social consensus that empowers society to impose more stringent regulations and sanctions.

To be effective, discourse ethics ideally requires stakeholders to have equal or similar salience. We require ethical principles that meet all stakeholders' needs. We now discuss stakeholders' dyadic interactions and focus on how they relate to the concepts that underlie ethical issues derived from the

Delphi study. We also identify ways to increase the salience of individual and society stakeholders and, ultimately, enable ethical discourse.

5.2 Interactions between Individuals and Organizations

Individuals represent key stakeholders in big data analytics in that they serve as both data sources for organizations and targets of decisions made using big data analytics, which is problematic because individuals often lack awareness of what data organizations collect about them and what happens to their data afterwards (Richards & King, 2014). Many big data trading practices lack transparency (Barocas & Nissenbaum, 2014). In many cases, organizations acquire data from individuals with only implicit consent, combine it from multiple sources, analyze it, use it, and/or sell it to third parties (Martin, 2015). As one of our panelists asked, “Do [individuals] know who holds the data [about them] and for what purpose?”. On the other hand, individuals who learn about big data analytics practices can rarely restrict and control the information that organizations create about them (e.g., individuals cannot update or delete data about themselves) (Clarke, 2016; Wigan & Clarke, 2013). In this case, our panelists commented: “How can I control who can/can't access my data?” and “Once the data goes into the aggregated big data set, how can a customer correct 'their' data if it is wrong?”. Unsurprisingly, this lack of control can lead to a lack of trust in organizations using big data. Panelists commented: “Problems arise from (the feeling of) constantly being tracked and those actions having consequences on many parts of one’s life” and “fear of my information being accessed by entities unknown”. Organizations need to recognize this lack of trust and develop reputations for ethical practices in big data analytics (Boyd & Crawford, 2012). Thus, in substance, organizations need to develop appropriate ethical governance informally and formally with appropriate values, norms, and shared beliefs (informal) together with policies and procedures (formal).

Organizations use analytics to create new knowledge about individuals, and exposing this new knowledge intentionally or unintentionally to others can harm individuals’ right to privacy (Barocas & Nissenbaum, 2014; Halavais, 2015). As one panelist noted: “Big data can lead to the discovery of information about the data subject that she herself is not aware of”. Organizations can use this knowledge to customize offers and manipulate individuals’ behavior for their own benefit (Zuboff, 2015). Analytics-based decision making in organizations creates “risks [for individuals] of being economically exploited or risks of being discriminated against (e.g., health insurance costs)” as one panelist cautioned. Individuals may suffer constraints on their freedom of choice and face discrimination from algorithms that may, perhaps unintentionally, profile them based on their race, income, gender, or social class (Ananny, 2016; Boyd & Crawford, 2012).

According to stakeholder theory, to increase their salience in interactions with organizations, individuals need to increase their power and urgency. They can do so in several ways. Individuals need to become more knowledgeable about big data analytics. They must understand the practices and consequences of big data analytics and then interact with organizations to ensure that the organizations establish ethical governance practices and that they have more access to and control over their personal data (Richards & King, 2014). Only then can individuals and organizations work together to ensure the quality of personal data and transparent data-trading practices (Martin, 2015). Furthermore, organizations need to develop decision-making practices that respect individuals’ freedom of choice and further build trust and reputation. Increasing the salience of individuals in their interactions with organizations can lead to mutual benefits and, by enabling ethical discourse, more ethical big data analytics practices to emerge.

5.3 Interactions between Organizations and Society

Big data analytics has far-reaching consequences for society via its introducing new forces and dynamics that influence equality and power relationships. These consequences can arise when a few large organizations dominate data trading in big data analytics, which creates knowledge asymmetries that can lead to a power imbalance and dominant entities’ surveilling societies (Boyd & Crawford, 2012; Crawford et al., 2014). Big data analytics challenges the basis for free markets in society by controlling and regulating behavior to create profits for organizations (Zuboff, 2015).

While organizations push boundaries by monetizing data about individuals to gain competitive advantage, current principles and guidelines that protect society lag behind technological developments (Metcalf & Crawford, 2016; Richards & King, 2014). According to one panelist, societies need to “design new laws and regulations for organizations that analyze big data”. Organizations that face competitive pressure to use big data analytics unethically may lack their own ethical governance practices, so society must not only promulgate principles and guidelines for ethical use but also exercise authority to enforce them and sanction violators.

According to stakeholder theory, to increase their salience in interactions with organizations, societies need to increase their power and urgency. Groups in societies must develop and implement effective principles and guidelines for organizations and, thereby, promote ethical governance practices. Societies should further develop these principles and guidelines into regulations and laws, such as national privacy regulations and laws, that they can use to sanction organizations that do not comply (Crawford & Schultz, 2014; Metcalf & Crawford, 2016). Two examples include the Australian Privacy Principles and the Australian privacy commissioner, who can investigate breaches of the privacy laws and apply sanctions, and the GDPR. Societal institutions need to interact with organizations to ensure that power imbalances do not become entrenched and that competitive pressures on organizations do not incentivize unethical practices (Martin, 2015). As a society, we must find ways for organizations to engage ethically in data trading without subjecting society to Orwellian-level surveillance. Increasing the salience of societies in their interactions with organizations can lead to mutual benefits and, by enabling ethical discourse, cause big data analytics practices that balance different stakeholders' interests more ethically to emerge (Markus & Topi, 2015).

5.4 Interactions between Individuals and Society

The practices and consequences of big data analytics highly affect both individuals and societies, yet they arguably have limited influence over shaping how big data analytics changes society and individuals' own lives. Our participants highlighted that "commercial needs predominantly" drive the benefits from big data analytics and that a clear need for "balancing of interests" in a context where a significant power imbalance may arise exists. Such balance requires individuals to better understand big data analytics practices and their consequences and to actively participate in societies to ensure they develop appropriate principles and guidelines. These principles and guidelines need to examine whether and how society at large can avoid unethical consequences, such as discrimination against individuals, (Wigan & Clarke, 2013). Societies need to ensure that they establish and enforce regulations and laws to sanction organizations that do not comply (Crawford & Schultz, 2014).

Participation in society has increasingly come to depend on using apps, social networks, and sensors almost to the point of coercion where providing one's data no longer represents a choice (Newell & Marabelli, 2015). Individuals have become more dependent on using big data analytics services without any awareness about the influence that big data analytics has on their lives and the consequences that result from such influence (Richards & King, 2014). In particular, societies need to understand the possibility and consequences of a surveillance economy (Zuboff, 2015). Panel members stated that societies must create "awareness of what might happen when big data gets analyzed" and "protect the citizens from abuse of big data analytics".

According to stakeholder theory, both individuals and societies need to increase their salience in interactions with organizations. Individuals should actively participate in developing principles and guidelines to ensure that societies establish regulations and laws with effective sanctions. Such sanctions will help to protect individuals' rights about data privacy and freedom of choice (Crawford & Schultz, 2014). Societies can provide education to help individuals better understand the benefits and costs of big data analytics (Zuboff, 2015). Societies need to find ways to share in the benefits of big data analytics without coercing individuals to provide data in order to participate in society. Increasing the salience of individuals and societies jointly in their interactions with organizations can lead to mutual benefits and, thus, enable an ethical discourse and a strong impetus for organizations to develop and adopt ethical big data analytics practices.

6 Conclusion

Big data analytics represents a complex social phenomenon with an inherent duality. It clearly offers opportunities to further advance human societies but also creates ethical challenges for the stakeholders involved. In this study, we use stakeholder theory to analyze the salience of each stakeholder involved in big data analytics and discourse ethics and stakeholder theory to discuss the dyadic interactions among the stakeholders.

6.1 Implications for Research

Two implications for research emerge from our study. First, we provide a stakeholder perspective on big data analytics and define it as a social process that arises from the interactions between multiple stakeholders. Such stakeholders include individuals who contribute their data, organizations that use big data, and societies that have the responsibility to govern, control, and shape this evolving sociotechnical phenomenon. Unlike the current, largely technical view of big data analytics, we use a stakeholder perspective to identify and conceptualize the ethical issues and challenges for each stakeholder. In particular, we focus on interactions between relevant stakeholders as a means to address ethical problems based on discourse ethics. By using stakeholder theory and discourse ethics, we provide researchers with a useful theoretical lens to further explore ethical issues in big data analytics. Second, we develop theoretical concepts that underlie the ethical issues associated with big data analytics for the individuals, organizations, and societies. In this regard, we address the research gap about the lack of theoretical concepts in the emerging big data analytics body of knowledge. Our Delphi study and concept development approach provides empirical evidence for the theoretical concept and establishes their relevancy to each stakeholder, particularly in enabling an ethical discourse.

6.2 Implications for Practice

Our findings provide stakeholders with the language and concepts necessary to confront ethical issues as they engage with and are impacted by big data analytics. They inform individuals about how big data analytics influences their lives and empower them to engage in balancing the positive and negative consequents of big data analytics. They inform organizations about the factors that engender ethical problems for individuals and society and also the factors to consider as they seek to use big data analytics in an ethical manner. Finally, they highlight how big data analytics influences society and how a society can control and shape it in a way that benefits all stakeholders in a balanced and fair manner.

6.3 Limitations and Future Work

We used a Delphi study to identify and rank theoretical concepts that explain the ethical issues of big data analytics for three stakeholder groups. In doing so, we focused on defining each concept and disaggregating the concepts to their various dimensions. We used stakeholder theory and discourse ethics to analyze the theoretical concepts and explain how stakeholder interactions might enhance the extent to which organizations ethically use big data analytics.

Our study has five limitations. First, we base our findings on the perceptions of a limited number of participants. The study included 34 academic and practitioner participants from around the world, which concurs with other Delphi studies in the IS literature (e.g., Keil et al., 2013; Schmiedel et al., 2013). The participants had diverse backgrounds and experience, which their roles evidence (see Appendix C). We demonstrate through fuzzy set analysis that we had sufficient coverage of the three stakeholder groups to provide broadly representative input into identifying and defining concepts, although, unsurprisingly, the organizational stakeholder group had the greatest coverage. Nevertheless, one can always improve the balance of perspectives in the panel. Future research could include participants with strong views about ethical issues for society in particular, such as social activists. Second, Delphi studies can face difficulty in achieving consensus (Schmiedel et al., 2013). Our consensus-finding process included convergence in both naming and defining the concepts and mean satisfaction rates that exceeded a threshold score for each concept over two rounds in the Delphi study. Although Paré et al. (2013) recommend at least six rounds for a full Delphi study, Keeney et al. (2006) argue that “response exhaustion” can occur after two rounds and certainly after four rounds. We included three rounds in our study as a compromise between these positions and also because we had achieved strong convergence in the concepts’ names and definitions and in satisfaction rates. However, we may have achieved better consensus with further rounds. Future research could include a greater number of rounds and a more demanding consensus process. Third, many of the concepts we defined already exist in previous literature. Although one may see this fact as a limitation, we argue that we grouped the concepts by stakeholder type and disaggregated the concepts to their various dimensions. Furthermore, we identified a rank for the concepts according to each stakeholder group to enable researchers and practitioners to prioritize their future work. Our main contribution lies in our defining and ranking the concepts in detail rather than simply identifying them, which creates the novelty in our findings. Future research could focus on particular stakeholder types or concepts and enhance our detailed definitions and rankings. Fourth, although we discuss dyadic stakeholder interactions and identify some interconnectivity between concepts, future research needs to expand our approach to analyze possible interactions between stakeholders. Finally, we focused the study on data about individuals that organizations collect. Research needs to examine the many other applications of big data analytics in other domains.

Despite these limitations, which many Delphi studies share (Worrell et al., 2013), our findings contribute to both research and practice. Additionally, our identifying and defining a set of concepts that underpin the ethical issues with big data analytics provides a sound base for future (both qualitative and quantitative) research. It provides qualitative researchers with a powerful lens through which to explore interactions between different stakeholders in various contexts, and it provides quantitative researchers with a set of concepts with strong face validity that they can further develop into measurable constructs. Future research can focus on designing artifacts that would facilitate ethical interactions between stakeholders in terms of the concepts we have developed.

In this paper, we provide a stakeholder perspective on big data analytics. We used a Delphi study to identify the key concepts that underlie ethical issues in big data analytics for three different stakeholders. We analyzed these concepts using stakeholder theory and discourse ethics to suggest how individuals, organizations, and society can better interact to ensure they more ethically use big data analytics.

References

Abbasi, A., Sarker, S., & Chiang, R. H. L. (2016). Big data research in information systems: Toward an inclusive research agenda. *Journal of the Association of Information Systems*, *17*(2), 1–32.

Ananny, M. (2016). Toward an ethics of algorithms: Convening, observation, probability, and timeliness. *Science Technology and Human Values*, *41*(1), 93-117.

Anderson, J., Rungtusanatham, M., & Schroeder, R. (1994). A theory of quality management underlying the deming management method. *The Academy of Management Review, 19*(3), 472-509.

Barocas, S., & Nissenbaum, H. (2014). Big data's end run around anonymity and consent. In L. Jane, V. Stodden, S. Bender, & H. Nissenbaum (Eds.), *Privacy, big data, and the public good* (pp. 44-75). New York, NY: Cambridge University Press.

Bélanger, F., & Crossler, R. E. (2011). Privacy in the digital age: A review of information privacy research in information systems. *MIS Quarterly*, *35*(4), 1017-1041.

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.

Chen, H., Chiang, R., & Storey, V. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, *36*(4), 1165-1188.

Clarke, R. (2016). Big data, big risks. *Information Systems Journal*, *26*(1), 77-90.

Constantiou, I. D., Kallinikos, J., & Kallinikos, J. (2015). New games, new rules: Big data and the changing context of strategy. *Journal of Information Technology*, *3017*(1), 44-57.

Crawford, K., Miltner, K., & Gray, M. L. (2014). Critiquing big data: Politics, ethics, epistemology. *International Journal of Communication*, *8*, 1663-1672.

Crawford, K., & Schultz, J. (2014). Big data and due process—toward a framework to redress predictive privacy harms. *Boston College Law Review*, *55*(1), 93-128.

Currie, W. L., & Seddon, J. J. M. (2017). The regulatory, technology and market "dark arts trilogy" of high frequency trading: A research agenda. *Journal of Information Technology*, *32*(2), 111-126.

Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science*, *9*, 458- 467.

Davenport, T. H., & Kudyba, S. (2016). Designing and developing analytics-based data products. *MIT Sloan Management Review*, *58*(1), 83-89.

Derikx, S., de Reuver, M., & Kroesen, M. (2016). Can privacy concerns for insurance of connected cars be compensated? *Electronic Markets*, *26*(1), 73-81.

Ekbia, H., Mattioli, M., Kouper, I., Arave, G., Ghazinejad, A., Bowman, T., Suri, V. R., Tsou, A., Weingart, S., & Sugimoto, C. R. (2015). Big data, bigger dilemmas: A critical review. *Journal of the Association for Information Science and Technology*, *66*(8), 1523-1545.

Freeman, R. E. (1984). Strategic management: A stakeholder theory. *Journal of Management Studies*, *39*(1), 1-21.

Galliers, R., Newell, S., Shanks, G., & Topi, H. (2017). Datification and its human, organizational and societal effects: The strategic opportunities and challenges of algorithmic decision-making. The Journal of Strategic Information Systems, *26*(3), 185-190.

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2012). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational Research Methods*, *16*(1), 15-31.

Goes, P. B. (2014). Big data and IS research. *MIS Quarterly*, *38*(3), iii-viii.

Günther, W. A., Rezazade Mehrizi, M. H., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems*, *26*(3), 191-209.

Habermas, J. (1990). Discourse ethics: Notes on a program of philosophical justification. In J. Habermas (Ed.), *Moral consciousness and communicative action*. Cambridge, MA: Polity Press.

Habermas, J. (1992). Discourse ethics, law and Sittlichkeit. In P. Dews (Ed.), *Autonomy and solidarity: Interviews with Jürgen Habermas*. London, UK: Verso.

Halavais, A. (2015). Bigger sociological imaginations: Framing big social data theory and methods. *Information Communication and Society*, *18*(5), 583-594.

Keeney, S., Hasson, F., & McKenna, H. (2006). Consulting the oracle: Ten lessons from using the Delphi technique in nursing research. *Journal of advanced nursing*, *53*(2), 205-212.

Keil, M., Lee, H. K., & Deng, T. (2013). Understanding the most critical skills for managing IT projects: A Delphi study of IT project managers. *Information and Management*, *50*(7), 398-414.

Lewis, P. V. (1985). Defining “business ethics”: Like nailing jello to a wall. *Journal of Business Ethics*, *4*(5), 377-383

Loebbecke, C., & Picot, A. (2015). Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda. *Journal of Strategic Information Systems*, *24*(3), 149-157.

Lyon, D. (2014). Surveillance, Snowden, and big data: Capacities, consequences, critique. *Big Data & Society*, *1*(2), 1-13.

Madsen, A. K. (2015). Between technical features and analytic capabilities: Charting a relational affordance space for digital social analytics. *Big Data & Society*, *2*(1).

Marjanovic, O., & Cecez-Kecmanovic, D. (2017). Exploring the tension between transparency and datafication effects of open government IS through the lens of complex adaptive systems. *Journal of Strategic Information Systems*, *26*(3), 210-232.

Markus, M. L. (2015). New games, new rules, new scoreboards: The potential consequences of big data. *Journal of Information Technology*, *30*(1), 58-59.

Markus, M. L. (2017). Datification, organizational strategy, and IS research: What’s the score? *Journal of Strategic Information Systems*, *26*(3), 233-241.

Markus, M. L., & Topi, H. (2015). *Big data, big decisions for science, society, and business.* Bentley University.

Martin, K. E. (2015). Ethical issues in the big data industry. *MIS Quarterly Executive*, *14*(2), 67-85.

McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review*, *90*(10), 61-68.

Metcalf, J., & Crawford, K. (2016). Where are human subjects in big data research? The emerging ethics divide. *Big Data & Society*, *3*(1).

Mingers, J., & Walsham, G. (2010). Towards ethical information systems: The contribution of discourse ethics. *MIS Quarterly*, *34*(4), 833-854.

Mitchell, R. K., Agle, B. R., & Wood, D. J. (1997). Toward a theory of stakeholder identification and salience: Defining the principle of who and what really counts. *Academy of Management Review*, *22*(4), 853-886.

Murdoch, T. B., & Detsky, A. S. (2013). The inevitable application of big data to health care. *JAMA*, *309*(13), 1351-1352.

Newell, S., & Marabelli, M. (2015). Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of “datification”. *Journal of Strategic Information Systems*, *24*(1), 3-14.

Paré, G., Cameron, A.-F., Poba-Nzaou, P., & Templier, M. (2013). A systematic assessment of rigor in information systems ranking-type Delphi studies. *Information & Management*, *50*(5), 207-217.

Price, R., & Shanks, G. (2005). A semiotic information quality framework: Development and comparative analysis. *Journal of Information Technology*, *20*(2), 88-102.

Ragin, C. C. (2008a). Qualitative comparative analysis using fuzzy sets (fsQCA). In B. Rihoux & C. C. Ragin (Eds.), *Configurational comparative methods: Qualitative comparative analysis* (QCA) and related techniques (vol. 51, pp. 87-121). London, UK: Sage.

Ragin, C. C. (2008b). *Redesigning social inquiry: Fuzzy sets and beyond*. Chicago, IL: University of Chicago Press.

Richards, N., & King, J. (2014). Big data ethics. *Wake Forest Law Review.*

Ross, A., & Lofthouse, G. (2005). *Reputation: Risk of risks*. London, UK]: The Economist Intelligence Unit.

Russom, P. (2011). *Big data analytics.* TWDI Research.

Schmidt, R., Lyytinen, K., Keil, M., & Cule, P. (2001). Identifying software project risks: An international Delphi study. *Journal of Management Information Systems*, *17*(4), 5-36.

Schmiedel, T., vom Brocke, J., & Recker, J. (2013). Which cultural values matter to business process management? Results from a global Delphi study. *Business Process Management Journal*, *19*(2), 292-317.

Scott, S. V, & Walsham, G. (2005). Reconceptualizing and managing reputation risk in the knowledge economy: Toward reputable action. *Organization Science*, *16*(3), 308-322.

Solove, D. J. (2013). Introduction: Privacy self-management and the consent dilemma. *Harvard Law Review*, *126*(7), 1880-1903.

Spiekermann, S., & Korunovska, J. (2017). Towards a value theory for personal data. *Journal of Information Technology*, *32*(1), 62-84.

Tene, O., & Polonetsky, J. (2013). Big data for all: Privacy and user control in the age of analytics. *Northwestern Journal of Technology and Intellectual Property Volume*, *11*(5), 239-271.

Wigan, M. R., & Clarke, R. (2013). Big data's big unintended consequences. *IEEE Computer*, *46*(6), 46-53.

Wixom, B. H., & Markus, L. (2017). To develop acceptable data use, build company norms. *Research Briefing of the Center for Information System Research*, *XVII*(4).

Wixom, B. H., & Ross, J. W. (2017). How to monetize your data? *MIT Sloan Management Review*, *58*(3), 10-13.

Worrell, J. L., Di Gangi, P. M., & Bush, A. A. (2013). Exploring the use of the Delphi method in accounting information systems research. *International Journal of Accounting Information Systems*, *14*(3), 193-208.

Yoo, Y. (2015). It is not about size: A further thought on big data. *Journal of Information Technology*, *30*(1), 63-65.

Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, *30*(1), 75-89.

Zwitter, A. (2014). Big data ethics. *Big Data & Society*, *1*(2).

Appendix A: Paré et al. (2013) Checklist for Rigor in Delphi Studies

1) Describe expert-recruitment and -selection process in detail

   We provide details in Section 3.1 of the process by which we recruited our expert panel. We first identified potential panel members using several criteria: a balance between academics and practitioners, a balance between information systems and other experts (for example legal), and the ability to adequately represent the views of individuals, organizations, and society.

2) Profile expert participants

   We discuss the profiles of the expert participants in Section 3.1 and include details of the profiles in Appendix C.

3) Initial request for participation, panel size, and retention rate

   We used email to send initial invitations to experts, which described the research project as the University of Melbourne research ethics process requires. Our sample size in the first round was 34 panel members, and we had a retention rate of 72 percent.

4) Pre-test Delphi instructions and data-collection instruments

   We pretested the Delphi instructions first with several academic colleagues and then with several external experts. We made minor changes to the instructions, although generally we found strong support for the clarity of the instructions and the data-collection instrument.

5) Experts describe and validate descriptions of their items in brainstorming phase

   We instructed experts to clearly define the concepts they identified in the first Delphi round. We did not give them a second opportunity in the first round to validate their descriptions, although we did so in the second round. We did not so in the first round due to concerns about "response exhaustion", which can occur even after two rounds (Keeney, Hasson, & McKenna, 2006).

6) Randomly order items in narrowing-down phase

   We categorized the items into the three stakeholder types (individual, organization, and society) and randomly ordered them in each category.

7) Justify modifications to full Delphi method

   The full Delphi method as Paré et al. (2013) describe it comprises at least six rounds to fully cover the recommended steps. However, using six steps contradicts Keeney et al.'s (2006) recommendations: they note the importance of "response exhaustion", which can occur after two rounds and certainly will occur after four rounds. We included three rounds in our study as a compromise between these positions and to have rounds for the three key phases: brainstorming, narrowing down, and ranking (Paré et al., 2013).

Appendix B: Data Structures

First-Order Categories	Second-Order Themes	Aggregate Dimensions
(1) Individuals restricting who can access their personal information, (2) individuals choosing what data they share with who, (3) Individuals choosing what data is aggregated about them	Ability to control the access to personal information	Privacy
(1) Individuals knowing and consenting to the use of personal information, (2) Individuals consenting to the secondary uses of personal information, (3) Individuals able to choose what the data is used for and by who	Ability to control the use of personal information	
(1) Individuals able to audit the quality of data available about them, (2) Individuals able to update and delete unreliable personal information, (3) Individuals able to correct aggregated data	Ability to modify personal information	
(1) Individuals able to genuinely consent for the collection of their personal data, (2) Individuals able to easily withdraw from participation in big data analytics services, (3) Individuals assured that their personal data will not be used for any purpose other than what she/he agreed to	Protection against unauthorized monitoring	Trust
(1) Individuals are only contacted by organizations they choose and for reasons they have agreed to, (2) Individuals are not subjected to unsolicited advertisements, emails and promotional offers based on data aggregated across services	Protection against unsolicited intrusions	
(1) Individuals having the confidence that organizations will protect their data against security breaches, (2) Individuals having the confidence that organizations will protect their data shared with other third parties or stored in cloud environments	Assurance of security of personal information	
(1) Individuals understanding how big data analytics services use their data, (2) Individuals cognizant of the consequences of loss of privacy and potential to be manipulated for marketing or political reasons, (3) Individuals understanding appropriate ways to engage with big data analytics services	Knowledge of big data analytics	Awareness
(1) Individuals understanding policies, regulations and laws that exist to protect them from the potential negative consequences of big data analytics	Knowledge of legal rights against big data analytics	
(1) Individuals knowing what data is collected about them, (2) individuals knowing who owns and controls their data, (3) individuals knowing who their data is shared with	Knowledge of what data is shared with who	
(1) Individuals targeted based on their past behaviour, and constrained in their exposure to new opportunities and experiences, (2) Individuals being profiled and categorized based on their age, gender, religion, location	Limiting individuals' choices	

Figure B1. Data Structures for Individuals' Ethical Issues

First-Order Categories	Second-Order Themes	Aggregate Dimensions
(1) Explicit, intelligent and easy to understand consent form, (2) Choice for individuals to genuinely participate or easily opt-out of data collection, (3) Clear communication of data collection purposes and practices	Explicit informed consent	Data Trading
(1) Transparency as to what data is sourced or collected about individuals, (2) Transparency as to what data is aggregated about individuals, (3) Transparency as to who the data is shared with and for what purposes	Transparency of data sourcing and sharing	
(1) Protecting individuals' identities in the big data value chain, (2) Reducing the risk of re-identification as the data is shared or sold with other third parties	Preserving anonymity	
(1) Shared governance of appropriate and acceptable data use practices in the big data value chain, (2) focal organization collecting the data exerting power over other parties in the value chain regarding the use and purposes of data, and further sharing of data.	Protecting data against unethical use in big data	
(1) Developing an ethical data culture (shared values, norms and beliefs) by educating and training employees, (2) Informing employees of the broader implications of big data analytics on individuals and society	Building ethical norms	Ethical Governance
(1) Establishing data ethics committees to monitor big data analytics practices, (2) Establishing rules and procedures regulating acceptable data practices within the organisation and in the big data value chain (3) dealing with conflicting laws and compliance with regulations in different locations	Establishing rules and procedures	
(1) Study and investigate the impact and consequences of big data analytics on different stakeholders particularly individuals and society, (2) Balance and equitably distribute the costs and benefits of big data analytics among the various stakeholders	Internalizing the costs	

Figure B2. Data Structures for Organizational Ethical Issues

First-Order Categories	Second-Order Themes	Aggregate Dimensions
(1) Lack of bottom-up culture development through exemplary practices within organizations, (2) inability to develop social norms in which employees naturally follow ethical data practices, (3) lack of power to enforce (top-down) use of ethical data practices	Difficulty in developing an ethical culture	Reputation
(1) Employees may take advantage of access to data for personal gain (2) inability to monitor and control the actions of all employees with access to data.	Exploitation of trust by rogue employees	
(1) Competitive business environment incentivizes organizations to reap the benefits of big data analytics without considering consequences for individuals or society	Competitive pressure	
(1) Creating guidelines for assessing big data quality, particularly unstructured and social data, (2) ensuring all employees understand what data items and sourcing practices are acceptable within the organization	Big data quality criteria	Data Quality
(1) Ensuring there are practices in place for auditing and assessing accuracy of aggregated data about individuals	Quality of aggregated data	
(1) Creation of data definitions when data is sourced or collected, (2) Capturing the context, including purpose, in which data is shared, (3) maintaining meaning of data as the data is shared and sold in the big data analytics value chain	Creating and maintaining metadata	
(1) Algorithms leading to unreliable outputs by use of subjective, historical data, (2) Algorithms lacking reliability as they rely on correlations and inductive data analysis rather than causal explanations (3) Lack of fit between the algorithms and decision tasks	Reliability of algorithms	Algorithmic Decision Making
(1) Big data analytics relies on automation, which omits the role of human decision maker, (2) Big data analytics uses visualizations to summarise patterns which can distort interpretations and introduce bias	Lack of human involvement	
(1) Humans decision makers and business managers of big data analytics services are unable in most cases to understand or interpret what algorithms are doing, (2) Responsibility of decisions made is not clear	Accountability of decisions	

Figure B2. Data Structures for Organizational Ethical Issues (Cont.)

First-Order Categories	Second-Order Themes	Aggregate Dimensions
(1) Few entities dominate access to big data, (2) Control and influence of a few technology entities over vast populations	Limited access to big data analytics	Power Imbalance
(1) Few entities knowing everything about individuals, their past choices, personalities and preferences (2) Individuals not knowing how big data analytics influence societies at scale.	Knowledge asymmetries	
(1) Governmental bodies, NGOs and associations creating principles and guidelines for organizations delineating the appropriate use of personal information	Establishing principles and guidelines	Principles and Guidelines
(1) Creating laws for use of big data analytics, with adequate sanctions applied when wrongdoings occur, (2) establishing authorities to enforce laws on data use, and (3) policing of the use of big data analytics by organizations	Existence of regulatory authorities	
(1) Oversight on how big data analytics should be used in societies, (2) Re-distribute the benefits of big data analytics appropriately between individuals, organizations and societies	Balancing the benefits and costs of big data	
(1) Large scale monitoring of individuals within the society, (2) Blurring between public and private spaces	Loss of privacy	Surveillance
(1) Manipulating and influencing individuals' choices and behaviour based on analysis of their past choices and behaviours, (2) Encouraging individuals to behave in a certain way that is beneficial for the political and economic objectives of big data analytics services	Regulated behavior	
(1) Participation in social networks, use of apps and other technologies only possible by individuals sharing their data, including location and other sensory data with big data analytics services.	Involuntary participation	Coercion
(1) Ubiquitous dependence of the individuals on the use big data analytics services without knowing the actual purposes, (2) Society driven by corporates and their marketing activities that would increase use of big data services across individuals	Dependence of individuals on big data analytics	

Figure B3. Data Structures for Societal Ethical Issues

Appendix C: Delphi Panel Descriptive Statistics

For our fuzzy set analysis, we created three sets that we labeled I, O, and S to cover the individual, organization, and society stakeholder perspectives, respectively. We then developed a calibration framework to classify our pool of 14 professional roles in terms of how experience in that role would entail exposure to ethical issues of the relevant stakeholder group. We then chose a fuzzy variation scale we expected to see in memberships scores. We chose a common variation of 1 (meaning fully-in), 0.67 (meaning more in than out), 0.33 (meaning more out than in), and 0 (meaning fully-out). We assigned each role membership scores corresponding to the extent to which the role would have membership of each of the three sets (I, O, S). Table C1 shows our calibration framework.

Table C1. Calibration Framework for Assigning Fuzzy Membership Score

Description of sets	Membership scores in sets I, O, S			
	1	0.67	0.33	0
Set I = set of panel members who have the experience to understand ethical issues for individuals	6-12, 14*	4, 5, 13	1, 2, 3	
Set O = set of panel members who have the experience to understand ethical issues for organizations	1, 2, 4, 5, 13, 14	3, 8-12	0	
Set S = set of panel members who have the experience to understand ethical issues for society	6-12, 14	4, 5, 13	1, 2, 3	

* We assigned codes to each role in our Delphi panel as following: business analytics and big data managers (1), chief data and analytics officers (2), senior data scientists (3), business analytics academics (4), information management professionals (5), privacy commissioners (6), ethics committee member of a professional body (e.g., ACM) (7), data-protection officers (8), technology ethics academics, (9), data regulators (10), digital law practitioner (11), privacy law practitioner (12), data ethics consultant (13), IT/ethics editors (14)

To assess coverage of the three stakeholder sets, we compiled data for each panel member based on Linkedin profiles or other publicly available information as to the professional roles in which they had experience. Two researchers assigned the membership scores and discussed the scores until they achieved 100 percent consensus. After we assigned membership scores, we ran descriptive fuzzy analysis to ensure we had minimum and sufficient representation of panel members in each of the sets. Table C2 shows the resulting descriptive statistics that the fsQCA fuzzy set analysis software produced (Ragin, 2008a). Based on the analysis, we had a balanced representation in our panel members with the experience to identify ethical issues for individuals, organizations, and society (coverage of each set ranged from 70 to 86 percent).

Table C2. Fuzzy Set Coverage Descriptive Statistics

Sets	Mean	Std. dev.	Minimum	Maximum	N cases
I (individual)	0.70	0.29	0.33	1	34
O (organization)	0.86	0.16	0.67	1	34
S (society)	0.70	0.29	0.33	1	34

Appendix D: Satisfaction Scores for Theoretical Concepts

Table D1. Second Round Satisfaction Scores for Theoretical Concepts

Concept	Mean*	STD
Concepts related to individuals		
Data ownership	3.63	1.13
Awareness	4.32	0.57
Data control	4.11	0.91
Trust	4.05	0.60
Privacy	4.26	0.96
Self-determination	3.84	1.09
Fear	3.63	0.98
Concepts related to organizations		
Data quality	4.21	0.95
Data sourcing	4.16	0.93
Data sharing	4.05	0.89
Algorithmic decision making	3.68	1.17
Presentation	3.63	1.18
Ethical capability	3.89	1.02
Ethical culture	4.26	0.64
Ethical data governance	4.32	0.80
Ethical performance	4.00	0.92
Reputation	3.89	0.79
Competitive pressure	4.05	0.89
Concepts related to society		
Power	3.84	0.99
Dependence	3.79	0.83
Social awareness	4.00	1.12
Surveillance	4.26	0.91
Principles and guidelines	4.16	0.74
Authority	4.16	0.99
Climate	3.68	0.86
* We measured satisfaction rate in the second round using a scale that ranged between 1 and 5 as a coarse-grain analysis was appropriate for this round.		

Table D2. Third Round Satisfaction Scores for Theoretical Concepts

Concept	Mean	STD
Concepts related to individuals		
Awareness	5.82	1.22
Trust	5.77	1.27
Privacy	5.5	1.59
Choice	5.09	1.66
Concepts related to organizations		
Data trading	5.36	1.73
Ethical governance	5.86	1.44
Data quality	5.82	1.22
Algorithmic decision making	5.41	1.5
Reputation	5.95	1.13
Concepts related to society		
Power imbalance	5.86	1.17
Coercion	5.77	1.02
Surveillance	6.18	0.91
Principles and guidelines	6.59	1.14
* We measured satisfaction rate in the third round using a scale that ranged between 1 and 7 as a finer-grain analysis was appropriate for this round.		

About the Authors

Ida Someh is a lecturer in the Business Information Systems discipline at the UQ Business School, The University of Queensland, Australia, and a research affiliate at the Centre for Information Systems Research (CISR), MIT Sloan School of Management, US. Her research focuses on organizational and societal impact of data, analytics and artificial intelligence. She completed her PhD in 2015 at The University of Melbourne and was awarded the best PhD thesis in Melbourne School of Engineering and the Vice Chancellor's PhD Prize at The University of Melbourne.

Michael Davern, CPA, is Chaired Professor of Accounting and Business Information Systems at the University of Melbourne where he holds appointments in both Business and Engineering. His research explores decision making in Accounting and Information Systems across a diverse range of contexts including analytics, business intelligence, ethics, risk management, IT Value, performance management and financial reporting. For over 25 years, both in Australia and overseas (New York University, University of Minnesota), he has conducted industry engaged research funded by both the corporate sector and Australian Research Council grants. Michael's work appears in leading journals in both information systems and accounting including *Journal of the Association for Information Systems, Journal of Management Information Systems, Journal of Information Technology, Decision Support Systems, Journal of Information Systems,* and *Abacus*, among others. A charter member of the Association for Information Systems, he co-founded the Cognitive Research special interest group.

Christoph Breidbach is a Lecturer at The University of Melbourne, School of Computing and Information Systems. He previously held a postdoctoral position at the University of California Merced and was a visiting researcher at IBM's Almaden Research Center. His empirical and conceptual research addresses the fundamental question of how information technology transforms service ecosystems, and is positioned at the intersection of the Business Information Systems and Service Science disciplines. His publications consistently appear in leading journals such as the *Journal of Service Research*, *Industrial Marketing Management, Managing Service Quality, Journal of Service Theory and Practice, Service Science*, as well as the *ICIS*, *ECIS, ACIS* and *HICSS Proceedings*. He serves on the Editorial Boards of the *Journal of Service Research*, *Journal of Service Theory and Practice*, and the *Journal of Business Research* and leads the AIS Special Interest Group *Services* as elected President.

Graeme Shanks is a Professorial Fellow in the School of Computing and Information Systems at the University of Melbourne. He was previously Associate Dean Research and Professor in the Faculty of Information Technology at Monash University. Before becoming an academic, he worked for a number of private and government organizations as a programmer, systems analysts and project leader. His research interests focus on the management and impact of information systems, ethics issues in big data analytics and enterprise architecture. He has published the results of his research widely in leading information systems journals and conferences.