Problems and Causes of Data Privacy in Big Data Systems in Brazil

Danilo F. Oliveira[a] and Edmir P. V. Prado[b]
Universidade de São Paulo, Brazil

Keywords: Data Privacy, Privacy Issues, Big Data Analytics, Systematic Literature Review.

Abstract: User interactions with computerized systems have led to ethical dilemmas in data use, such as privacy violations. Occasionally, organizations' interests may conflict with the users' privacy interests. Ethical dilemmas arise from this conflict. Furthermore, there is no consensus between organizations and users on the ethical use of data. It is difficult to achieve data privacy in Information Systems that use Big Data Analytics (ISBDA). On the order hand, there is not enough research in the literature on data privacy issues in ISBDA. This study aims to analyze data privacy problems in ISBDA, and their causes, in the Brazilian context. We conducted a systematic literature review to find data privacy problems and their causes. This is exploratory research performed with 16 experts in data privacy and ISBDA, using the Delphi technique for data collection. We identified nine data privacy problems which have seven causes. The research contributed to managerial and organizational practices by associating data privacy problems and their causes.

1 INTRODUCTION

Nowadays, a considerable portion of human interactions is recorded in digital media (Norris & Soloway, 2009). Therefore, Wu *et al.* (2014) and Kitchin (2014) claim that we live in an era of rapid technological transformations in data analysis and processing. The consequence of this technological advance can be positive or negative. However, this depends more on how the technology is applied.

User interactions with computerized systems have led to technical challenges, such as storing and processing large volumes of data, and to ethical dilemmas in the use of data. While most technical challenges have been solved in the academy and IS industry (Kitchin, 2014), ethical dilemmas continue with the same concerns cited by Conger, Loch and Helft (1995), such as privacy violations, and ownership of data, ideas, processes, software code.

Occasionally, organizations' interests may conflict with users' privacy interests. From this conflict arise ethical dilemmas. Furthermore, there is no consensus between organizations and users on the ethical use of data. In this context, Barker *et al.* (2009) claim that several parameters must be considered to understand and assess data privacy risks. However, in Brazil, using personal data with deviation from validity is characterized as a violation of a principle of good faith (ethical) (BRASIL, 2018).

Stahl and Wright (2018) studied ethics in IS, and the topic of data protection and privacy was the most highlighted. However, Singh *et al.* (2018) concluded that there is not enough research in the literature on data privacy issues in Information Systems that use Big Data Analytics (ISBDA). Furthermore, it is difficult to achieve data privacy in ISBDA (Ying & Grandison, 2017), and there is a lack of adequate privacy protection strategies (Wang, 2018). Similarly, Joshi and Kadhiwala (2017) concluded that more research is needed on data privacy management.

In this context, the study of data privacy problems is relevant. This problem intensifies when two aspects are considered: data privacy in ISBDA, as these systems' architectures are quite varied and are used for decision-making (Shaytura *et al.*, 2016), and the Brazilian reality, which has low digital competitiveness compared to developed countries (IMD World Digital, 2020). This fact is corroborated by Abouelmehdi *et al.* (2017), who claim that privacy and data security issues pose the greatest risk in ISBDA.

[a] https://orcid.org/0000-0002-8663-9467
[b] https://orcid.org/0000-0002-3505-6122

Oliveira, D. and Prado, E.
Problems and Causes of Data Privacy in Big Data Systems in Brazil.
DOI: 10.5220/0011766200003467
In *Proceedings of the 25th International Conference on Enterprise Information Systems (ICEIS 2023) - Volume 1*, pages 115-122
ISBN: 978-989-758-648-4; ISSN: 2184-4992

This study aims to analyze data privacy problems in ISBDA, and their causes, in the Brazilian context. Based on this goal, two specific objectives were defined: (1) to identify and describe data privacy problems and their causes based on the literature; and (2) to analyze these problems and causes with experts on data privacy that work in Brazil.

The scope of this study refers to ISBDA but does not include research that discusses data privacy in the context of the internet of things or blockchain, as these topics have specific challenges. Likewise, this research only considers data security issues that directly impact data privacy.

2 LITERATURE REVIEW

This section addresses topics used in this research and found in the literature. The first describes issues related to data privacy and the second to the technological environment of ISBDA.

2.1 Data Privacy

The concept of privacy has broad and diffuse definitions in the literature (Stutzman & Hartzog, 2012). According to Barker *et al*. (2009), it is often assumed that privacy is a globally uniform concept, but this is not always true. The concepts that form the idea of privacy are the right to be left alone, secrecy, control over one's personal information, and intimacy (Solove, 2002). Hartzog (2018) recognizes that there is disagreement in the definition of privacy. This author defines privacy in the IS area as user control over system settings, as it is widely adopted by academics, executives, legislators, regulators, and judges.

For Schaub, Konings and Weber (2015) IS are present in various situations in the daily lives of citizens. This has numerous privacy implications, as these systems can gather and exchange comprehensive information from users with people or companies anywhere in the world. As a consequence, ensuring data protection and privacy has become an issue for companies that use their customers' personal data in their services (Ahmadian et al., 2018). For this reason, the design of an IS needs to consider privacy issues. This idea is related to the concept of "privacy by design", which is a software engineering approach in which privacy is required to be considered at all stages of the software development process (Cavoukian, 2012).

Information security is another important aspect of IS projects. It is divided into three pillars: confidentiality, integrity, and availability (Chen & Zhao, 2012). Confidentiality exists when access to information is restricted to only those who need it. Integrity refers to incorruptible information, and availability means that information must be available to those who need access to it. However, it is important to highlight the difference between data privacy and information security, as this research only addresses the issue of data privacy. Data privacy is limited to the scope of individuals, not organizations. Example: A company's financial information may be confidential, but it has nothing to do with the privacy of individuals. Therefore, a leak of this type of information constitutes a security incident but not a privacy incident. Thus, it is possible to have a privacy violation, such as misuse of personal data, without a security incident within the organization.

2.2 Big Data Analytics

There is no consensus in the literature on the definition of big data. A common and widely accepted definition is the 3Vs, cited by McAfee and Brynjolfsson (2012): volume, variety, and velocity. That is, big data refers to a large volume of data coming from several different sources and in extremely short time intervals. On the other hand, the term "Analytics" is defined by Oxford University (2020) as the analysis of data and statistics carried out systematically by computational means. Analytics is commonly related to big data because big data alone is of little use (Gandomi & Haider, 2015). That is, the potential of big data is only harnessed when used in decision-making.

The union of these two concepts gave rise to BDA (big data analytics), which is a sub-process of extracting insights from big data (Jagadish *et al*., 2014). According to Ranjan and Foropon (2021), despite the growing number of organizations launching BDA initiatives, they have limitations when trying to convert the potential of these initiatives into business value. These authors concluded that organizations of different sizes, structures, and sectors have great difficulties with BDA.

The analysis of the BDA environment is not restricted to the amount of data and its processes for extracting insights. Fan, Han and Liu (2014) claim that part of the causes or solutions of data privacy problems may be linked to the architecture design, or the technologies adopted. According to these authors, the characteristics of big data generate challenges, such as high computational cost, algorithmic instability, and difficulty in aggregating data from multiple sources that use different technologies.

3 RESEARCH METHOD

This is exploratory research with a quantitative approach (Creswell & Creswell, 2021). The methodological procedures are described below in two phases: in the first one, we describe the procedures for identifying problems and causes, and in the second one, we describe the procedures for analyzing problems and causes.

3.1 Identification of Problems and Causes

In this phase, we carried out a bibliographic search, through a systematic literature review (SLR). We adopted the procedure described by Kitchenham *et al.* (2009). The SLR protocol started with the selection of scientific databases. The databases used were: ACM Digital Library (https://dl.acm.org), IEEE Xplore (http://ieeexplore.ieee.org), Scopus (http://www.scopus.com), and Web of Science (https://apps.webofknowledge.com).

RSL aimed to identify data privacy issues and their causes in ISBDA. The data privacy issues in this research focus on people and not on organizations, as it is a study on privacy that impacts individuals' private life. On the other hand, the causes are related to organizations and the way people interact with them. Based on this goal, we defined the following questions for research in the databases: 1) What are the data privacy issues in ISBDA? (2) What are the identified causes for these problems?

For a research to be selected for the systematic review, it was mandatory to meet all eight inclusion criteria (IC): IC1, contain a reference to "privacy" in the title or keywords; IC2, contain a reference to BDA or related terms in the title or keywords; IC3, contains a reference to "problems" or related terms in the title or keywords; IC4, the source of the study must be a conference or journal; IC5, the document type must be a journal or conference article; IC6, the publication must be in English; IC7, the research field of the publications must include IS; and IC8, published from 2016 onwards, to ensure recent research.

For a research to be excluded, it is sufficient to meet an exclusion criterion (EC): CE1, duplicate document in the databases; CE2, access not allowed and not be found in other sources; CE3, not having privacy in ISBDA as an object of study; CE4, research focus being on the internet of things technologies, blockchain, or artificial intelligence; and CE5, not having the objective of studying data privacy problems and their causes in ISBDA.

The data extraction and synthesis strategy were based on the approach suggested by Keshav (2007), and the data extracted from the text were recorded in electronic spreadsheets and text documents.

3.2 Analysis of Problems and Causes

In this phase, we carried out empirical field research with experts in data privacy and ISBDA. We used the Delphi technique for data collection. The Delphi technique is a group facilitation technique through an iterative process of several rounds of questionnaire application designed to transform expert opinion into group consensus (Hasson, Keeney, & McKenna, 2000). An adequate procedure for applying the Delphi technique requires the definition of:

Criteria for the Selection of Panelists. According to Powell (2003), panelists must have experience in the research topic. Furthermore, the group must have diverse trades and professions, as heterogeneous groups produce more quality solutions (Delbecq, Gustafson, & Van De Ven, 1985). Based on these guidelines, we defined the following selection criteria: each panelist must have at least five years of experience with data privacy, and the group of panelists must have a diverse profession.

The Number of Panelists. The number of panelists can vary from 15 to more than 100 (Powell, 2003). However, most of the time it is between 15 and 20 panelists (Hsu & Sandford, 2007). For this research, we defined a minimum of 15 panelists.

Data Processing. We used Kendall's coefficient of agreement (W) to measure the agreement of the panelist's opinions (Schmidt, 1997). The interpretation of this coefficient is shown in table 1.

Table 1: Interpretation of Kendall's coefficient of agreement.

W [0;1]	Agreement
<= 0,1	Very weak
> 0,1 e <= 0,3	Weak
> 0,31 e <= 0,5	Medium
> 0,5 e <= 0,7	Strong
> 0,7	Very strong

Source: Adapted from Schmidt (1997)

We used the W coefficient as a criterion to define the need for a new round or the end of the panel. Thus, the criterion for ending the panel is to achieve more than 80% of agreement. That is, the W coefficient must be bigger than 0.5 for 80% of the analyzed causes. If this level of agreement is not achieved in the third round, the panel must be ended with a divergence between the panelists.

4 RESULTS

The results are lodged in two topics: in the first one, we introduce the problems and causes found in the literature, and in the second one, we introduce the result of the Delphi panel with data privacy experts in Brazil.

4.1 Identification of Problems and Causes

We performed a synthesis of privacy issues in ISBDA using semantic content analysis (Bardin, 2011). In this work, a problem is considered an undesirable or harmful situation to individuals, and that has been caused by a violation of privacy in ISBDA. The summary of problems is shown below with the references.

Threat to Life and Freedom (P1). This issue refers to the threats that individuals may experience due to data privacy incidents. These threats can come from the government itself, in the case of countries with fragile democracies.

Bullying and Discrimination (P2). This is another issue that can be triggered by data privacy incidents. It is possible that an individual suffers psychological violence and has segregation treatment due to sexual, racial, and religious differences disclosed by leaking sensitive personal information.

Reputation (P3). Data privacy incidents can lead to embarrassment or reputational damage to individuals, which can cause irreversible damage to the reputation and esteem of individuals.

Negotiation (P4). Disadvantages in negotiations may occur in cases of data privacy incidents. For example, in the purchase or sale of assets and salary negotiations, among others.

Frauds (P5). Fraud and other crimes can be facilitated or made possible by data privacy incidents. Individuals may be deceived for the benefit of another individual or third-party organizations.

Loss of Anonymity (P6). An individual may lose anonymity at a given time or situation because of data privacy incidents.

Re-identification (P7). Re-identification of anonymized data may occur in several ways, but some of the main ways involve data cross-referencing. Data privacy incidents can facilitate the re-identification of anonymized data by the release of new sensitive information from individuals.

Unauthorized Access (P8). These are incidents related to data theft or unauthorized access.

Illegal Surveillance (P9). Individuals may experience unlawful surveillance by other individuals, organizations, or governments.

We used semantic content analysis (Bardin, 2011) to categorize the various causes found in the literature. In this work, the cause is defined as the origin of a data privacy problem in ISBDA. The summary of causes is shown below with the references.

Vulnerability (C1). Represents causes associated with vulnerability and lack of security in organizations that handle data. Factors such as external attacks by hackers or malware and weak encryption keys are included. Other factors encourage or facilitate attacks, such as the high market value of data (healthcare), low concern for security in specific industries, and use of third-party tools.

Management of ISBDA (C2). Causes associated with inefficient data management and non-compliance with good practices and regulations in organizations. This includes a lack of purpose and transparency in the use of data, a change in the purpose of data use, and improper retention of data.

Technical Challenges (C3). Causes associated with technical difficulties in protecting privacy in ISBDA, mainly due to volume, speed, and variety of data types. A major technical challenge related to ISBDA is dealing with many data coming from several different sources. Furthermore, anonymizing this data in storage, transmission, or publication is highly complex. Other technical challenges are the use of personally identifiable information and the improper publication of data.

User Skills (C4). The lack of skill of ISBDA users, such as misinformation, ignorance, or lack of care regarding privacy, can lead to a lack of control over their data. The lack of consent to the use of data by the user is also part of this type of cause.

Control of Access (C5). This cause refers to poor management of access to data, such as illegal access, unauthorized access, or overly granular access. Also, this cause includes improper access by third parties and insufficient access control over the organization's data.

Management and Culture (C6). This cause is associated with organizational and cultural deficiencies in data management. This category includes inadequate accountability, lack of internal regulations, malicious behavior by employees or third parties working for the company, lack of a privacy culture, lack of support from senior management, and lack of technical training for teams.

Inference (C7). They represent causes associated with the improper re-identification of supposedly anonymized data. This includes inference from original data that can be cross-referenced with other, more granular databases; or from new data.

4.2 Analysis of Problems and Causes

We use the Delphi panel to make the association between causes and problems. Panelists were selected from a survey of data management professionals working in Brazil on the LinkedIn social network. 59 professionals were selected, of which 16 participated in the Delphi panel. All 16 panelists had at least five years of professional experience, with seven of them performing technical activities and nine performing management activities. Furthermore, nine panelists worked in the field of data analysis, three in data science, twelve in data engineering, and eight in data governance. Thus, the sample is in accordance with the necessary minimum number of panelists and with the expected diversity of professional activities regarding data.

Panelists received the questionnaire and instructions on how to complete it by e-mail. In the first round, only one cause (14.3% of the causes) had a W agreement coefficient above 0.5. According to the criterion adopted, it was necessary to carry out a second round.

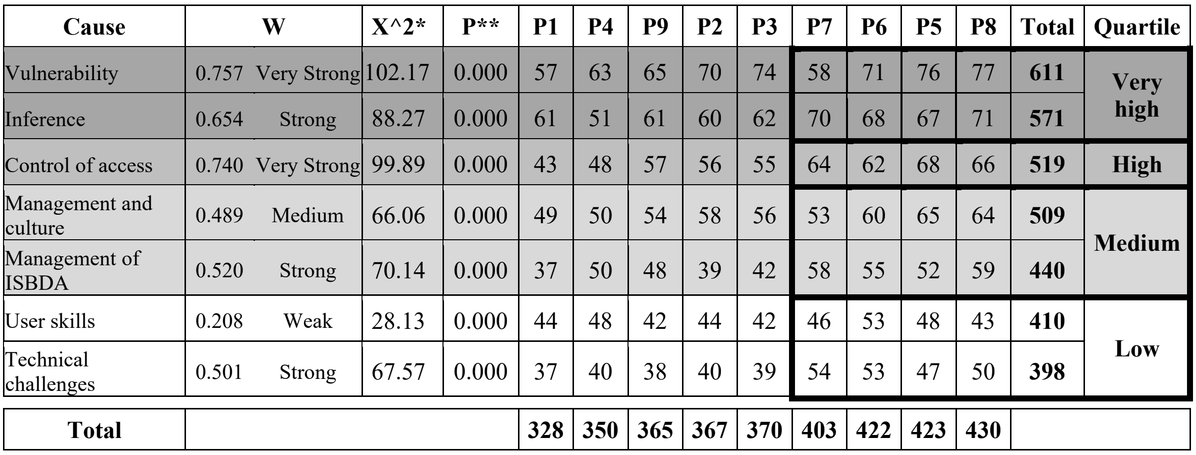
In the second round, 11 panelists agreed to change their answers, and the level of agreement increased. Six causes (85.7% of the causes) had a W concordance coefficient above 0.5. According to the criterion adopted, it was not necessary to carry out additional rounds, and the panel ended with a convergence of opinions among the panelists.

Figure 1 shows the results. In five causes, there was strong or very strong agreement; in one cause, the agreement was medium. We used the median statistic to highlight the highest-scoring problems: P5, P6, P7, and P8. The causes that most address the problems were classified by the quartile statistics, as also indicated in Figure 1.

5 DISCUSSIONS

In this section, we analyze the outcomes of the Delphi panel. Two types of analysis were performed. The first analysis was applied to the entire group of panelists, and the second to subgroups of panelists according to the moderating variables.

As shown in Figure 1, the main problems pointed out by the panelists were P5, P6, P7, and P8, as these problems had scores above the median (509). Similarly, the main causes pointed out by the panelists were C1, C5, and C7, as these causes had scored above the median (370). We used the quartile statistics and identified that causes C1 and C7 had very high importance, followed by the cause C5 with high importance. We also emphasize that the degree



Cause	W		X^2*	P**	P1	P4	P9	P2	P3	P7	P6	P5	P8	Total	Quartile
Vulnerability	0.757	Very Strong	102.17	0.000	57	63	65	70	74	58	71	76	77	**611**	**Very high**
Inference	0.654	Strong	88.27	0.000	61	51	61	60	62	70	68	67	71	**571**	
Control of access	0.740	Very Strong	99.89	0.000	43	48	57	56	55	64	62	68	66	**519**	**High**
Management and culture	0.489	Medium	66.06	0.000	49	50	54	58	56	53	60	65	64	**509**	**Medium**
Management of ISBDA	0.520	Strong	70.14	0.000	37	50	48	39	42	58	55	52	59	**440**	
User skills	0.208	Weak	28.13	0.000	44	48	42	44	42	46	53	48	43	**410**	**Low**
Technical challenges	0.501	Strong	67.57	0.000	37	40	38	40	39	54	53	47	50	**398**	
Total					**328**	**350**	**365**	**367**	**370**	**403**	**422**	**423**	**430**		

* Chi-square statistic; ** P value. Values less than or equal to 0.05 have a statistical significance level of 5%.

Figure 1: Delphi panel outcome.

of agreement for these three causes was very strong or strong (according to Kendall's coefficient of agreement, represented by W), and the outcome has a statistical significance level of 5% (p-value).

We analyzed the ranking of causes and problems through the moderating variables – a position held and type of expertise – and observed that the ranking is very similar (figure 2). Only panelists belonging to the category of Managers showed a lower convergence (66.7%) with the category of Technicians. Therefore, it is possible to infer that the position held by the panelist can influence the ranking of problems and causes. Other surveys in Brazil, using the Delphi technique, also showed differences between managers and technicians (Souza, 2012; Ayabe, 2021).

The problems most highlighted by experts were unauthorized access to data, fraud, and other crimes. While problems such as the threat to life or liberty and illegal surveillance were less highlighted. These last problems are more persistent in countries with non-democratic political regimes and were not highlighted by specialists in Brazil.

The causes most associated with the problems were vulnerability (C1), inference (C7), control of access (C5) management, and culture (C6). On the other hand, technical challenges (C3) had a lower score. We can infer that the technical challenges of BDA reported in the literature, such as the 3Vs (McAfee and Brynjolfsson, 2012), are less relevant to the Brazilian context than, for example, the management of security systems.

6 CONCLUSIONS

The research has contributed to knowledge in the field of data privacy and the context of developing countries, such as Brazil. In this section, we describe the objectives achieved, the limitations and contributions of the research, and the next steps of the research based on the results.

6.1 Research Goals

The goal of this research was to analyze data privacy issues in ISBDA. We achieved this goal through the application of a Delphi panel and the outcomes are summarized below:

Data Privacy Issues and their Causes. A literature search identified nine data privacy issues in ISBDA and seven causes of these issues.

Analysis of Problems and Causes. The analysis was performed by experts using a Delphi panel. Experts agreed on six of the seven causes. Furthermore, the agreement between the panelists remained high regardless of their professional activities, but there was a difference of opinion between specialists with a managerial role and with a technician role.

6.2 Research Limitations

Sample. This research was carried out with 16 professionals working in the field of data privacy in Brazil. These experts were selected based on their professional relationship with the research authors. Therefore, the results cannot be generalized.

Causes	All	Function		All x Function			Expertise		All x Expertise		
		Manager	Technician				Generalist	Non-generalist			
	(A)	(B)	(C)	AxB	AxC	BxC	(D)	(E)	AxD	AxE	DxE
C1	1	1	1	0	0	0	1	1	0	0	0
C7	2	2	2	0	0	0	2	2	0	0	0
C5	3	3	4	0	1	1	3	3	0	0	0
C6	4	4	3	0	1	1	4	4	0	0	0
C2	5	3	6	2	1	3	6	5	1	0	1
C4	6	7	5	1	1	2	5	7	1	1	2
C3	7	6	7	1	0	1	7	6	0	1	1
Quantity of differences				**4**	**4**	**8**			**2**	**2**	**4**
Degree of concordance				**83,3**	**83,3**	**66,7**			**91,7**	**91,7**	**83,3**

Figure 2: Ranking of causes with the entire sample and by subgroups.

Interpretation. The different causes and problems identified in the literature were grouped based on the content analysis technique, which has subjective characteristics, that is, different researchers could find different categories.

Technological Evolution. The evolution of information technologies can change the importance of the causes and problems identified in this research.

6.3 Contributions

The research contributed to the field of data privacy, as it identified the problems and causes of data privacy reported in the literature, elaborating a summary of them, which allows new studies to address the problems and causes most highlighted in the literature.

The research contributed to managerial and organizational practices through the identification and association of causes and problems. From this association, organizations can prioritize data privacy protection actions in ISBDA in order to optimize efforts and minimize risks. Among the four main causes identified, two are related to management aspects. This draws attention for organizations to consider not only the use of technology but also information security management practices.

6.4 Proposals

This research had an exploratory approach and associated data privacy problems with their causes. On the other hand, it did not describe how the causes contribute to the problems, nor did it propose the adoption of technical or managerial solutions to the problems. Therefore, the next steps of the research aim to identify a set of solutions and good practices that address the causes of the problems identified in this research.

REFERENCES

Ahmadian, A. S., Strüber, D., Riediger, V., & Jürjens, J. (2018). Supporting privacy impact assessment by model-based privacy analysis. *Proceedings of the ACM Symposium on Applied Computing*, 1467–1474. https://doi.org/10.1145/3167132.3167288

Ayabe, F. (2021). Fatores críticos de sucesso para terceirização de tecnologia da informação no setor público brasileiro. https://teses.usp.br/teses/disponiveis/100/100131/tde-16102018-102401/en.php

Bardin, L. (2011). *Análise de conteúdo* (Issue 70).

Barker, K., Askari, M., Banerjee, M., Ghazinour, K., MacKas, B., Majedi, M., Pun, S., & Williams, A. (2009). A data privacy taxonomy. *Lecture Notes in Computer Science*, *5588 LNCS*, 42–54. https://doi.org/10.1007/978-3-642-02843-4_7

Brasil. (2018). *Lei Geral de Proteção de Dados Pessoais*. Diário Oficial Da União. http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm

Cavoukian, A. (2012). Privacy by design [leading edge]. *IEEE Technology and Society Magazine*, *31*(4), 18–19. https://doi.org/10.1109/MTS.2012.2225459

Chen, D., & Zhao, H. (2012). Data security and privacy protection issues in cloud computing. *ICCSEE 2012*, *1*(973), 647–651. https://doi.org/10.1109/ICCSEE.2012.193

Colesky, M., Hoepman, J. H., & Hillen, C. (2016). A Critical Analysis of Privacy Design Strategies. *IEEE SPW 2016*, 33–40. https://doi.org/10.1109/SPW.2016.23

Conger, S., Loch, K. D., & Helft, B. L. (1995). Ethics and information technology use: a factor analysis of attitudes to computer use. *Information Systems Journal*, *5*(3), 161–183. https://doi.org/10.1111/j.1365-2575.1995.tb00106.x

Constantiou, I. D., & Kallinikos, J. (2015). New games, new rules: Big data and the changing context of strategy. *Journal of Information Technology*, *30*(1), 44–57.

Cooper, A. (2012). What is "Analytics"? Definition and Essential Characteristics. *CETIS Analytics Series*, *1*(5), 1–10. http://publications.cetis.ac.uk/2012/521

Creswell, J. W., & Creswell, J. D. (2021). *Projeto de pesquisa: Métodos qualitativo, quantitativo e misto*. Penso.

Delbecq, A. L., Gustafson, D. H., & Van De Ven, A. H. (1985). *Group Techniques for Program Planning: A Guide to Nominal Group and Delphi Processes*. https://doi.org/10.1177/105960117600100220

Fan, J., Han, F., & Liu, H. (2014). Challenges of Big Data analysis. *National Science Review*, *1*(2), 293–314. https://doi.org/10.1093/nsr/nwt032

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Google. (n.d.). *Google Trends*. https://trends.google.com/trends/explore?date=today 5-y&q=Big Data, Data Analytics

Google. (2020). *Google Translator*. http://translate.google.com/

Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing*, 32(4), 1008-1015.

Hartzog, W. (2018). The Case Against Idealising Control. *European Data Protection Law Review*, *4*(4), 423–432. https://doi.org/10.21552/edpl/2018/4/5

Hsu, C. C., & Sandford, B. A. (2007). The Delphi technique: Making sense of consensus. *Practical Assessment, Research and Evaluation*, 12(10), 1-8.

IMD World Digital. (2020). IMD World Digital Competitiveness Ranking 2020. *IMD World Competitiveness Center*, 180.

Jagadish, H. V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., Ramakrishnan, R., & Shahabi, C. (2014). Big data and its technical challenges. *Communications of the ACM*, *57*(7), 86–94. https://doi.org/10.1145/2611567

Keshav, S. (2007). How to read a paper. *ACM SIGCOMM Computer Communication Review*, *37*(3), 83–84. https://doi.org/10.1145/1273445.1273458

Kitchenham, et al. (2009). Systematic literature reviews in software engineering – A systematic literature review. *Information and Software Technology*, *51*(1), 7–15. https://doi.org/10.1016/j.infsof.2008.09.009

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data and Society*, *1*(1). https://doi.org/10.1177/2053951714528481

Mcafee, A., & Brynjolfsson, E. (2012). Spotlight on Big Data Big Data: The Management Revolution, 2012. *Harvard Business Review*, *October*, 1–9.

Müller, O., Junglas, I., Brocke, J. Vom, & Debortoli, S. (2016). Utilizing big data analytics for information systems research: Challenges, promises and guidelines. *European Journal of Information Systems*, *25*(4), 289–302. https://doi.org/10.1057/ejis.2016.2

Norris, C., & Soloway, E. (2009). A disruption is coming. A primer for educators on the mobile technology revolution. *Mobile Technology for Children*, 83–98. https://doi.org/10.1016/B978-0-12-374900-0.00005-3

Oxford University. (2020). *Oxford English Dictionary*.

Powell, C. The Delphi technique: Myths and realities. (2003). *Journal of Advanced Nursing*, 41(4), 376-382.

Ranjan, J., & Foropon, C. (2021). Big Data Analytics in Building the Competitive Intelligence of Organizations. *International Journal of Information Management*, *56*, p. 102231. https://doi.org/10.1016/j.ijinfomgt.2020.102231

Schaub, F., Konings, B., & Weber, M. (2015). Context-Adaptive Privacy: Leveraging Context Awareness to Support Privacy Decision Making. *IEEE Pervasive Computing*, *14*(1), 34–43. https://doi.org/10.1109/MPRV.2015.5

Shaytura, S. V., Stepanova, M. G., Shaytura, A. S., Ordov, K. V., & Galkin, N. A. (2016). Application of Information-Analytical Systems. *Journal of Theoretical and Applied Information Technology*, *90*(2), 10-22.

Schmidt, R. C. (1997). Managing Delphi surveys using nonparametric statistical techniques. *Decision Sciences*, 28(3), 763-774.

Solove, D. J. (2002). Conceptualizing privacy. *California Law Review*, *90*(4), 1087–1155. https://doi.org/10.2307/3481326

Souza, A. M. (2012). Uso de SGBDs nas organizações: uma aplicação em banco de dados não relacionais. https://teses.usp.br/teses/disponiveis/100/100131/tde-26112013-181716/en.php

Stahl, B. C., & Wright, D. (2018). Ethics and Privacy in AI and Big Data: Implementing Responsible Research and Innovation. *IEEE Security and Privacy*, *16*(3), 26–33. https://doi.org/10.1109/MSP.2018.2701164

Stutzman, F., & Hartzog, W. (2012). Boundary regulation in social media. *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, 769–778. https://doi.org/10.1145/2145204.2145320

Wall, J. D., Lowry, P. B., & Barlow, J. B. (2016). Organizational violations of externally governed privacyand security rules: Explaining and predicting selective violations under conditions of strain and excess. *Journal of the Association for Information Systems*, *17*(1), 39–76.

Wang, K. (2018). A survey on risks of big data privacy. *Advances in Intelligent Systems and Computing*, 580, 161-167.

Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2014). Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering*, *26*(1), 97–107. https://doi.org/10.1109/TKDE.2013.109

Ying, S., & Grandison, T. (2017). Big data privacy risk: Connecting many large data sets. IEEE International Conference on CIC, p. 86-91.