EmAIT

Journal homepage: http://publisher.uthm.edu.my/ojs/index.php/emait
e-ISSN : 2773-5540

Emerging Advances in Integrated Technology

Conceptual Paper: Sentience of Big Data towards User Privacy Concerns and Online Self-Disclosure Activities

A Ismail[1*], M R Hamzah[1], H Hussin[1]

[1]Faculty of Applied and Human Sciences,
Universiti Malaysia Perlis Pengkalan Jaya, Jalan Alor Setar-Kangar, 01000 Kangar, Perlis, MALAYSIA

*Corresponding Author

DOI: https://doi.org/10.30880/emait.2021.02.02.006
Received 27 September 2021; Accepted 18 October 2021; Available online 15 November 2021

Abstract: Big data allows widespread use and exchange of user data, and this will lead to the possibility of privacy breaches. Governments and corporations will incorporate personal data from different sources and learn a great deal about people and in turn, raise concerns about privacy. This paper will provide a conceptual understanding on the antecedents towards user privacy concerns and online self-disclosure activities, which are the knowledge and perceived risks of big data. In this paper, big data knowledge is hypothesized to decrease privacy concerns, meanwhile perceived risks is suggested to increase the outcome. Based on the framework, propositions are formulated as a basis for the study that will follow.

Keywords: Big data knowledge, big data perceived risks, online privacy concerns, self-disclosure

1. Introduction

Since the beginning of internet era, privacy has been one of the main concerns among the users. With big data being used vastly by corporations, extracted from information available everywhere, there is a need to understand whether the providers of those information; the generators themselves, understand the risks they are facing with the provision of those data. According to Alashoor et al. [1], there is proof of increased privacy concerns among users of the internet and social networking sites (SNS), and this is a significant problem when we recognize the reality of the digital age today. In particular, these issues are justified because the online activities of users are not only monitored, but most sites also use private information for targeted ads, and a lot of big companies such as Google, Yahoo, Microsoft and Facebook share their gathered customer data with many of their associated companies [2], [3].

Any kind of information, regardless of its form, or whether they are containing official or personal data; can be easily captured, extracted, analysed, and used by a third party once they are published online. SNS maintains massive databases containing a vast volume of private data collected [1]. They place users in jeopardy by actually risking them to identity fraud, humiliation, and other dangers as malicious parties breach or quickly access the information stored in these databases [4], [5]. Such widespread use and exchange of customer data includes the possibility of data breaches and the related negative psychological effects on people [2]. Governments and corporations will incorporate personal data from different sources and learn a great deal about what we are doing, where we go, who our acquaintances are, and also what our interests are [5], and this would, in turn, raise concerns about privacy.

2. Big Data and Its Definition

For a number of information systems, information overload is a serious issue. In that sense, the word 'big data' was coined for the first time in 1997 [6]. Data volume began to outstrip storage system capacities due to prompt growth, necessitating the creation of computer database systems capable of accommodating available data resources without the need to discard any of the collected data [7]. In a study published by Bryson et al. [8] the value of real-time analytics

*Corresponding author: adilaismail@unimap.edu.my

and the ability to extract only the details, patterns, and relationships required for a specific research project was highlighted. It has everything to do with the relatively latest idea 'big data' [7].

Big data is characterised as an information asset with such a high volume, velocity, and variety that its transformation into value necessitates the use of specialised technology and analytical methods [9]. It is an unstructured data collection that can be captured, preserved, processed, as well as handled in vast quantities. This vast volume of data will be worthless before it is inspected and analysed by big data analytics, that will expose previously unknown associations, unseen trends, and other useful knowledge [10]. These data include structured, semi / unstructured data, incomplete, or multimedia such as text, video, image, in various formats. Text messages, documents saved, emails, posts in social media, videos / audios / photographs, graphs, and other outputs from different forms of computer-generated data from sensor devices, RFID tags, machine logs, mobile phone global positioning system (GPS) signals, and DNA analysis devices are all used to collect this information [11].

3. Online Privacy Associated with Big Data

Big data is being mentioned everywhere, and among the top users and exploiters of big data are business organizations. Davenport [12] these companies monitor, research, and analyse social media users' search, sharing, and interest trends, and then use that data to push / advertise similar interests to potential customers. The data analysis will enable them to decide if the feedback about their products and brands in social media such as on Facebook pages / Twitter / blogs are on the whole positive or negative. Such data assists in assessing stakeholders, decision processes, and the requirements and timeframes for making resolutions [12]. One of the main advantages of big data is that it helps companies to capture, store, analyse, and handle the information of their users. Analysing social media data will help companies better understand consumer behaviour and target goods and services [13].

Today, human actions leave imprints that can be easily recorded, preserved, and aggregated [14]. New technologies allow the collection and efficient use of big data to increase knowledge understanding, speed decision-making, and provide numerous opportunities for social interaction [15]. It is important to remember, however, that allowing others to gather our information and openly do stuff with it poses a major risk. Social media, especially SNS with large user bases like Facebook, LinkedIn, Twitter, and Instagram, is an ideal information hunting ground. Big data, according to Chandler [16], contains information from a variety of sources, such as social media, mobile phones, visualising, mapping, and recording equipments, and the number of data-sharing devices is increasingly rising. According to Illmer [17], the majority of current attacks simply use social media sites as a distribution tool, but researchers are now predicting that sophisticated attacks on social media platforms would be able to monitor a user's contacts, venue, and even business activities. This knowledge can then be used to create tailored promotional campaigns for individual people, or even to assist in the instigation of crime in the virtual or real world [17]. Big data management applications, especially those requiring the processing of personally identifying details and other confidential data, have major social ramifications [7].

Sætra [18] has come to the conclusion that our liberty is under attack through big data. He clarified here that big data violates privacy and allows for surveillance. Big data has contributed to new data analytics and data analysis standards, according to Bello-Orgaz et al. [19]. In order to make use of the full benefit of big data, mass amount of information is needed for analysis and processing. This information may be gathered from those transactions we made from our daily activities, and one of them is from our SNS. In the rapidly growing archive collections, we leave endless signs of our everyday lives [18]. The data gathered by corporations and the government is important, and consumers are willing to provide them with access to different resources without understanding it. Big data in contemporary culture means processing large quantities of data from people, including the most private information, and therefore, as huge amounts of private data are obtained, privacy is, of course, under threat.

4. Big Data familiarity towards Privacy Concerns and Online Self-disclosure

Individuals' understanding of the term big data, and its relevance to them and to companies, as well as the four main implications of big data analytics, which include collection, storage, sorting, and sharing, is referred to as familiarity with big data [1]. Many people are now conscious that online companies are profiting from their personal information, which explains their worries over privacy and transparency policies [20]–[23]. Researchers also expect that as people become more aware of big data, they will be more concerned with privacy [24], [25]. Users should be conscious, though, that big data may provide them with potential benefits [1]. This means that the sentience of big data construct has two contrasting emphases, which are big data knowledge and perceived big data risks.

The rise of big data and its far-reaching consequences would almost definitely alter our views of data protection in the social media environment [1]. The issue of big data protection has already been addressed by a range of researchers and industry leaders [1], [5], [26]–[29]. Individuals have no awareness and concern about big data and its consequences, according to Clemons et al. [24]. According to Wieczorkowski and Polak [7], the collection and application of big data poses concerns regarding the real world's perception of privacy and the indication of a balance between the effective operation of the state or corporation and the security of human rights.

Alashoor et al. [1] argued that the users who are aware of big data and the opportunities it can offer are less likely to be worried with privacy because they realise that in order to benefit from today's technologies, they must give up some individual privacy. This forecast is in line with research results that suggest a negative relationship between internet literacy and privacy concerns [30]–[32]. Krasnova et al. [33] also found some evidence for a correlation between privacy concerns and understanding of how social media platforms manage information. The studies by Westin [34] and Sheehan [35] showed that at least half of the population were ready to share their information in return for rewards. According to Wieczorkowski and Polak [7], when it comes to issues about the general public's protection, privacy breaches are acknowledged. While there is no direct connection between this partnership and awareness of big data methods, the respondents are likely aware of the emerging danger to public safety and the feasibility of avoiding it by using mass data processing methods. According to the findings of a survey conducted by Wieczorkowski and Polak [7] showed that users have a high degree of confidence in government agencies, as in a high degree of recognition of privacy invasion for common social purposes.

Being aware of the various uses of personal data at the same time, on the other hand, may lead to heightened questions regarding personal privacy and disclosure habits [36]. According to Kamakshi [37] users of information technology and other people whose data is stored in information systems perceive a threat to their privacy. Recent polls shore up these conflicting personal views regarding governmental surveillance among US citizens: they find that while most people approve of governmental surveillance for security purposes, they are also worried about the safety of personal correspondence [38]. In addition, Wieczorkowski and Polak [7] also mentioned that personal information is gathered and stored by different public agencies to ensure the integrity of the state. These agencies derive data from open databases such as Closed Circuit Television (CCTV), aerial photos and satellite images, public internet information, and so on, as well as private data gathered by exclusive privileges provided by numerous agencies such as the police and special services, such as data from telephony: location and billings; private internet content: e-mails, various archives, and so on. It is also reported that these data could have been collected improperly by those organisations.

Furthermore, there is also a concern regarding the lower degree of support of big data use, relating to the use of personal data for commercial or advertisement purposes. In the corporate sector, systems are used to manage data that is directly connected to a particular user. Several examples of data processing practises collected by organisations that are likely to be used for commercial purposes were identified by Wieczorkowski and Polak [7];

- The collection of information about shopping and person location is related to the tracking and processing of financial purchases made by payment cards.
- In order to monitor the whole network, real-time monitoring of a power grid and power usage consists in the collection of data on human actions, such as by turning on the electrical devices.
- The electronic processing of accurate medical reports requires the management of confidential details about particular patients' health.
- Traffic monitoring, which involves the automated tracking of vehicle position, is connected to the collection of data regarding human actions and location.

5. Conceptual Framework

Personal data has been processed for a long time, but big data has opened up a world of possibilities for handling individual private information, not just cumulative data on culture, in the last few years. Its most pressing challenge right now is maintaining an acceptable degree of privacy. As previously mentioned, there is a conflict with users to approve or support the use of their data for particular reasons, as they are also worried with the gathering company violating their privacy. Thus, knowledge of big data and its advantages is suggested to contribute hypothetically to reducing privacy concerns. Otherwise, perceived big data risks will lead to growing privacy concerns.

The framework proposed in this paper is adapted from a previous work by Alashoor et al. (2017), which proposed that the familiarity of big data will have differing effects towards privacy concerns and two self-disclosure outcomes (self-disclosure accuracy and self-disclosure concerns). According to the authors, as people become more aware of big data, they will be more concerned with privacy. However, they also argued that argued that the users who are aware of big data and the opportunities it can offer are less likely to be worried with privacy because they realise that in order to benefit from today's technologies, they must give up some individual privacy. This means that in their proposed framework, the familiarity of big data construct has two contrasting emphases: (1) awareness of big data and (2) awareness of big data implications.

Agreeing with Alashoor et al., the authors proposed a modified framework, which is more simplified since it will only be tested towards one final outcome. The proposed framework is shown in Figure 1.

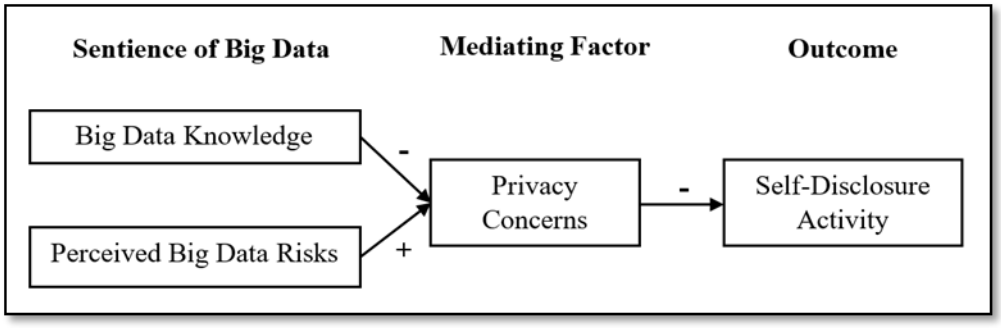

Fig. 1 - Proposed framework

Big data has opened up a world of possibilities for handling individual private information, not just cumulative data on culture, in the last few years. Its most pressing challenge right now is maintaining an acceptable degree of privacy. As previously mentioned, there is a conflict with users to approve or support the use of their data for particular reasons, as they are also worried with the gathering company violating their privacy. In this framework, the knowledge of big data and its advantages is suggested to contribute hypothetically to reducing privacy concerns. Otherwise, perceived of big data risk will lead to growing privacy concerns. Disclosing information online would be hampered by a higher privacy concern. In this paper, it is concluded that privacy concerns will negatively impact self-disclosure activities among users.

6. Summary

The testable propositions presented in this conceptual framework offer an opportunity for further investigation on the sentience of big data towards self disclosure activities, especially in through variety of research designs and settings. The survey research designs employing social media users would best match the requirements for validating the proposed framework, and prospective researchers intending to adopt the model should also consider incorporating additional dimensions of other factors that could have been other antecedents or predictors. The studies can then provide education policy makers, educators, and those interested in the area with an understanding of how the users' knowledge of big data may influence their future decision to self-disclose online.

References

[1] T. Alashoor, S. Han, R.C. Joseph, Familiarity with big data, privacy concerns, and self-disclosure accuracy in social networking websites: An APCO model. Commun. Assoc. Inf. Syst. 41, 62–96, 2017

[2] J.H. Benamati, Z.D. Ozdemir, J. Smith, An empirical test of an Antecedents - Privacy Concerns - Outcomes model. J. Inf. Sci. 43, 583–600, 2016

[3] J. Gomez, T. Pinnick, A. Soltani, The current state of web privacy, data collection, and information sharing. KnowPrivacy http://knowprivacy.org/report/KnowPrivacy_Final_Report.pdf, 2009

[4] B. Choi, Z.J. Jiang, B. Xiao, S.S. Kim, Embarrassing exposures in online social networks: An integrated perspective of privacy invasion and relationship bonding. Inf. Syst. Res. 26, 675–694, 2015

[5] H.J. Watson, Tutorial: Big data analytics: Concepts, technologies, and applications. Commun. Assoc. Inf. Syst. 34, 1247–1268, 2014

[6] M. Cox, D. Ellsworth, Application-controlled demand paging for out-of-core visualization. in Proceedings of the IEEE Visualization Conference 235–244, 1997. doi:10.1109/visual.1997.663888

[7] J. Wieczorkowski, P. Polak, Big data and privacy: The study of privacy invasion acceptance in the world of big data. Online J. Appl. Knowl. Manag. 5, 57–71, 2017

[8] S. Bryson, D. Kenwright, M. Cox, D. Ellsworth, R. Haimes, Visually exploring gigabyte data sets in real time. Commun. ACM 42, 82–90, 1999

[9] A. De Mauro, M. Greco, M. Grimaldi, A formal definition of big data based on its essential features. Libr. Rev. 65, 122–135, 2016

[10] E.M. Tachizawa, M.J. Alvarez-Gil, M.J. Montes-Sancho, How “smart cities” will change supply chain management. Supply Chain Manag. 20, 237–248, 2015

[11] E. Bertino, P. Bernstein, D. Agrawal, S. Davidson, U. Dayal, M. Franklin, J. Gehrke, L. Haas, A. Halevy, J. Han, Challenges and opportunities with big data. A Community White Pap. Dev. by Lead. Res. Across United States 11, 72–74, 2011

[12] T.H. Davenport, How strategists use ‘big data’ to support internal business decisions, discovery and production. Strateg. Leadersh. 42, 45–50, 2014

[13] H.J. Esfahani, K. Tavasoli, A. Jabbarzadeh, Big data and social media: A scientometrics analysis. Int. J. Data Netw. Sci. 145–164, 2019 doi:10.5267/j.ijdns.2019.2.007
[14] Z. Tufekci, Big questions for social media big data: Representativeness, validity and other methodological pitfalls. in Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014, 2014
[15] S. Chauhan, N. Agarwal, A.K. Kar, Addressing big data challenges in smart cities: A systematic literature review. Info 18, 73–90, 2016
[16] D. Chler, A world without causation: Big data and the coming of age of post humanism. Millenn. J. Int. Stud. 43, 833–851, 2015
[17] A. Illmer, Social media: A hunting ground for cybercriminals. Technology of Business, News, 2016
[18] H.S. Sætra, Freedom under the gaze of Big Brother: Preparing the grounds for a liberal defence of privacy in the era of big data. Technol. Soc. 58, 101160, 2019
[19] G. Bello-Orgaz, J.J. Jung, D. Camacho, Social big data: Recent achievements and new challenges. Inf. Fusion 28, 45–59, 2016
[20] B. Choi, L. Land, The effects of general privacy concerns and transactional privacy concerns on Facebook apps usage. Inf. Manag. 53, 868–877, 2016
[21] H. Krasnova, E. Kolesnikova, O. Günther, 'It won't happen to me!': Self-disclosure in online social networks. in 15th Americas Conference on Information Systems 2009, AMCIS 2009 vol. 4 2559–2567, 2009
[22] N.K. Malhotra, S.S. Kim, J. Agarwal, Internet users' information privacy concerns, IUIPC: The construct, the scale, and a causal model. Inf. Syst. Res. 15, 336–355, 2004
[23] J. Turow, M. Hennessy, Internet privacy and institutional trust: Insights from a national survey. New Media Soc. 9, 300–318, 2007
[24] E.K. Clemons, J. Wilson, Investigations into consumers preferences concerning privacy: An initial step towards the development of modern and consistent privacy protections around the globe. in Proceedings of the Annual Hawaii International Conference on System Sciences 4083–4092, 2014. doi:10.1109/HICSS.2014.504
[25] W. Xie, K. Karan, Consumers' privacy concern and privacy protection on Facebook in the era of Big Data: Empirical evidence from college students. J. Interact. Advert. 19, 187–201, 2019.
[26] M. Altman, A. Wood, D.R. O'Brien, U. Gasser, Practical approaches to big data privacy over time. Int. Data Priv. Law 8, 29–51, 2018
[27] D. Boyd, K. Crawford, Critical questions for big data. Information, Commun. Soc. 15, 662–679, 2012.
[28] D. Breznitz, M. Murphree, S. Goodman, Ubiquitous data collection: Rethinking privacy debates. Computer, Long. Beach. Calif. 44, 100–102, 2011
[29] J.P. Shim, A.M. French, C. Guo, J. Jablonski, Big data and analytics: Issues, solutions, and ROI. Commun. Assoc. Inf. Syst. 37, 797–810, 2015
[30] T. Dinev, P. Hart, Internet privacy concerns and social awareness as determinants of intention to transact. Int. J. Electron. Commer. 10, 7–29, 2006
[31] L. Baruh, E. Secinti, Z. Cemalcilar, Online privacy concerns and privacy management: A meta-analytical review. J. Commun. 67, 26–53, 2017
[32] W. Hong, F.K.Y. Chan, J.Y.L. Thong, Drivers and inhibitors of internet privacy concern: A multidimensional development theory perspective. J. Bus. Ethics, 2019 doi:10.1007/s10551-019-04237-1
[33] H. Krasnova, N.F. Veltri, W. El Garah, Effectiveness of justice-based measures in managing trust and privacy concerns on social networking sites: An intercultural perspective. Commun. Assoc. Inf. Syst. 35, 83–108, 2014
[34] A.F. Westin, Privacy in the workplace: How well does American law reflect American values? Chic. Kent. Law Rev. 72, 271–283, 1996
[35] K.B. Sheehan, Toward a typology of internet users and online privacy concerns. Inf. Soc. 18, 21–32, 2002
[36] H. Krasnova, O. Günther, S. Spiekermann, K. Koroleva, Privacy concerns and identity in online social networks. Identity Inf. Soc. 2, 39–63, 2009
[37] P. Kamakshi, Survey on big data and related privacy issues. Int. J. Res. Eng. Technol. 3, 68–70, 2014
[38] G. Gao, What Americans think about NSA surveillance, national security and privacy. Pew Res. Cent., 2015