Data Privacy, What Still Need Consideration in Online Application System?

Dewi Puspasari
Faculty of Computer Science,
Universitas Indonesia
Kampus UI Depok, West Java,
16424, Indonesia
dewi.puspasari81@ui.ac.id

Adhiawan Soegiharto
Faculty of Computer Science,
Universitas Indonesia
Kampus UI Depok, West Java,
16424, Indonesia
adhiawan@gmail.com

Achmad Nizar Hidayanto
Faculty of Computer Science,
Universitas Indonesia
Kampus UI Depok, West Java,
16424, Indonesia
nizar@cs.ui.ac.id

Qorib Munajat
Management School,
Lancaster University,
Bailrigg, Lancaster
LA1 4YX, United Kingdom
q.munajat@lancaster.ac.uk

Abstract

This paper aims to conduct an analysis and exploration of matters that still needs to be considered in relation to data privacy in the online application system. This research is still a preliminary study. We conduct research related to data privacy using systematic literature review approach (SLR). By using SLR stages, we made a synthesis of 44 publications from Scopus Database Online that were released in the range 2015 - 2019. Based on this study, we found six points to consider in data privacy, namely security and data protection, user awareness, risk management, control setting, ethics, and transparency.

Keywords: Data privacy, systematic literature review, personal data, online application system, online system network

Introduction

Data at this time has been considered as a valuable and important asset (Baillie et al. 1994; Reinsel et al. 2018; Tapsell et al. 2018). This is because data can be a basis for strategic business decision making and can also provide insight in finding business opportunities (Reinsel et al. 2018). Data asset here includes data that is personal or data that can be associated or attached to someone (Lopes and Quaresma 2016; Tapsell et al. 2018).

Personal data is now more easily obtained by certain parties with the rise of social media and online application systems such as marketplaces, online transportation, and online loan services (Klukovich et al. 2016; Mostafa et al. 2017). Thus personal data is more prone to be misused (Schuppler et al. 2018; Shabtai et al. 2012). For this reason, there appears an emphasis on the term data privacy in which a person has the right to reject and close information attached to him (Korže and Čertanec 2017).

On the surface of the community it is as if deliberately sharing personal data because they think personal data is safe both on social media and on the online application system they use. In fact, this personal data is often misused by individuals who are traded to those who use it, among others, to examine consumer behavior, influence one's political direction, design political campaigns, to criminal acts such as credit card burglary and extortion (Shabtai et al. 2012).

The importance of protecting personal data is increasingly echoed after the scandal that befell Facebook with the sale of user data to Cambridge Analytica (Isaak and Hanna 2018; Srivastava and Geethakumari 2016). In Indonesia, data stored online is also widely misused, especially in the case of online loans and other cases relating to financial services (Majumdar et al. 2018). The case involving Cambridge Analytica and Facebook in 2018 did shake the world. The complete picture of the case is documented in full in the film "The Great Hack". In the documentary that the use of millions of users' personal data has been going on for years and only then revealed (Livemint 2019). Cases of misuse of personal data of its users apparently also involve Google and Twitter (Curran 2018; TechSpot 2019).

Meanwhile, awareness of personal data has begun to be intensified, including the presence of data privacy days or data protection days commemorated every January 28. The existence of this anniversary encourages awareness of the importance of data protection, including personal data, both by the institution and each individual (Vervier et al. 2017). Awareness of the importance of protecting personal data is widespread. Sharing personal data or vice versa, refusing to share data is privacy for everyone, including in the cyber world.

The important role of personal data is also recognized by the Indonesian government. Moreover, every year there are cases of misuse of personal data reported by the public. In 2019 there was a lot of news about Civil Registry Office's data leaks in the form of residence identification numbers and ID numbers. This data is said to be traded, some are used for extortion, although later the Civil Registry Office's dismissed the issue (Sekretariat Kabinet Republik Indonesia 2019). Based on Legal Aid, there are three thousand reports of data misuse by online lending institutions in Indonesia (Katadata 2019). Not to mention other cases of violations in other places and those that have not been reported. Personal data that are generally misused include telephone numbers, identity cards, ID numbers, and data on credit/banking cards (Hukum Online 2018).

Because we consider data protection to be important, therefore, in this study there is one thing that we want to explore with the explanatory method by taking a systematic literature approach. We ask one question, which is, "What are the things that still need to be considered in the activities of protecting personal data in an online application system, relation to data privacy?" The output of this research is any area that needs to be considered in relation to data privacy, especially with regard to personal data.

Literature Study

What is data privacy? Privacy according to Merriam-Webster is something whose use is intended to be limited to only certain people or groups. In relation to technological developments, the term and scope of privacy are also widespread. Now, privacy is not only about something physical and action, but also in the form of information or data. This privacy data has to do with privacy that is defined by Westin (1967), namely the demands of individuals, groups, or institutions to do and determine their own how, when, and to what extent information about them is communicated to other parties. Seeing from this definition, data privacy is related to access rights and control of information (Mai 2016).

General Data Protection Regulation (GDPR), the regulation of European Union law on data protection and privacy which is often a reference in data privacy and protection of personal data, provides a definition of data privacy as the freedom granted to make their own decisions about who can process data them and for what purpose (GDPR.EU 2019).

The definition of data privacy is generally associated with personal information that can characterize an individual (Mai 2016). The type of data that is of concern in terms of their use or relating to data privacy is personal data. Therefore, privacy data is closely related to personal data. Personal data is defined as all information that has a connection with identity or as natural can identify someone either directly or indirectly (Klosek 2000). This definition is similar to that stated in GDPR, personal data is any information relating to someone that can be identified directly or indirectly (European Parliament and of the Council 2016). Whereas in Indonesia, personal data based on the Population Administration Law is certain personal data that is stored, maintained, and kept truthful and protected by confidentiality (DPR 2006).

Personal data is based on the level of confidentiality and its importance is divided into four, namely insensitive data, sensitive data, quasi-identifiers, and explicit identifiers (Nataraj Venkataramanan 2016). Insensitive data is data that is easily accessed, for example gender. Sensitive data is data that has confidential information about the owner's records, for example health issues, financial status, and income. Whereas Quasi-identifiers are attributes that include demographic, geographical, telephone and e-mail address information. While explicit identifiers are attributes that are attached to someone directly. Examples are name, identity card, insurance ID, and social security number, driver license.

According to GDPR personal data includes names and email addresses, location information, ethnicity, gender, biometric data, religious beliefs, web cookies, and opinions. Pseudonym data can also be entered into personal data if it is relatively easy to identify someone from it (European Parliament and of the Council 2016). Whereas in Indonesia, personal data that is protected under the Population Administration Act includes family card numbers, employment numbers; ID number; date/month/year of birth; information on physical and/or mental disability; parent identification number; and some notes on important events (DPR 2006).

The discovery of the internet caused the issue of data privacy to be important. Especially with the internet access gadget. The internet is referred to as a source of information about individuals (Klosek 2000). This data is collected by the internet through surveys, cookies, pages that need to be registered and so on (Klosek 2000). At this time information is also being collected by various mobile applications embedded in the device. For this reason, protection of personal data is important. GDPR recommends that the personal data be processed legally, fairly and transparently, collected with the stated, explicit and legal purpose. Personal data may only be stored in the long term for archiving and public use, research purposes and statistical purposes (GDPR.EU 2019).

Research relating to personal data has been booming lately in a decade. This is because data privacy is something important related to technological progress. One paper discusses about user's awareness of their personal data in the online system (Hossain and Zhang 2015). They conducted the study by distributing questionnaires to 377 users who were familiar with online social networks (OSN) such as Facebook and Twitter. Based on this research 80 percent considered OSN had not provided adequate privacy controls (Hossain and Zhang 2015). Other studies discuss children's online privacy (Minkus et al. 2015). Some parents volunteered to share their children's data, even if for example the only viewing arrangement was arranged by a close friend. This is risky because crime against children is currently increasing, parents also need to protect their children's personal data (Minkus et al. 2015).

Research Methodology

In this study the author explores anything that still needs to be considered in relation to data privacy and protection of personal data, especially in relation to the online application system. To answer this problem, the authors conducted a series of research stages. This research methodology began with a systematic literature study of the latest research in the realm of computer science about personal data. Systematic literature review (SLR) is a research method that is widely used. This method is generally used in the fields of health and medicine, as well as in science. But later its use expanded, including in the field of computer science. The SLR method is suitable for preliminary research to find out about trends in a particular discipline, to clarify preliminary research, and to identify and interpret the state of the art on a topic (Kitchenham and Brereton 2013).

SLR can also be referred to as a form of secondary study by conducting a series of activities, beginning with the identification, analysis, and interpretation of all the evidence that has been obtained related to certain research questions that are not biased (Kitchenham et al. 2009). The stages can be repeated. This SLR method is suitable for researchers who want to know the current issues raised by researchers in a particular discipline or field, and aim to synthesize, avoid subjectivity, and bias (Kitchenham and Brereton 2013).

We chose to use the SLR approach to find current issues about data privacy. We hope to find a state of the art about research related to data privacy at this time. For this reason, we used the term 'data privacy' when searching using SLRs. We initially found 72,713 documents on Scopus electronic database. This

research began to climb in 2004 with 1.409 publications and continued to grow with 2.184 publications in 2007. Furthermore, research in this field had decreased before then returned to be interesting to study. The topic of privacy data again boomed in 2010 with 2.584 publications, then continued to climb until its peak reached 9.007 in 2019. The trends and figures on privacy data can be seen in the Figure 1.

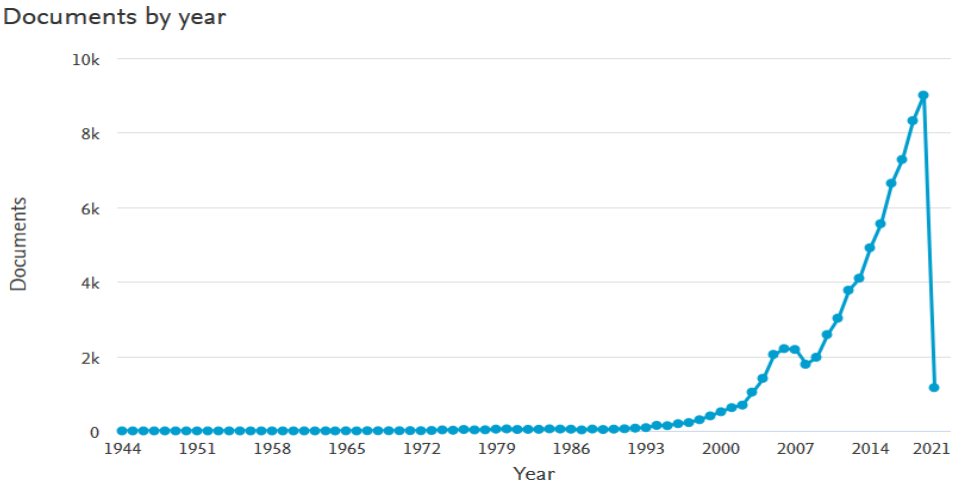

Figure 1. The Number of Publication on Data Privacy in Scopus Based By Year

We then carried out the screening process by providing several criteria and limits. We filtered based on the criteria for inclusion and exclusion as shown in the Table 1. We chose the latest publication, which ranges from 2015 to 2020. We also chose publications that fall into the field of computer science that written in English. We also limited the publication of results to a paper conference or journal. We also made restrictions so that the paper discusses more about the use of data privacy in the online system. From these limits we found 195 publications.

Table 1. Inclusion and Exclusion Criteria in SLR Process

Inclusion Criteria	Exclusion Critera
In the discipline area of Computer Science	Outside of Computer Science
Data privacy in online application system	Not relevant, for example only discuss the meaning of data privacy
2015 – 2020 range	Duplicated publications
Written in English language	Using any language
Type publication are papers, journals, or books review	Publications cannot be accessed by online

We then read the title of publication one by one. But apparently there were still some lecture notes that went into it. From the title screening, we found 141 publication, then we conducted abstract screening. We collected 96 suitable publications from this process. Next, we carried out final screening process by reading in full paper. We found 44 papers, then we made the summary paper one by one, we also did the synthesis and categorization. The stages in this SLR can be seen in the Figure 2.

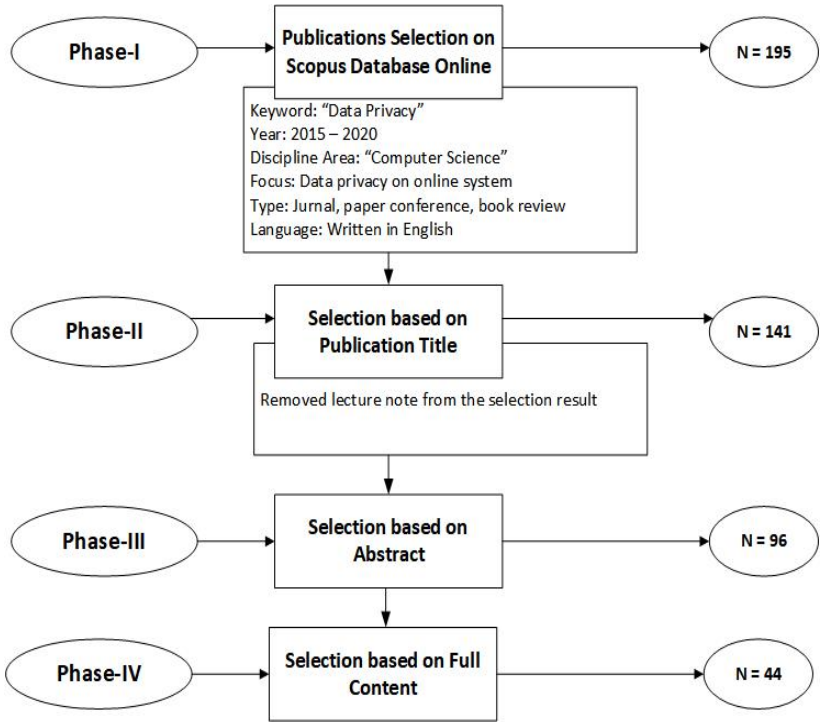

Figure 2. Systematic Literature Review Stages

Result

Of the 44 publications we obtained with the systematic literature review, most of them were conference papers, namely 43 papers and 1 in the form of journals. The average publication was issued in 2016 and 2017, with details: 2015 as many as 6 pieces (13.6%), 2016 consisting of 15 pieces (34.1%), 2017 as many as 16 pieces (36.4%), 2018 consists of 3 units (6.8%), and 2019 consists of 3 publications (6.8%). we do not find relevant publications for 2020.

The topics discussed in this publication about data privacy vary, but still have a common thread with the online system. Online technologies related to data privacy are mobile and web applications, applications that use cloud computing, online system networks, also big data. Part of this publication discusses the online system network (OSN) because OSN such as Facebook, Twitter, Instagram, and LinkedIn, have many loyal users using it.

Other topics aside from discussing online technologies are related to the topics of risk and risk mitigation; data security protection; access control policy; transparency of online service providers, user awareness, and the ethics of those who store and use personal data. One publication can discuss more than one issue, for example the level of user awareness regarding their personal data and their expectations of online service provider transparency. After the 44 publications are sorted into various categories above, we then synthesize them. The topic categories and summaries can be seen in Table 2.

A. Data Privacy Related to Online System Technology

Online system technology continues to grow, with technological advances and faster internet access. Currently various applications can also be accessed with mobile technology and big data technology is increasingly being used.

1) Online system network

Online system network (OSN) is used by millions of users from various countries. Users vary from teenagers to adults, not even children. It is fun to socialize with OSN, but there are risks lurking about the user's personal data (Hossain and Zhang 2015; Kulal and Dhamdhere 2017; Kumar et al. 2017; Luma et al. 2019; Masoumzadeh and Cortese 2017; Polakis et al. 2016; Revathi and Suriakala 2018; Srivastava and Geethakumari 2016; Symeonidis et al. 2016; Tambe and Vora 2017). Personal data of vulnerable

users is sold or used by third parties (Nalinipriya and Asswini 2016; Nandhini and Das 2016; Srivastava and Geethakumari 2016).

Their worries increase when there are cases of data misuse by several OSN (Shinjo et al. 2016; Umair et al. 2017). On the one hand this opens the awareness of users about the importance of reading the OSN privacy policy and regulating who can see their personal data (Albertini et al. 2017; Costa 2016; Hossain and Zhang 2015; Ilia et al. 2017; Klukovich et al. 2016; Minkus et al. 2015; Petkos et al. 2015; Tsirtsis et al. 2016; Van Der Valk et al. 2016). On the other hand they hope OSN is also law-abiding and transparent in the matter of the use of personal data of its users (Hossain and Zhang 2015; Mostafa et al. 2017; Polakis et al. 2016).

2) Mobile and web based application

At this time web-based applications can also generally be accessed by using a mobile device. These applications range from game applications, education, shopping for children's games, health, health insurance, and so on (Hölzl et al. 2016; Thao et al. 2018; Yee 2017a; Zhang et al. 2016, 2015).). Some applications also have features to share with OSN. With the increasing variety of applications embedded in the device, concerns have arisen that the user's data remains stored and then used as material to examine user habits, sell them to third parties, or be tapped to commit criminal acts (Aktypi et al. 2017; Alsalamah 2017; Hung et al. 2016; Leung et al. 2016; Yildirim and Varol 2019).

One of which was discussed is the Fitbit application, users of this application start to worry if the membership of a community increases (Zhang et al. 2015)). They are wary when conversations or historical data on the application are misused (Zhang et al. 2015). There is also research that discusses the importance of protecting one's medical record data when using health insurance applications because it is sensitive data (Zhang et al. 2016). Children's personal data are also prone to be misused when he interacts with smart toys, or online games (Hung et al. 2016).

3) Cloud computing

At present the use of cloud computing more and more with the convenience offered. There are many services that use cloud technology, such as e-voting (Grewal et al. 2015; Sedky and Hamed 2015). Of course this technology has risks because the stored data can be accessed by unauthorized parties (Mijuskovic and Ferati 2016).

4) Big data

At this time many companies are using big data technology for the benefit of companies, both those that are for the public interest, or for economic purposes. The use of personal data is actually something that must be in accordance with the guidelines, relating to the rules of data privacy (Vervier et al. 2017). Lately, there have been many cases of student data being used for educational data mining purposes, where the data is entered in sensitive data (Barril and Tan 2017). Another sensitive big data issue is research using patient data and medical records (Purandhar and Saravana Kumar 2019).

B. Data Privacy Related to Risk

With the increasing dependence of the community on the internet, intentionally or unintentionally their personal data is vulnerable exposed (Aktypi et al. 2017; Burbach et al. 2018; Hossain and Zhang 2015; Kumar et al. 2017; Pirzada et al. 2019; Purandhar and Saravana Kumar 2019; Symeonidis et al. 2016). The risk of exposure to this personal data from cases of buying and selling of their data, fraud, ID and password theft, social engineering attacks, SQL injection attacks, XSS attacks, fake friend profiles, recommendation systems, etc. (Alsalamah 2017; Hölzl et al. 2016; Hung et al. 2016; Leung et al. 2016; Luma et al. 2019; Malloy et al. 2017; Nalinipriya and Asswini 2016; Nandhini and Das 2016; Pirzada et al. 2019; Tsirtsis et al. 2016; Yildirim and Varol 2019). For this reason it is important to do mitigation to minimize the risk of private or sensitive data (Yee 2017a).

C. Data Privacy Related to Data Protection

Protection of personal data that is spread across the online system when it is important that it is not misused (Yee 2017b). This personal data can be in the form of names and inherent attributes including health records, web search records, location, conversation data, and sound cards. Now various data protection technologies are present. One data protection model is to blur the data. The process of blurring this data can be done by encryption(Grewal et al. 2015; Klukovich et al. 2016; Kulal and Dhamdhere 2017; Sedky and Hamed 2015), data masking (Degadwala and Gaur 2017) or by anonymization techniques (Srivastava and Geethakumari 2016; Thao et al. 2018; Zhang et al. 2016). In the process of obscuring data with anonymization techniques, not only anonymized data, nodes and attributes in an OSN graph also need to be anonymized (Srivastava and Geethakumari 2016). In masking techniques can also be done on the data attributes in the form of images (Degadwala and Gaur 2017).

Other data blurring techniques are data sanitization. This system uses the substitution method to clear keywords. Because nouns and verbs provide the most information in a sentence, they will be treated as keywords and the rest of the words will be treated as function words. Keywords will be sanitized using Stanford natural language processing (Tambe and Vora 2017).

Other data protection proposals by using decentralized social networking services use virtual private networks so that data is controlled and does not leave a group (Shinjo et al. 2016). It is also proposed to conduct periodic Fraud Assessment and Detection, for example by regularly checking and verifying fake links and fake friend profiles (Nandhini and Das 2016; Tsirtsis et al. 2016), measuring user exposure through periodic privacy exposure metrics (Masoumzadeh and Cortese 2017), using a data security algorithm for securing personal data pribadi (Pirzada et al. 2019), or using the Logic Rule Generation algorithm to be able to find and analyze the nature of user vulnerabilities (Revathi and Suriakala 2018).

D. Data Privacy Related to Access Control Regulation and Setting Control

Data privacy is related to controls. Users may refuse or grant access to their personal data. For this reason, each OSN must provide control arrangements for who can see the user's personal data and what data can be seen(Albertini et al. 2017; Hossain and Zhang 2015; Klukovich et al. 2016). When this feature is available, users feel safer when sharing information (Van Der Valk et al. 2016). A study proposes a collaborative multi-party access control model that allows all users associated with these resources to participate in access control policy specifications (Ilia et al. 2017).

E. Data Privacy Related to Transparency

Regarding data privacy, some OSN users consider OSN to be not transparent in their data usage policies (Hossain and Zhang 2015). This mistrust is triggered by the many cases of data breaches by selling user data to third parties and various cases that threaten users. For this reason, users expect that there are terms and conditions that mention data privacy and if data is used by the OSN provider (Mostafa et al. 2017).

F. Data Privacy Related to Ethics

The use of personal data under the GDPR is permitted if intended for public purposes and for statistical purposes required by the region or country. However, there are various other provisions relating to research ethics, such as maintaining data confidentiality, respecting privacy, not selling or sharing it with other parties, and so on (Polakis et al. 2016).

G. Data Privacy Related to User Awareness

With many cases of violations in the use of user data, users are increasingly aware of the threatening risks when they give too much personal data to the public. Most users begin to be aware and aware of the attributes of their data that enter sensitive data (Hossain and Zhang 2015; Mijuskovic and Ferati 2016; Umair et al. 2017; Zhang et al. 2015). However, sometimes parents forget that they are neglectful of the privacy of their children's data(Minkus et al. 2015; Tsirtsis et al. 2016). Although for example the visibility has been set for only close people but the child's data remains something risky (Minkus et al. 2015). Some are actually aware of privacy issues, but they then voluntarily share them with certain rewards (Vervier et al. 2017). Because user awareness is important, a study proposes a framework for

measuring privacy awareness in three dimensions, namely visibility, level of control, and privacy score (Petkos et al. 2015).

Table 2. Summary of Each Category

Topic	Summary	Key Point	Publications
Online System Technology			
• Online System Network (OSN)	OSN is used by millions of users with various threats to the user's personal data. Moreover, lately there have been many cases of misuse of user data by several OSNs. On the one hand this opens the awareness of users about the importance of reading OSN's privacy policy and regulates who can see their personal data, as well as the expectations of OSN to be transparent in the use of user data.	User awareness, data protection, risk, control setting, transparency	(Albertini et al. 2017; Hossain and Zhang 2015; Ilia et al. 2017; Klukovich et al. 2016; Kulal and Dhamdhere 2017; Kumar et al. 2017; Luma et al. 2019; Masoumzadeh and Cortese 2017; Minkus et al. 2015; Mostafa et al. 2017; Nalinipriya and Asswini 2016; Nandhini and Das 2016; Petkos et al. 2015; Polakis et al. 2016; Revathi and Suriakala 2018; Shinjo et al. 2016; Srivastava and Geethakumari 2016; Symeonidis et al. 2016; Tambe and Vora 2017; Tsirtsis et al. 2016; Umair et al. 2017; Van Der Valk et al. 2016)
• Mobile and Web Based Application	With the increasing variety of applications, including those embedded in the device, there is a concern that the user's data remains stored and then used as material to examine user habits, sell them to third parties, or be tapped to commit criminal acts.	Risk, data misuse, data protection	(Aktypi et al. 2017; Alsalamah 2017; Hölzl et al. 2016; Hung et al. 2016; Leung et al. 2016; Thao et al. 2018; Yee 2017a; Yildirim and Varol 2019; Zhang et al. 2016, 2015)
• Cloud Computing	This cloud technology has risks because stored data can be accessed by unauthorized parties.	Data security and protection	(Grewal et al. 2015; Mijuskovic and Ferati 2016; Sedky and Hamed 2015)
• Big Data	Utilization of personal data must be in accordance with the guidelines, relating to the data privacy rules. Lately there have been many cases of student data being used for educational data mining purposes, where the data is entered in sensitive data.	Ethics	(Barril and Tan 2017; Purandhar and Saravana Kumar 2019; Vervier et al. 2017)
Risk	The risk of exposure to personal data varies, from cases of buying and selling of their data, fraud, ID and password theft, social engineering attacks, SQL injection attacks, XSS attacks, fake friend profiles, recommendation systems, etc.	Risk, mitigation	(Aktypi et al. 2017; Alsalamah 2017; Burbach et al. 2018; Hölzl et al. 2016; Hossain and Zhang 2015; Hung et al. 2016; Kumar et al. 2017; Leung et al. 2016; Luma et al. 2019; Malloy et al. 2017; Nalinipriya and Asswini 2016; Nandhini and Das 2016; Pirzada et al. 2019; Purandhar and Saravana Kumar 2019; Symeonidis et al. 2016; Tsirtsis et al. 2016; Yee 2017a; Yildirim and Varol 2019)
Data Protection	Data protection can be done by obscuring data such as encryption techniques, data masking, data	Data blurring, fraud	(Degadwala and Gaur 2017; Grewal et al. 2015; Klukovich et al. 2016; Kulal and Dhamdhere

Topic	Summary	Key Point	Publications
	sanitization, or anonymization techniques. Other methods are using fraud assessment and detection, also by privacy exposure metrics.	detection, security	2017; Masoumzadeh and Cortese 2017; Nandhini and Das 2016; Pirzada et al. 2019; Revathi and Suriakala 2018; Sedky and Hamed 2015; Shinjo et al. 2016; Srivastava and Geethakumari 2016; Tambe and Vora 2017; Thao et al. 2018; Tsirtsis et al. 2016; Yee 2017b; Zhang et al. 2016)
Access Control Regulation and Control Setting	Data privacy is related to controls, users may refuse or grant access to their personal data. For this reason, each OSN must provide control settings for who can see the user's personal data and what data can be seen.	Access control, control settings	(Albertini et al. 2017; Hossain and Zhang 2015; Ilia et al. 2017; Klukovich et al. 2016; Van Der Valk et al. 2016)
Transparency	Users expect that there are terms and conditions that state if data is used by the OSN provider.	Transparency	(Hossain and Zhang 2015; Mostafa et al. 2017)
Ethics	The use of personal data is permitted if it is intended for public purposes and for statistical purposes required by the region or country. Related to research ethics, such as maintaining data confidentiality, respecting privacy, not selling or sharing it with other parties, and so on.	Ethics	(Polakis et al. 2016)
User Awareness	With many cases of violations in the use of user data, users are increasingly aware of the threatening risks when they give too much personal data to the public. There are also those who are aware of privacy issues, but they then voluntarily share them with certain rewards.	User awareness	(Hossain and Zhang 2015; Mijuskovic and Ferati 2016; Minkus et al. 2015; Petkos et al. 2015; Tsirtsis et al. 2016; Umair et al. 2017; Vervier et al. 2017; Zhang et al. 2015)

Based on the information in Table 2, there are several key points that are often reviewed in each topic. Key points that are widely reviewed are about risk and mitigation, security and data protection, user awareness, control settings and access control, transparency, fraud detection, and ethics. Because there are several terms that are similar and can be combined, we propose six key points that still need to be considered in maintaining data privacy when using online applications. The key points are security and data protection, user awareness, control settings, risk management, transparency, and ethics as in Figure 3. Fraud assessment and detection can be a part of risk management and security and data protection. While, access control of data is included in the control setting.

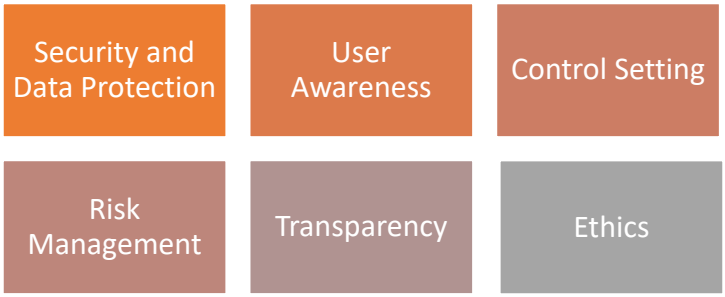

Figure 3. Key Points in Data Privacy

Conclusion and Future Work

Until now there are still many cases of misuse of personal data in the online system. Several cases were revealed that increased user awareness of the importance of protecting personal data. The also demanded service providers to respect privacy data.

Based on research using a systematic review, we found 44 publications (2014-2019) that discussed data privacy. After we categorized and synthesized them, we found six key points that must be considered when using an online application system related to data privacy. These six points are security and data protection, user awareness, control settings, risk management, transparency, and ethics.

This research is still a preliminary study, so there are still many things that can be explored based on this research. Next, we will conduct a gap analysis between key points in data privacy and personal data protection regulations in Indonesia.

References

Aktypi, A., Nurse, J. R. C., and Goldsmith, M. 2017. "Unwinding Ariadne's Identity Thread: Privacy Risks with Fitness Trackers and Online Social Networks," *MPS 2017 - Proceedings of the 2017 Workshop on Multimedia Privacy and Security, Co-Located with CCS 2017* (2017–Janua), pp. 1–11. (https://doi.org/10.1145/3137616.3137617).

Albertini, D. A., Carminati, B., and Ferrari, E. 2017. "Privacy Settings Recommender for Online Social Network," *Proceedings - 2016 IEEE 2nd International Conference on Collaboration and Internet Computing, IEEE CIC 2016*, IEEE, pp. 514–521. (https://doi.org/10.1109/CIC.2016.079).

Alsalamah, A. 2017. "Security Risk Management in Online System," *Proceedings - 2017 5th International Conference on Applied Computing and Information Technology, 2017 4th International Conference on Computational Science/Intelligence and Applied Informatics and 2017 1st International Conference on Big Data, Cloud Compu*, IEEE, pp. 119–124. (https://doi.org/10.1109/ACIT-CSII-BCD.2017.59).

Baillie, C. F., Hawick, K. A., and Johnston, D. A. 1994. "Quenching 2D Quantum Gravity," *Physics Letters B* (Vol. 328). (https://doi.org/10.1016/0370-2693(94)91481-8).

Barril, J. F. H., and Tan, Q. 2017. "Integrating Privacy in Architecture Design of Student Information System for Big Data Analytics," *2017 2nd IEEE International Conference on Cloud Computing and Big Data Analysis, ICCCBDA 2017* (1), IEEE, pp. 139–144. (https://doi.org/10.1109/ICCCBDA.2017.7951899).

Burbach, L., Nakayama, J., Plettenberg, N., Ziele, M., and Valdez, A. C. 2018. "User Preferences in Recommendation Algorithms," *RecSys 2018 - 12th ACM Conference on Recommender Systems*, pp. 306–310. (https://doi.org/10.1145/3240323.3240393).

Costa, L. 2016. *Data Protection Law, Processes and Freedoms BT - Virtuality and Capabilities in a*

World of Ambient Intelligence: New Challenges to Privacy and Data Protection. (https://doi.org/10.1007/978-3-319-39198-4_6).

Curran, D. 2018. "Are You Ready? This Is All the Data Facebook and Google Have on You | Dylan Curran | Opinion | The Guardian," *The Guardian*, pp. 1–12. (https://www.theguardian.com/commentisfree/2018/mar/28/all-the-data-facebook-google-has-on-you-privacy, accessed March 11, 2020).

Degadwala, S. D., and Gaur, S. 2017. "An Efficient Privacy Preserving System Using VCS, Block DWT-SVD and Modified Zernike Moment on RST Attacks," *2017 International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies, ICAMMAET 2017* (2017–Janua), pp. 1–4. (https://doi.org/10.1109/ICAMMAET.2017.8186685).

DPR. 2006. "Laws of The Republic Indonesia Number 23 of 2006 Regarding of Population Administration."

European Parliament and of the Council. 2016. "Art. 4 GDPR – Definitions | General Data Protection Regulation (GDPR)," *European Parliament and of the Council*. (https://gdpr-info.eu/art-4-gdpr/, accessed March 12, 2020).

GDPR.EU. 2019. "A Guide to GDPR Data Privacy Requirements," *A Guide to GDPR Data Privacy Requirements*. (https://gdpr.eu/data-privacy/, accessed March 12, 2020).

Grewal, G. S., Ryan, M. D., Chen, L., and Clarkson, M. R. 2015. "Du-Vote: Remote Electronic Voting with Untrusted Computers," *Proceedings of the Computer Security Foundations Workshop* (2015–Septe), IEEE, pp. 155–169. (https://doi.org/10.1109/CSF.2015.18).

Hölzl, M., Roland, M., and Mayrhofer, R. 2016. "Real-World Identification: Towards a Privacy-Aware Mobile EID for Physical and Offline Verification," *ACM International Conference Proceeding Series*, pp. 280–283. (https://doi.org/10.1145/3007120.3007158).

Hossain, A. A., and Zhang, W. 2015. "Privacy and Security Concern of Online Social Networks from User Perspective," *ICISSP 2015 - 1st International Conference on Information Systems Security and Privacy, Proceedings*, SCITEPRESS, pp. 246–253. (https://doi.org/10.5220/0005318202460253).

Hukum Online. 2018. "Yuk Simak, Perlindungan Data Pribadi Yang Tersebar Di Beberapa UU." (https://www.hukumonline.com/berita/baca/lt5aa2522899af7/yuk-simak--perlindungan-data-pribadi-yang-tersebar-di-beberapa-uu/, accessed March 11, 2020).

Hung, P. C. K., Fantinato, M., and Rafferty, L. 2016. "A Study of Privacy Requirements for Smart Toys," *Pacific Asia Conference on Information Systems, PACIS 2016 - Proceedings* (June).

Ilia, P., Carminati, B., and Ferrari, E. 2017. *SAMPAC : Socially-Aware Collaborative Multi-Party Access Control*, pp. 71–82.

Isaak, J., and Hanna, M. J. 2018. "User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection," *Computer* (51:8), IEEE, pp. 56–59. (https://doi.org/10.1109/MC.2018.3191268).

Katadata. 2019. "LBH Terima 3.000 Aduan Pelanggaran Fintech Pembiayaan." (https://katadata.co.id/berita/2019/02/04/lbh-terima-3000-aduan-pelanggaran-fintech-pembiayaan, accessed March 11, 2020).

Kitchenham, B., and Brereton, P. 2013. "A Systematic Review of Systematic Review Process Research in Software Engineering," *Information and Software Technology* (55:12), Elsevier B.V., pp. 2049–2075. (https://doi.org/10.1016/j.infsof.2013.07.010).

Kitchenham, B., Pearl Brereton, O., Budgen, D., Turner, M., Bailey, J., and Linkman, S. 2009. "Systematic Literature Reviews in Software Engineering - A Systematic Literature Review," *Information and Software Technology* (51:1), Elsevier B.V., pp. 7–15.

(https://doi.org/10.1016/j.infsof.2008.09.009).

Klosek, J. 2000. "Data Privacy in the Information Age," *Quorum Books*, p. 251. (https://books.google.com.my/books?id=18HSi5ekRbcC&printsec=frontcover&dq=data+privacy&hl=en&sa=X&ved=0ahUKEwjJqu2pxvfeAhUMiXAKHc1PBGYQ6AEIKjAA#v=onepage&q&f=false, accessed March 12, 2020).

Klukovich, E., Erdin, E., and Gunes, M. H. 2016. "POSN: A Privacy Preserving Decentralized Social Network App for Mobile Devices," *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2016*, IEEE, pp. 1426–1429. (https://doi.org/10.1109/ASONAM.2016.7752436).

Korže, B., and Čertanec, A. 2017. "Protecting Personal Data in the Context of Interoperability among Organizations for Protection and Rescue," *International Data Privacy Law* (7:4). (https://doi.org/10.1093/idpl/ipx017).

Kulal, N., and Dhamdhere, V. 2017. "Technique for Preserving Privacy on Friend Recommendation System by Using Naive Bayes Classifier in OSN," *Proceedings of the 2017 International Conference on Intelligent Computing and Control Systems, ICICCS 2017* (2018–Janua), pp. 315–319. (https://doi.org/10.1109/ICCONS.2017.8250734).

Kumar, H., Jain, S., and Srivastava, R. 2017. "Risk Analysis of Online Social Networks," *Proceeding - IEEE International Conference on Computing, Communication and Automation, ICCCA 2016*, IEEE, pp. 846–851. (https://doi.org/10.1109/CCAA.2016.7813833).

Leung, C., Ren, J., Choffnes, D., and Wilson, C. 2016. "Should You Use the App for That? Comparing the Privacy Implications of App-and Web-Based Online Services," *Proceedings of the ACM SIGCOMM Internet Measurement Conference, IMC* (14–16–Nove), pp. 365–372. (https://doi.org/10.1145/2987443.2987456).

Livemint. 2019. "What The Great Hack Tells Us about Data Privacy." (https://www.livemint.com/mint-lounge/features/what-the-great-hack-tells-us-about-data-privacy-1565946607143.html, accessed March 11, 2020).

Lopes, S., and Quaresma, R. 2016. "Data Privacy in Interoperability Environments -A Case Study in the Portuguese Healthcare Sector," *16th Portuguese Association for Information Systems Conference, CAPSI 2016* (16), pp. 43–54. (https://doi.org/10.1109/INFCOM.2011.5934930).

Luma, A., Abazi, B., and Aliu, A. 2019. "An Approach to Privacy on Recommended Systems," *3rd International Symposium on Multidisciplinary Studies and Innovative Technologies, ISMSIT 2019 - Proceedings*, IEEE, pp. 1–5. (https://doi.org/10.1109/ISMSIT.2019.8932805).

Mai, J.-E. 2016. *The Information Society An International Journal Big Data Privacy: The Datafication of Personal Information*. (https://doi.org/10.1080/01972243.2016.1153010).

Majumdar, S., Tripathy, S., Majumdar, S., Tripathy, S., Abubakar, L., and Handayani, T. 2018. "IOP Conference Series: Earth and Environmental Science Financial Technology: Legal Challenges for Indonesia Financial Sector Related Content Data Protection in Financial Technology Services: Indonesian Legal Perspective Dian Purnama Anugerah and Masitoh Indriani-Service Sector Performance: A Critical Review Performance Evaluation of Indian Education Sector Using Interpretive Structural Modelling Financial Technology: Legal Challenges for Indonesia Financial Sector," *IOP Conf. Ser.: Earth Environ. Sci. 175 IOP Conf. Series: Earth and Environmental Science* (175), p. 12204. (https://doi.org/10.1088/1755-1315/175/1/012204).

Malloy, M., Barford, P., Alp, E. C., Koller, J., and Jewell, A. 2017. "Internet Device Graphs," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Part F1296), pp. 1913–1921. (https://doi.org/10.1145/3097983.3098114).

Masoumzadeh, A., and Cortese, A. 2017. "Towards Measuring Knowledge Exposure in Online Social Networks," *Proceedings - 2016 IEEE 2nd International Conference on Collaboration and*

Internet Computing, IEEE CIC 2016 (Section VI), IEEE, pp. 522–529. (https://doi.org/10.1109/CIC.2016.080).

Mijuskovic, A., and Ferati, M. 2016. "Cloud Storage Privacy and Security User Awareness: A Comparative Analysis between Dutch and Macedonian Users," *International Journal of Human Capital and Information Technology Professionals* (7:3), pp. 1–18. (https://doi.org/10.4018/IJHCITP.2016070101).

Minkus, T., Liu, K., and Ross, K. W. 2015. "Children Seen but Not Heard: When Parents Compromise Children's Online Privacy," *WWW 2015 - Proceedings of the 24th International Conference on World Wide Web*, pp. 776–786. (https://doi.org/10.1145/2736277.2741124).

Mostafa, S. A. M., Noori, S. R. H., and Jafreen, S. 2017. "Transparency-A Key Feature Integration in Existing Privacy Frameworks for Online User," *IWCI 2016 - 2016 International Workshop on Computational Intelligence* (December), IEEE, pp. 74–78. (https://doi.org/10.1109/IWCI.2016.7860342).

Nalinipriya, G., and Asswini, M. 2016. "A Survey on Vulnerable Attacks in Online Social Networks," *Proceedings 2015 - IEEE International Conference on Innovation, Information in Computing Technologies, ICIICT 2015*, IEEE, pp. 1–6. (https://doi.org/10.1109/ICIICT.2015.7396102).

Nandhini, M., and Das, B. B. 2016. "An Assessment and Methodology for Fraud Detection in Online Social Network," *2016 2nd International Conference on Science Technology Engineering and Management, ICONSTEM 2016*, IEEE, pp. 104–108. (https://doi.org/10.1109/ICONSTEM.2016.7560932).

Nataraj Venkataramanan, A. S. 2016. "Data Privacy: Principles and Practice," *Chapman and Hall/CRC*, p. 232. (https://books.google.co.id/books?id=lpWKDQAAQBAJ&printsec=frontcover&dq=%27data+privacy%27&hl=id&sa=X&ved=0ahUKEwjv3v-LmpToAhU2yTgGHW89CUAQ6AEIbzAH#v=onepage&q='data privacy'&f=false, accessed March 12, 2020).

Petkos, G., Papadopoulos, S., and Kompatsiaris, Y. 2015. "PScore: A Framework for Enhancing Privacy Awareness in Online Social Networks," *Proceedings - 10th International Conference on Availability, Reliability and Security, ARES 2015*, IEEE, pp. 592–600. (https://doi.org/10.1109/ARES.2015.80).

Pirzada, S. J. H., Murtaza, A., Liu, J., and Xu, T. 2019. "The Parallel CMAC Authentication Algorithm," *2019 IEEE 11th International Conference on Communication Software and Networks, ICCSN 2019*, IEEE, pp. 800–804. (https://doi.org/10.1109/ICCSN.2019.8905326).

Polakis, I., Maggi, F., Zanero, S., and Keromytis, A. D. 2016. "Security and Privacy Measurements in Social Networks: Experiences and Lessons Learned," *Proceedings - 3rd International Workshop on Building Analysis Datasets and Gathering Experience Returns for Security, BADGERS 2014*, IEEE, pp. 18–29. (https://doi.org/10.1109/BADGERS.2014.9).

Purandhar, N., and Saravana Kumar, N. M. 2019. "Review of Data Extraction, Segregation Privacy with Big Data Analytics in the Online Health Care Systems," *Proceedings of the International Conference on Intelligent Sustainable Systems, ICISS 2019* (Iciss), IEEE, pp. 193–197. (https://doi.org/10.1109/ISS1.2019.8907973).

Reinsel, D., Gantz, J., and Rydning, J. 2018. *The Digitization of the World From Edge to Core*, (November). (https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf).

Revathi, S., and Suriakala, M. 2018. "An Intelligent and Novel Algorithm for Securing Vulnerable Users of Online Social Network," *Proceedings of the 2nd International Conference on Computing Methodologies and Communication, ICCMC 2018* (Iccmc), IEEE, pp. 214–219.

(https://doi.org/10.1109/ICCMC.2018.8487760).

Schuppler, C., Leitner, M., and Rinderle-Ma, S. 2018. "Privacy-Aware Data Assessment of Online Social Network Registration Processes," *CODASPY 2018 - Proceedings of the 8th ACM Conference on Data and Application Security and Privacy* (2018–Janua), pp. 167–169. (https://doi.org/10.1145/3176258.3176950).

Sedky, M. H., and Hamed, E. M. R. 2015. "A Secure E-Government's e-Voting System," *Proceedings of the 2015 Science and Information Conference, SAI 2015*, IEEE, pp. 1365–1373. (https://doi.org/10.1109/SAI.2015.7237320).

Sekretariat Kabinet Republik Indonesia. 2019. "No Leaks on Personal Data Database: Home Ministry." (https://setkab.go.id/en/no-leaks-on-personal-data-database-home-ministry/, accessed March 11, 2020).

Shabtai, A., Elovici, Y., and Rokach, L. 2012. "Data Leakage Detection/Prevention Solutions," *SpringerBriefs in Computer Science* (9781461420521), pp. 17–37. (https://doi.org/10.1007/978-1-4614-2053-8_4).

Shinjo, Y., Kainuma, N., Nobori, D., and Sato, A. 2016. "Magic Mantle Using Social VPNs against Centralized Social Networking Services," *2016 14th Annual Conference on Privacy, Security and Trust, PST 2016*, IEEE, pp. 341–348. (https://doi.org/10.1109/PST.2016.7906984).

Srivastava, A., and Geethakumari, G. 2016. "Determining Privacy Utility Trade-off for Online Social Network Data Publishing," *12th IEEE International Conference Electronics, Energy, Environment, Communication, Computer, Control: (E3-C3), INDICON 2015*, IEEE, pp. 1–6. (https://doi.org/10.1109/INDICON.2015.7443693).

Symeonidis, I., Tsormpatzoudi, P., and Preneel, B. 2016. "Collateral Damage of Online Social Network Applications," *ICISSP 2016 - Proceedings of the 2nd International Conference on Information Systems Security and Privacy* (June), pp. 536–541. (https://doi.org/10.5220/0005806705360541).

Tambe, P., and Vora, D. 2017. "Data Sanitization for Privacy Preservation on Social Network," *International Conference on Automatic Control and Dynamic Optimization Techniques, ICACDOT 2016*, IEEE, pp. 972–976. (https://doi.org/10.1109/ICACDOT.2016.7877732).

Tapsell, J., Akram, R. N., and Markantonakis, K. 2018. "Consumer Centric Data Control, Tracking and Transparency - A Position Paper," *Proceedings - 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications and 12th IEEE International Conference on Big Data Science and Engineering, Trustcom/BigDataSE 2018*, IEEE, pp. 1380–1385. (https://doi.org/10.1109/TrustCom/BigDataSE.2018.00191).

TechSpot. 2019. "Twitter 'accidentally' Misused User Data to Sell Targeted Ads." (https://www.techspot.com/news/82267-twitter-accidentally-misused-user-data-sell-targeted-ads.html, accessed March 11, 2020).

Thao, T. P., Makanju, A., and Kubota, A. 2018. "Anonymous and Analysable Web Browsing," *2017 IEEE 36th International Performance Computing and Communications Conference, IPCCC 2017* (2018–Janua), pp. 1–8. (https://doi.org/10.1109/PCCC.2017.8280466).

Tsirtsis, A., Tsapatsoulis, N., Stamatelatos, M., Papadamou, K., and Sirivianos, M. 2016. "Cyber Security Risks for Minors: A Taxonomy and a Software Architecture," *Proceedings - 11th International Workshop on Semantic and Social Media Adaptation and Personalization, SMAP 2016*, IEEE, pp. 93–99. (https://doi.org/10.1109/SMAP.2016.7753391).

Umair, A., Nanda, P., and He, X. 2017. "Online Social Network Information Forensics: A Survey on Use of Various Tools and Determining How Cautious Facebook Users Are?," *Proceedings - 16th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, 11th IEEE International Conference on Big Data Science and Engineering and*

14th IEEE International Conference on Embedded Software and Systems, pp. 1139–1144. (https://doi.org/10.1109/Trustcom/BigDataSE/ICESS.2017.364).

Van Der Valk, R. V. R., Helms, R. W., Van De Wetering, R., Bex, F. J., and Corten, R. 2016. "Feeling Safe? Privacy Controls and Online Disclosure Behavior," *24th European Conference on Information Systems, ECIS 2016* (June).

Vervier, L., Zeissig, E. M., Lidynia, C., and Ziefle, M. 2017. "Perceptions of Digital Footprints and the Value of Privacy," *IoTBDS 2017 - Proceedings of the 2nd International Conference on Internet of Things, Big Data and Security* (IoTBDS), pp. 80–91. (https://doi.org/10.5220/0006301000800091).

Yee, G. O. M. 2017a. "Model for Reducing Risks to Private or Sensitive Data," *Proceedings - 2017 IEEE/ACM 9th International Workshop on Modelling in Software Engineering, MiSE 2017*, IEEE, pp. 19–25. (https://doi.org/10.1109/MiSE.2017.6).

Yee, G. O. M. 2017b. "Adding Privacy Protection to Distributed Software Systems," *ICETE 2017 - Proceedings of the 14th International Joint Conference on e-Business and Telecommunications* (4:Icete), pp. 351–358. (https://doi.org/10.5220/0006434903510358).

Yildirim, N., and Varol, A. 2019. "A Research on Security Vulnerabilities in Online and Mobile Banking Systems," *7th International Symposium on Digital Forensics and Security, ISDFS 2019*, IEEE, pp. 1–5. (https://doi.org/10.1109/ISDFS.2019.8757495).

Zhang, A., Bacchus, A., and Lin, X. 2016. "A Fairness-Aware and Privacy-Preserving Online Insurance Application System," *2016 IEEE Global Communications Conference, GLOBECOM 2016 - Proceedings*, IEEE, pp. 1–6. (https://doi.org/10.1109/GLOCOM.2016.7841495).

Zhang, J., Dibia, V., Sodnomov, A., and Lowry, P. B. 2015. "Understanding the Disclosure of Private Healthcare Information within Online Quantified Self 2.0 Platforms," *Pacific Asia Conference on Information Systems, PACIS 2015 - Proceedings*.