The Defining Approaches and Practical Paradox of Sensitive Data: An Investigation of Data Protection Laws in 92 Countries and Regions and 200 Data Breaches in the World

MIN WANG[1,2]
Wuhan University, China

ZUOSU JIANG
Central China Normal University, China

Results of juristic investigations into 92 countries and regions demonstrate that sensitive data are viewed as a core area of both privacy and data protection. The latest data protection laws in most countries and regions have defined sensitive data. A comparative study of all of these definitions or classifications revealed two defining approaches: "EU standard" and "EU standard plus criminal records." A practical examination of the 200 biggest data breaches worldwide that have occurred since 2004 showed categories of personal data, particularly financial data, although not defined as sensitive in most countries, have been violated most. This paradox can be interpreted by universal principles of laws and ethics, such as human dignity and sacredness of life. Finally, we suggest a comprehensive understanding of data privacy as both a personality and a quasiproperty right and provide recommendations for defining sensitive data in legislation.

Keywords: privacy, sensitive data, data protection, universal principle, human dignity

Privacy is and "will continue to be a crucial social value" (Trepte & Reinecke, 2011, p. 7), which depends largely on "the contemporary technological developments" (Soffer & Cohen, 2015, p. 145). Technological convenience, however, has come "at the expense of our digital privacy right" (Sisk, 2016, p.

Min Wang: carterwm@qq.com
Zuosu Jiang: 19530921@qq.com
Date submitted: 2017–01–13

[1] Some of the data and analysis has previously appeared in the Chinese language publication *Xinwenjie.* Wang, M. (2017). 敏感数据的定义模型与现实悖论：基于 92 个国家隐私相关法规以及 200 个数据泄露案例的分析[The defining models and practical paradox of sensitive data: An analysis of privacy-related laws in 92 countries and 200 data breaches]. *Xinwenjie*, *6*, 2–10.

[2] We thank the three anonymous reviewers for their helpful comments on a draft of this work, and the editors for their kindness and patience throughout the whole process. This study is supported by the National Social Science Foundation of China (Grant 17CXW027) and the Major Program of Key Research Center of Humanities and Social Sciences, Ministry of Education of China (Grant 16JJD860004).

101). In particular, the rise of social media and the advances in ICTs have made it virtually impossible for anyone "to be let alone" (Freivogel, 2015, p. 303). The impact of the digital age is so deep, pervasive, and universal that a single or regional privacy and data protection law, or the applications of regional concepts or principles, is unlikely to adequately address the paradoxes between privacy and security, efficiency, innovation, transparency, and free speech (Baker, 2013; Friedewald & Pohoryles, 2014; Hiranandani, 2011; Nissenbaum, 2009; Solove, 2004; Volokh, 2000). Etzioni (2015) calls for "an evolution" in both the conceptual and the legal status of privacy (p. 1267). Moreover, it is increasingly necessary to develop universal principles to balance individual privacy with personal, commercial, public, and global interests in the context of frequent cross-border data flow, and to establish a safe and integrated digital market for the sake of globalized data-driven business, just as the "Digital Single Market" of the European Commission.

In the latest EU General Data Protection Regulation (EU Regulation for short), the right to be forgotten is probably the most controversial provision, which would allow individual Internet users to demand removal of their personal data by websites, search engines, and data controllers (Reding, 2012). The most complex part of the removal requests is to determine "what constitutes a valid request for removal" (Myers, 2014, p. 56). The guidelines released in November 2014 by the Working Party, an independent European advisory body on privacy and data protection, provide common criteria to evaluate a removal request, and the most important one is whether the data about the person are "sensitive" (Article 29 Data Protection Working Party, 2014).

Some kinds of data are more sensitive than others (Etzioni, 2015; Pesciotta, 2012; Schwartz & Solove, 2011). This sensitivity differs in two aspects: At the basic level, sensitive data are separate from nonpersonal data, such as "pseudonymized data" and "anonymized data" (Zuiderveen, Van, & Gray, 2015, p. 2077); the second level is that it distinguishes from other personal data that are deemed less private or special, such as "basic or factual data" (Taddicken, 2014, p. 270). Al-Fedaghi (2007) points out that the sensitivity of personal data is one of the most important factors in determining an individual's perception of privacy, and the gradation of sensitivity could decide the security level that one controls access to such data. On the contrary, for data subjects, a failure to reasonably secure sensitive personal data would "lead to more potential exposure" (Harris, 2016, p. iv). The loss of sensitive data is a significant concern for individuals whose data may be "at risk of a breach" (Photopoulos, 2011, p. 3). This reveals a relation between sensitive data and data breaches. Ojanen (2014) argues that sensitive data are a "core area" (p. 534) of both privacy and data protection, and thus deserve special or stricter protection in legislation. This might be the reason why most of the countries and regions in the world investigated by us have defined sensitive data in their data protection laws, although the definitions sometimes differ.

Therefore, there are a few questions to be investigated: Which categories of personal data are the most commonly recognized as sensitive data in theoretical legislation among the world's countries and regions? What are the most frequently violated categories of personal data in practice that need special protection? What can account for differences, if any, between theoretical definitions and practical violations? To address these research questions, we next present a theoretical framework and literature review. In the third section, we describe the method and process about how the data were collected, compared, and analyzed. The fourth section explores the findings following the empirical work. In the fifth section, we discuss our research findings, explain the differences between theoretical definitions and practical violations, and offer a new defining framework of sensitive data. In the final section, we present the implications of our study for research and practice, limitations of our work, and directions for future research.

Theoretical Framework and Literature Review

This study employed a theoretical framework of "sensitive data versus universal principle." In this section, we review the literature concerning the concept and classification of sensitive data, coupled with the universal principle of human dignity.

The Concept

The concept that some kinds of personal data are more sensitive than others often has been articulated by privacy scholars and operationalized by lawmakers using a variety of terms (Etzioni, 2015). In 1993, Westin used the Medical Sensitivity Index to summarize the results of his privacy-related surveys and to show trends in medical privacy concerns (Kumaraguru & Cranor, 2005), being one of the earliest proponents of individual differences in privacy sensitivity (Hurwitz, 2013). Other terms, such as "intimate information" (Nissenbaum, 1998, p. 559) and "highly intrusive data" (Pesciotta, 2012, p. 237), have also been applied to indicate sensitive data. Some scholars have defined sensitive data in terms of "the level of risk to one's privacy or the extent of harm to one's privacy" (Cate, Cullen, & Mayer-Schonberger, 2013, p. 11), and others have measured it by "the degree of intrusiveness upon personal privacy" (Pesciotta, 2012, p. 238). Although literature differs on the definition of sensitive data, there seems to be a basic consensus: All nonpersonal data are inherently not sensitive, and that "not all personal data are of the same gradation of sensitivity" (Al-Fedaghi, 2007, pp. 165–166).

At the legislative level, the Organization for Economic Cooperation and Development (OECD) Guidelines first introduced the concept of sensitive data, but failed to "achieve consensus on which categories of data deserve special protection" (McCullagh, 2007, p. 190). Thereafter, sensitive data were defined or categorized in international law through the Council of Europe Convention in 1981, the UN Guidelines in 1990, and the EU Data Protection Directive 95/46/EC (EU Directive for short) in 1995 (see Table 1). There has been an evolution in the concept of sensitive data and a broadening of the scope to include more types of data. The current legal framework, the EU Directive, prohibits "the processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life" (European Parliament and the Council of the European Union, 1995, Article 8). This is the "EU standard" of sensitive data, containing seven categories of personal data. The standard excludes criminal conviction, which is listed in the Council of Europe Convention, although it is now specially protected by the right to be forgotten (Rosen, 2012). The major differences between the UN Guidelines and EU standard lie in color and health data. In the latest EU Regulation, the categories of genetic data and biometric data have been added to the EU standard.

Table 1. Categories of Sensitive Data in International Legislation.

Legislation	Definition of sensitive data	Legislation	Definition of sensitive data
OECD Guidelines (1980)	None	EU Directive (European Parliament and the Council of the European Union, 1995)	• Racial or ethnic origin • Political opinions • Religious beliefs • Philosophical beliefs • Sex life • Health data • Trade-union membership
Council of Europe (1981)	• Racial origin • Political opinions • Religious beliefs • Criminal convictions • Sexual life • Health data		
UN Guidelines (UN General Assembly, 1990)	• Racial or ethnic origin • Political opinions • Religious beliefs • Philosophical beliefs • Sex life • Color • Membership in an association or trade union	EU Regulation (European Parliament and the Council of the European Union, 2016)	• Racial or ethnic origin • Political opinions • Religious beliefs • Philosophical beliefs • Sex life/sexual orientation • Health data • Trade-union membership • Genetic data • Biometric data

The Classification

To identify and classify such a core area is far more complex than what has been found by previous studies. As early as 1972, Bing (1972) attempted to categorize all personal data according to their sensitivity. However, the approach was quickly abandoned because of vagueness of boundaries (McCullagh, 2007). According to McCullagh (2007), there are two approaches to classify the categories of sensitive data: context-based and purposed-based. Indeed, some scholars believe that sensitivity of personal data varies from context to context (Fule & Roddick, 2004). Even the OECD Guidelines adopt a contextual approach and do not classify special categories of sensitive data (OECD, 1980). Nevertheless, Wacks (1989) argues that what changes from context to context is not the degree of data sensitivity, but the extent of an individual's attitude toward its use, and that the nature of the information does not change. Other scholars simply "do not exclude the possibility of 'context-free' sensitivity" (Al-Fedaghi & Al-Azmi, 2012, p. 123). Sariyar, Schluender, Smee, and Suhr (2015) raise several types of data sensitivity, such as legal, ethical, social, and contextual sensitivities.

In the present research, we chiefly discuss legal sensitivity or the sensitive data defined in legislation, not to consider the context, or "the values of the society" (Etzioni, 2015, p. 1278) or "normative culture" (Etzioni & Rice, 2015, p. 8). According to the latest DLA Piper (2017), most countries and regions in the world have defined or classified sensitive data in their data protection laws. The main principle under which these definitions are addressed is nondiscrimination and protection of human dignity

(Commission Nationale de l'Informatique et des Libertés, 2013). For example, the UN Guidelines state, "data likely to give rise to unlawful or arbitrary discrimination, including information on racial or ethnic origin, colour, sex life, political opinions . . . should not be compiled" (UN General Assembly, 1990, p. 2). McCullagh (2007) also mentions "freedom of political activity" (p. 193) in the United Kingdom's existing categories of sensitive data. This freedom can basically be viewed as a part of human dignity.

Universal Principle of Human Dignity

The universal ethical principle of human dignity has substantial presence in global legislation, especially in Europe (Neal, 2014). Mamberti (2012) even argues that the ultimate and essential goal of all law is "to promote and to guarantee the dignity of the human person" (p. 1). The first sentence in The Universal Declaration of Human Rights is a "recognition of the inherent dignity and of the equal and inalienable rights of all members of the human family" (United Nations, 1948). Analogously, the EU Charter of Fundamental Rights also declares in the prominent place that "human dignity is inviolable" (European Commission, 2009). In the European Convention on Human Rights, there is a "full recognition of the inherent dignity of all human beings" (Council of Europe, 2010, p. 52). Just as Christians (2010) points out, as an ethical protonorm,[3] human dignity has been translated into a legal framework with practical universal application. Relating human dignity and privacy, Bloustein (1964) believes privacy is an aspect of human dignity. Today, privacy is an important factor that determines autonomy and free will, and is increasingly linked to the issues of human dignity (Commission Nationale de l'Informatique et des Libertés, 2013). This point might explain why the EU standard pays special attention to personal data concerning individual dignity, such as religious beliefs, political opinions, racial/ethnical identity, and philosophical beliefs.

Other universal principles that might be connected with the definition and classification of sensitive data are also explored in the following sections. Such an exploration is based on an investigation and a comparison of the worldwide data protection laws. Scholars such as Greenleaf (2014, 2015) and organizations such as DLA Piper (2017) have completed extensive empirical work in this field. However, there has been little research done to compare sensitive data. The present research investigates and compares the definitions of sensitive data in 92 countries and regions to identify the most commonly recognized and newly emerging categories and to conceptualize these categories and their defining approaches. Moreover, given that a failure to secure sensitive data would lead to more breaches (Harris, 2016), the loss of sensitive data is a concern of being at risk of a breach (Photopoulos, 2011). Whereas Malheiros, Preibusch, and Sasse (2013) point out the relationship between sensitive data and data breaches, this research also makes a comparative analysis of categorizations of sensitive data and breaches of data. This comparison tries to examine whether the categories or list of sensitive data should be upgraded or a new defining approach should be adopted. Overall, these two comparisons aim to provide an evolving understanding of sensitive data that needs continuous evaluation.

[3] The term *protonorm* was coined by Clifford G. Christians. According to Christians (2015), *proto* in Greek means *underneath*, and the notion of protonorm is "a way of rooting our universals in ontology rather than in the rationalist propositions of the Western tradition" (p. 342).

Method and Process

We adopted a three-step comparative method: (1) Collect and compare the definitions of sensitive data in privacy-related laws of the world to identify the most commonly recognized categories in legislation, (2) collect and analyze the biggest data breaches of the world to identify the most commonly violated categories of personal data in practice, and (3) compare the categories of sensitive data in legislation and data of breaches in practice to examine the differences between the results of the first two steps. More specifically, the process can be divided into three steps.

The First Step

We studied the definitions of sensitive data in 92 countries and regions through DLA Piper's (2017) latest handbook, *Data Protection Laws of the World*. By May 2017, this book had set out an overview of the key privacy and data protection laws and regulations across 92 jurisdictions, which cover not only the major countries in every continent but also most of the countries in the world. According to Banisar (2016), "over 100 countries and independent jurisdictions and territories" (p. 1) around the world had adopted comprehensive data/privacy laws to protect personal data by the end of November 2016. Therefore, the research sample is representative.

First, we reviewed the overviews of each country and region, examined whether there was a definition of sensitive data, and figured out how many countries and regions have defined sensitive data. Second, we listed all the categories of personal data in each country or region's definition of sensitive data, and calculated the frequency of each category recognized by all countries and regions with definitions. Based on the results, we analyzed the defining approaches.

Because the definitions of each country and region presented by *Data Protection Laws of the World* are clear and distinct, we used those definitions to categorize the types of sensitive data and calculated their frequency. In this process, we addressed variants and ambiguity:

- Some have defined "sexual orientation" as a type of data, whereas others have classified it as "sexual preference." Although there are some differences, we put both of them into the category "sex life."
- According to the Article 29 Data Protection Working Party (2011), "moral belief," similar to "philosophical belief," is different from "religious belief" and "moral or emotional characteristic." These were separated into different categories.
- Personal data concerning physical or mental health ("health data" for short), genetics, and biometrics are independent in most legislative definitions.
- Financial data, including bank account, credit or debit card, or other payment instrument details, are categorized as "income data."

The Second Step

Next, we collected and examined the 200 biggest data breaches on record throughout the world from April 2004 to November 2016 to see which categories of personal data were leaked the most or more likely to be violated. We chose to explore the data breaches from 2004 chiefly because (1) by 2004, the Internet was in wide use after 10 years of development, and the global legislation on data protection (especially the major countries such as EU member states, Australia, Canada, Japan, and the United States) was almost complete, and (2) given the widespread usage of e-mail and blogs since 2004, the problem of data and privacy violation has become more serious.

First, we collected the data breaches from the annual news reports concerning the "10 biggest data breaches of the year" released by the mainstream media outlets (including *The New York Times*, CNN, BBC, *The Guardian*, Reuters, ITRC website, etc.), and summarized the 200 biggest cases of breaches in total from April 2004 to November 2016 according to the size of the data leak. Specifically, we summarized one major category of lost data from each breach case, figured out its total number of leaked data points, and sorted out the 200 biggest data breaches. Second, we calculated the total lost number of each data category in the 200 breaches, sorted each category by the leak size, and found out which categories of personal data were violated most.

There were usually several categories of data leaked in one case. We mainly chose one principal category either according to each one's number of losses or in light of the organization or sector involved. For example, data leaked from hospitals and health departments largely belong to health data. We generally summarized six primary categories of personal data from the breaches to maintain coding consistency, namely, (1) e-mail addresses/online accounts, (2) social security numbers/personal details, (3) debit/credit card information, (4) e-mail passwords, (5) health records, and (6) full bank account details. Both credit card information and full bank account details are part of the financial data category.

The Third Step

Based on the results of the first two steps, we compared the most commonly defined sensitive data and the most commonly violated personal data, and investigated the differences between the two parts. Because both parts need special protection by laws, the results of comparison could test whether the current legislative categories of sensitive data are reasonable or need to be extended. Universal principles proposed by Western scholars, such as protonorms and human dignity, were applied to analyze and explain the results and differences.

Results

The first major finding relates to definitions and classifications of personal data. After examining the latest privacy-related laws in 92 countries and regions (DLA Piper, 2017), we found that 80.43% of the countries and regions (n = 74) defined sensitive information/data or classified special categories of personal data, whereas the remaining 19.57% (n = 18) of countries and regions did not have legal definitions of sensitive information/data or an equivalent (see Table 2).

Table 2. Countries and Regions With and Without Definitions of Sensitive Information/Data.

With definitions (*n* = 74)	Without definitions (*n* = 18)
Angola, Argentina, Australia, Austria, Azerbaijan, Belgium, Bosnia and Herzegovina, Brazil, Bulgaria, Cape Verde, Chile, Mainland China, Colombia, Costa Rica, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Ghana, Gibraltar, Greece, Guernsey, Honduras, Hungary, Iceland, India, Ireland, Israel, Italy, Japan, Jersey, Latvia, Lesotho, Lithuania, Luxembourg, Macau, Macedonia, Madagascar, Malaysia, Malta, Mauritius, Mexico, Monaco, Montenegro, Morocco, Netherlands, Nigeria, Norway, Peru, Philippines, Poland, Portugal, Romania, Russia, Qatar, Seychelles, Slovak Republic, South Africa, South Korea, Spain, Sweden, Switzerland, Taiwan, Trinidad and Tobago, Turkey, UAE—Dubai (DIFC), Ukraine, United Kingdom, United States, Uruguay	Bahrain, Belarus, British Virgin Islands, Canada, Cayman Islands, Egypt, Indonesia, Hong Kong, New Zealand, Saudi Arabia, Serbia, Singapore, Thailand, Pakistan, Panama, UAE—General, Venezuela, Zimbabwe

Second, we found that of the 74 countries and regions with defined sensitive information/data or classified special categories of personal data, there were 33 categories of personal data that have been defined as sensitive.

Third, on closer examination of these 33 data categories, the majority of countries and regions (>50%) recognized the following eight categories as sensitive: (1) physical or mental health, (2) religious beliefs or affiliations, (3) political opinions or membership, (4) sexual life, (5) race or ethnicity, (6) trade union, (7) philosophical or moral beliefs, and (8) criminal records or proceedings or administrative proceedings. Another 25 categories of data about genetics, biometrics, financial information, social welfare, personality, ID numbers, children's information, and so on, also appeared in some countries' sensitive list (see Table 3).

Table 3. Categories of Personal Data Defined as Sensitive by 74 Countries and Regions.

Category	Countries and regions defining data category as sensitive (*n*)	Frequency (%)
Physical or mental health	72	97.30
Religious beliefs or affiliations	71	95.95
Political (ideological) opinions or membership	69	93.24
Sexual life, orientation, preference, or practices	68	91.89
Racial/ethnic origin	66	89.19
Trade/labor union membership	60	81.08
Philosophical or moral beliefs	54	72.97
Criminal records or proceedings, or administrative proceedings	40	54.05
Genetic information	23	31.08
Biometrics	18	24.32
Marital status or family matters, private life	9	12.16

Government numbers, licenses, social welfare	5	6.76
Personality, moral or emotional characteristics, personal habits	5	6.76
Financial/income/accounts, debit/credit cards	4	5.41
Identification number	4	5.41
Personal information of a child	4	5.41
Tax arrears or returns	2	4.05
Passwords	2	2.70
Phone number	2	2.70
Other (e.g., abnormal addiction, age, child adoption, credit worthiness, domestic violence, notarial acts, education, home address, registered domicile, personal and familiar heritage, personal electronic address, professional/trade association, social status, student data, etc.)	14	18.92

The category of data concerning criminal records and administrative proceedings, which ranked eighth in the sensitive data list, was the major divergence among EU member states or even all of the 74 countries and regions with sensitive data definitions. In the European Union, 17 member states (60.71%) classified these data as special and sensitive, and the other 11 states (39.29%) did not recognize the category; in the 74 countries and regions, 50% defined it as sensitive (see Figure 1).

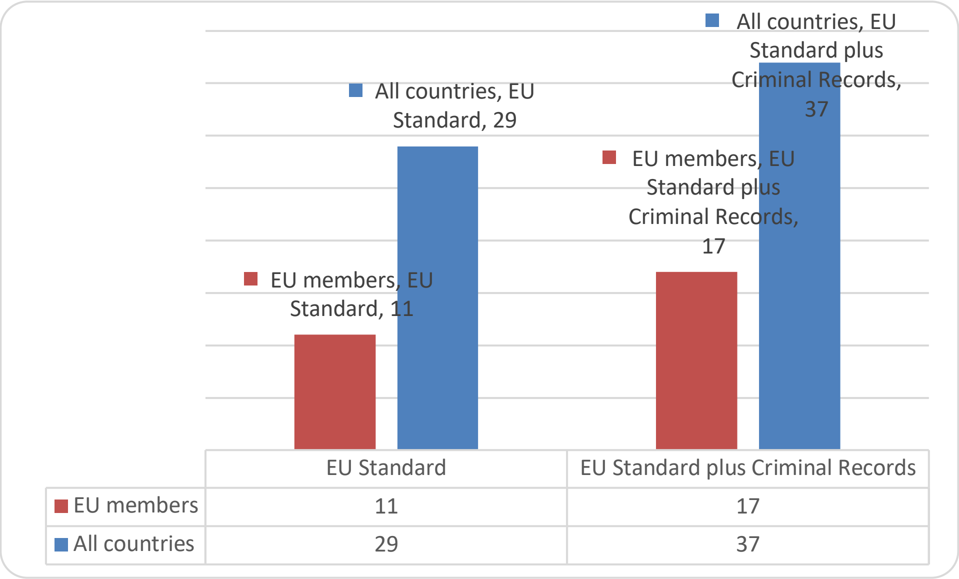

Figure 1. Countries and regions of EU standard and EU standard plus criminal records.

A final major finding was identified after collecting and analyzing the 200 biggest data breaches (loss or thefts) from 2004 to 2016 throughout the world. The main data stolen or lost were classified into six categories: (1) e-mail addresses/online accounts, (2) ID/social security numbers/personal details, (3) debit/credit card information, (4) passwords, (5) health records, and (6) full bank account details. The total number of breach cases involved with each of the six categories was 51, 77, 34, 6, 16, and 16, respectively (see Table 4). The accumulated number of lost records regarding these six categories was 601,504,476, 869,307,903, 572,455,544, 98,841,170, 28,054,947, and 393,961,000, respectively. The total records of lost financial data, including debit/credit card information and full bank account details, amounted to 966,416,544, accounting for 38.1% of the total loss and ranking first among all the categories. Data concerning ID/social security numbers/personal details took up 34.3% of the total loss and ranked second. Lost health records occupied only 0.1%, being the least leaked category (see Figure 2).

Table 4. Number of Breach Cases and Number of Leaked Records for Each Category of Personal Data.

	Breach cases		Leaked records	
Category	*n*	%	*n*	%
E-mail addresses/online accounts	51	25.5	601,504,476	23.7
ID/social security numbers/personal details	77	38.5	869,307,903	34.3
Debit/credit card information	34	17.0	572,455,544	22.6
Passwords	6	3.0	98,841,170	3.9
Health records	16	8.0	28, 054,947	0.1
Full bank account details[a]	16	8.0	393,961,000	15.5
Total	200	100.0	2,536,070,093	100.1

[a]Full bank account details include the account number, the password, card verification value numbers, and billing addresses. There were mainly 16 breach cases, with the number of lost data concerning full bank account details ranging from 40,000 to 160,000,000. The total number of records leaked is the accumulated summation of the number in all cases, which is approximately 393,961,000.

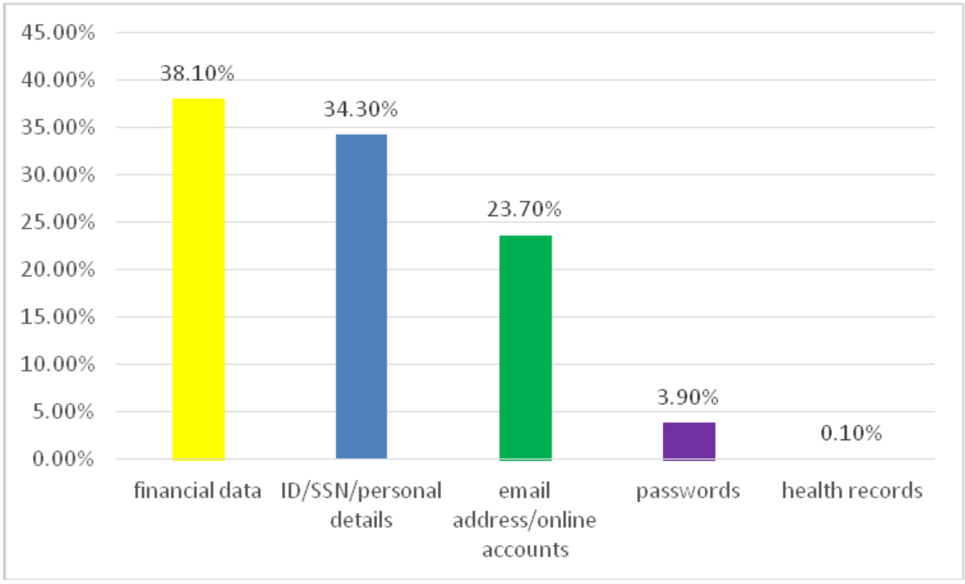

Figure 2. The percentage of lost records for each category of personal data.

Discussion

This might be the first large-scale comparative research on the definitions of sensitive data among different countries and regions. With all research questions investigated, the research outlines a list of special categories of personal data and two approaches of defining sensitive data, provides an overview of lost personal data from the biggest data breaches worldwide, and explains the practical paradox between defined sensitive data and breached personal data by adopting common protonorms. In addition, this research makes other recommendations for upgrading the definition of sensitive data based on juristic investigations and practical observations.

Defining Approaches and the Practical Paradox

In theory, the results of juristic observations demonstrate that the majority of the countries and regions around the world have defined sensitive data through two approaches: the EU standard and the EU standard plus criminal records. Moreover, the EU definition of sensitive data is so prevalent that most countries and regions have classified their sensitive data or special categories of personal data based on the EU standard. The major divergence lies in data concerning criminal records, which is defined as sensitive by more than half of the countries and regions investigated. This defining approach can be summarized as "EU standard plus criminal records," which is adopted by the majority of countries and regions (see Figure 1). These two approaches contribute to a comprehensive and overall understanding of the connotation and denotation of sensitive data, establishing a theoretical foundation for the clarification and development of the definition of sensitive data. The EU Directive also provides exemptions in cases of "(a) national security; (b) defense; (c) public security; (d) the prevention, investigation, detection, and prosecution of criminal offences; (e) monetary, budgetary, and taxation matters" (European Parliament and the Council of the European Union, 1995, Article 13, para. 1). Thus, there could be some restrictions in the actual implementation and enforcement of data protection, regardless of what kind of personal data are involved or how sensitive the data are.

Second, health data, although most widely defined or classified in theory as sensitive, are less likely to be stolen, whereas financial data, although not among the top-10 recognized sensitive data, are the most likely to be violated in practice. This is the practical paradox of sensitive data.

What can account for this paradox and explain the differences between theoretical definitions and practical data breaches? We can analyze it from two perspectives.

Theoretically, the study argues that the universal protonorms, such as the sacredness of life and human dignity, are the very essence and theoretical basis for the EU standard. One rationale behind the seven special categories of data in the EU standard stems from the presumption that the "misuse or leakage of these data could result in more severe impact or consequences on the data subject than that of other, 'normal' personal data" (Article 29 Data Protection Working Party, 2011, p. 4). China's guideline of data protection also defines sensitive data as personal data "the leakage or alteration of which may result in adverse impact to the data subject" (DLA Piper, 2017, p. 100). Misuse or leakage of sensitive data, such as health data or sexual practices, may have a long-term and irreversible impact on the individual's life health, dignity, reputation, or property. Thus, universal principles of the sacredness of life and human dignity justify the special protection of such personal data. According to Christians (2010), these protonorms ground a responsibility that is global in scope and self-evident regardless of cultures and ideologies. This argument can account for the prevalent recognition of health data as sensitive data. Another rationale in the EU standard is nondiscrimination, which is closely connected with human dignity, as previously discussed.

Practically, personal data contain tremendous financial value and belong to individual property (Murphy, 1995). For example, the Singapore Commission has stated that "personal data is of a sensitive financial nature in its enforcement" (DLA Piper, 2017, p. 416), although currently there is no definition of sensitive data in Singapore. In the age of big data when data are basic units for almost every kind of application, the result of data loss and breaches is always for commercial use or even fraudulent transaction. According to the Consumer Sentinel Network, a Federal Trade Commission database of consumer complaints, identity theft was the number-one concern expressed in 2014, and thieves tried to use stolen personal data to commit credit card, phone, bank, and loan frauds in 47.3% of the identity theft cases in the United States (U.S. Federal Trade Commission, 2015). Identity fraudsters bilked $16 billion from 12.7 million U.S. consumers in 2014. This was actually an improvement over the year before, when fraudsters stole $18 billion from 13.1 million victims (Holmes, 2015). E-commerce, e-marketing, data trade, and fraudulent transactions largely result from personal data thefts and loss. From this perspective, the fact that financial records and information are at the highest risk of being violated and stolen can be explained by the property nature and potential financial value of personal data.

In sum, the practical paradox lies in looking at personal data from two separate perspectives. The EU standard of sensitive data is theoretically defined on the basis of privacy as a human right or personality right, and most data breaches are induced practically by personal data as a property. Seen from the EU standard, the growing property attribute of personal data in the present digital world has been greatly ignored in legislation. This ignorance can be dated to the era before the EU Directive became an influential force in 1990s. At that time, computers and the Internet were not widely used. People communicated and conducted business in traditional ways, such as face to face or by cable telephones. With the rapid developments especially in ICTs, new communicating ways and business models that fundamentally influence individuals' personal lives and business processes have emerged (Tene, 2011). At present, the economic value of personal data has been greatly highlighted. Paying less attention to the property attribute of privacy, particularly personal data, will lead to more significant economic and property loss.

Other Recommendations for Defining Sensitive Data

Investigations into various privacy-related laws show that technological developments have produced new kinds of data, such as genetic and biometric data, which should also be considered sensitive. In the 1990s, the public rarely knew about the biometric and genetic data (Jain, Ross, & Nandakumar, 2011). However, biometric technology has become so pervasive that 23 countries investigated define genetic data as sensitive and 18 countries provide special protection to biometric data. Although the EU standard does not include either genetic or biometric data, such newly generated personal data should also be better protected because they are "closely related with personal identity and commercial activity" (Jain, Bolle, & Pankanti, 2006, p. ix). Being unique and permanent, biometric data once lost, stolen, copied, or forged will lead to "irreversible identity theft" and "financial loss" (Campisi, 2013, p. v). Moreover, biometric data often act as access to a subject's identity, thereby involving other personal data, such as personality and health information or bank accounts (Kindt, 2013, p. 89). As a consequence, the latest EU Regulation prohibits "the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person" (European Parliament and the Council of the European Union, 2016, p. 38). Besides biometric and genetic data, new forms of data as well as data processing should be provided with specific safeguards to avoid privacy intrusion and economic damage.

Furthermore, certain categories of sensitive data, such as philosophical or moral beliefs, health data, criminal conviction, private life, and social status, are somewhat obscure or hard to identify in practice. For example, Simonton (1976) examined and listed 16 philosophical beliefs developed by thinkers in history, including empiricism, materialism, idealism, singularism, universalism, ethics of happiness, ethics of love, and so forth, and a court in the United Kingdom recognized belief in climate change as a philosophical belief (Article 29 Data Protection Working Party, 2011). The concept of philosophical belief seems very broad and obscure. Other important categories are also broad. For instance, health data can include genetic and biometric data, and genetic data can also be part of biometrics. As for criminal conviction, some scholars argue that there should be an explanation about whether the conviction is the first time or a second time. If these terms are not clearly or legally defined, the specific legal protection of sensitive data can be difficult to enforce and implement in practical cases and real life.

In addition, advances in ICTs not only produce new categories of data but also bring changes in generational expectations. We found that four countries classify the personal data of a child as sensitive because of special protection for youth. Actually, the newer generation, which has grown up in the past decade, has deeply integrated the ICTs into their daily lives (Tene, 2011). This means that the prevalent ICTs have changed, and the users of these ICTs have also changed to be "digital natives" (Palfrey & Gasser, 2013, p. 1). These digital natives may have different perceptions of privacy compared with individuals using ICTs in 1990s. The seven categories of sensitive data in accordance with the OECD Guidelines and the EU Directive might not be that special, private, or intimate to them. As a matter of fact, expectations may vary because the generations could have different expectations about the general definition of privacy and sensitive data in particular (Ashraf, 2015). The current legislative framework also needs to be upgraded and improved to protect privacy in accordance with the personal data that the new generation values.

Conclusion

By examining the data protection laws in 92 countries and regions, we found that the majority of countries and regions in the world have defined or classified sensitive data in their data protection laws. There are typically two defining approaches by these countries and regions, namely, the EU standard and the EU standard plus criminal records. Both frameworks tend to view sensitive data more as a personality right and neglect its increasing property attribute. An investigation of the 200 biggest data breaches worldwide indicates that this negligence might most frequently result in financial loss and economic damage given that financial data top all of the six data categories that have been stolen in the breaches. Although recognized as sensitive in few data protection laws, financial data have been most widely violated in daily life. This is the practical paradox: What needs special protection in practice is not granted special protection in legislation. A weakness of the present research is that we identify sensitive data mainly according to the theoretical definitions in legislation, without referring to the context in which such data are used.

Cases and reports show that the leakage of personal data has been frequently involved with commercial activities or fraudulent transactions. To ensure a safe and sustainable digital market, we suggest other categories of personal data. In particular, because it has been breached so often, financial data should also be included as sensitive. An evolving understanding of sensitive data might need to address their property attributes under the universal principles of human dignity and life sacredness by taking financial data as both a personality and a quasiproperty right. Just as Conroy (2012) argues, "The personality and property aspects of the privacy right are inextricably intertwined" (p. 19). However, the present research did not have enough evidence to determine which categories of personal data can represent a property of the individuals and which can be included as sensitive. This is a limitation of the research.

Juristic investigations also show that the terms and concepts used in the definitions of sensitive data are either obscure or too broad to identify. Moreover, the privacy-related legislation in general and data protection in particular should be modified to keep up with the advances in ICTs and changes in the perception of data subjects. On April 27, 2016, the European Union modified such laws and adopted the

EU Regulation. Compared with the EU Directive, the regulation has made key changes to strengthen the online privacy rights and data protection in the digital age. An example change is the recognition of new statutory rights such as the right to be forgotten. However, there is still no specific progress on the definition or classification of sensitive data, except for recognition of biometric and genetic data. A failure to distinguish some of these similar data categories is also a weakness of our research.

Meanwhile, the worldwide problem of security is prominent. In a risk society, there should be some exemptions in cases of common good and public interest. One could ask which defining approach is more reasonable in such a context: the EU standard or the EU standard plus criminal records. In other words, to what degree or under what circumstance should the data concerning criminal records be viewed as sensitive? This can be a key topic of future research. In addition, just as not every category of personal data should be protected in the same way or to the same degree, different individuals and data subjects can also be classified in the legislation and enforcement of data protection. For instance, personal information of a child (under parental control or 13 years old) is viewed as sensitive data in Ghana, Lesotho, South Africa, and the United States. Privacy of special individuals and data subjects, such as children, the disabled, or even the deceased, is not addressed, a limitation that could be explored through future research.

References

Al-Fedaghi, S. (2007). How sensitive is your personal information? In *Proceedings of the 2007 ACM Symposium on Applied Computing* (pp. 165–169). New York, NY: Association for Computing Machinery. doi:10.1145/1244002.1244046

Al-Fedaghi, S., & Al-Azmi, A. A. R. (2012). Experimentation with personal identifiable information. *Intelligent Information Management*, *4*(4), 123–133. doi:10.4236/iim.2012.44019

Article 29 Data Protection Working Party. (2011, April 20). Advice paper on special categories of data ("sensitive data"). Retrieved from http://ec.europa.eu/justice/data-protection/article-29/documentation/other-document/files/2011/2011_04_20_letter_artwp_mme_le_bail_directive_9546ec_annex1_en.pdf

Article 29 Data Protection Working Party. (2014, November 28). Guidelines on the Implementation of the Court of Justice of the European Union Judgment on "Google Spain and Inc. v. Agencia Española de Protectión de Datos (AEPD and Mario Costeja González" C-131/12. Retrieved from http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-zecommendation/files/2014/wp225_en.pdf

Ashraf, S. (2015). Generational expectations: Does the global generation of social media users view privacy differently than the generation before them? In W. Babcock & W. Freivogel (Eds.), *The SAGE guide to key issues in mass media ethics and law* (pp. 419–436). Thousand Oaks, CA: SAGE Publications.

Baker, S. A. (2013). *Skating on stilts: Why we aren't stopping tomorrow's terrorism*. Stanford, CA: Hoover Press.

Banisar, D. (2016, November 28). National comprehensive data protection/privacy laws and bills 2016 map. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1951416

Bing, J. (1972). Classification of personal information, with respect to the sensitivity aspect. In *Proceedings of the First International Oslo Symposium on Data Banks and Societies* (pp. 98–150). Oslo, Norway: Universitetsforlaget.

Bloustein, E. J. (1964). Privacy as an aspect of human dignity: An answer to Dean Preosser. *New York University Law Review*, *39*(6), 962–1007.

Campisi, P. (2013). *Security and privacy in biometrics*. London, UK: Springer.

Cate, F. H., Cullen, P., & Mayer-Schonberger, V. (2013). *Data protection principles for the 21st century: Revising the 1980 OECD Guidelines*. Redmond, WA: Microsoft Corporation.

Christians, C. G. (2010). The ethics of universal being. In S. J. Ward & H. Wasserman (Eds.), *Media ethics beyond borders: A global perspective* (pp. 10–17). London, UK: Routledge.

Christians, C. G. (2015). The ethics of dignity in a multicultural world. In B. Shan & C. Christians (Eds.), *The ethics of intercultural communication* (pp. 337–355). New York, NY: Peter Lang.

Commission Nationale de l'Informatique et des Libertés. (2013). "Privacy towards 2020"—42 experts share their visions of the future of privacy with the French regulation authority. Retrieved from https://www.cnil.fr/sites/default/files/typo/document/CAHIER_IP_EN.pdf

Conroy, A. M. (2012). Protecting your personality rights in Canada: A matter of property or privacy? *Western Journal of Legal Studies, 1*(1), 1–22.

Council of Europe. (1981, January 28). Convention for the protection of individuals with regard to automatic processing of personal data, Chapter 2, Article 6. Retrieved from https://rm.coe.int/1680078b37

Council of Europe. (2010, June 1). Convention for the protection of human rights and fundamental freedoms. Retrieved from http://www.echr.coe.int/Documents/Convention_ENG.pdf

DLA Piper. (2017, May 15). *Data protection laws of the world handbook*. Retrieved from https://www.dlapiperdataprotection.com/system/modules/za.co.heliosdesign.dla.lotw.data_protection/functions/handbook.pdf?country=all

Etzioni, A. (2015). A cyber age privacy doctrine: More coherent, less subjective, and operational. *Brooklyn Law Review, 80*(4), 1263–1308.

Etzioni, A., & Rice, C. J. (2015). *Privacy in a cyber age: Policy and practice*. London, UK: Palgrave Macmillan.

European Commission. (2009, December). EU Charter of Fundamental Rights. Retrieved from http://ec.europa.eu/justice/fundamental-rights/charter/index_en.htm

European Parliament and the Council of the European Union. (1995, October 24). Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. *Official Journal of European Union*. Retrieved from http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML

European Parliament and the Council of the European Union. (2016, April 27). General data protection regulation, Article 9, para. 1. Retrieved from http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf

Freivogel, W. H. (2015). The right to be let alone. In W. A. Babcock & W. H. Freivogel (Eds.), *The SAGE guide to key issues in mass media ethics and law* (pp. 303–318). Thousand Oaks, CA: SAGE Publications. doi:10.4135/9781483346540.n27

Friedewald, M., & Pohoryles, R. J. (Eds.). (2014). *Privacy and security in the digital age*. London, UK: Routledge.

Fule, P., & Roddick, J. F. (2004, January). Detecting privacy and ethical sensitivity in data mining results. In *Proceedings of the 27th Australasian conference on Computer science-Volume 26* (pp. 159–166). Darlinghurst, Australia: Australian Computer Society, Inc.

Greenleaf, G. (2014). Sheherezade and the 101 data privacy laws: Origins, significance and global trajectories. *Journal of Law, Information and Science*, *23*(1), 4–49. doi:10.2139/ssrn.2280877

Greenleaf, G. (2015). *Global data privacy laws 2015: 109 countries, with European laws now a minority* (UNSW Law Research Paper No. 2015-21). Retrieved from https://ssrn.com/abstract=2603529

Harris, K. D. (2016, February). California data breach report. Retrieved from https://oag.ca.gov/sites/all/files/agweb/pdfs/dbr/2016-data-breach-report.pdf

Hiranandani, V. (2011). Privacy and security in the digital age: Contemporary challenges and future directions. *The International Journal of Human Rights, 15*(7), 1091–1106. doi:10.1080/13642987.2010.493360

Holmes, T. E. (2015, September 16). Credit card fraud and ID theft statistics. Retrieved from http://www.creditcards.com/credit-card-news/credit-card-security-id-theft-fraud-statistics-1276.php#7-consumer-sentinel-data-book

Hurwitz, J. B. (2013). User choice, privacy sensitivity, and acceptance of personal information collection. In S. Gutwirth, R. Leenes, P. de Hert, & Y. Poullet (Eds.), *European data protection: Coming of age* (pp. 295–312). Amsterdam, Netherlands: Springer.

Jain, A., Bolle, R., & Pankanti, S. (Eds.). (2006). *Biometrics: Personal identification in networked society* (Vol. 479). London, UK: Springer Science & Business Media.

Jain, A., Ross, A. A., & Nandakumar, K. (2011). *Introduction to biometrics*. London, UK: Springer Science & Business Media.

Kindt, E. (2013). *Privacy and data protection issues of biometric applications: A comparative legal analysis*. Berlin, Germany: Springer.

Kumaraguru, P., & Cranor, L. F. (2005). Privacy indexes: A survey of Westin's studies. Retrieved from http://www.cs.cmu.edu/~ponguru/CMU-ISRI-05-138.pdf

Malheiros, M., Preibusch, S., & Sasse, M. A. (2013). "Fairly truthful": The impact of perceived effort, fairness, relevance, and sensitivity on personal data disclosure. In M. Huth, N. Asokan, S. Capkun, I. Flechais, & L. Coles-Kemp (Eds.), *Trust and trustworthy computing* (pp. 250–266). London, UK: Springer.

Mamberti, D. A. (2012, September). Statement at the 67th ordinary session of the General Assembly of the United Nations on the rule of law at the national and international levels. Retrieved from https://www.un.org/ruleoflaw/files/Statement%20by%20Holy%20See.pdf

McCullagh, K. (2007). Data sensitivity: Proposals for resolving the conundrum. *Journal of International Commercial Law and Technology*, *2*(4), 190–201.

Murphy, R. S. (1995). Property rights in personal information: An economic defense of privacy. *Georgetown Law Journal, 84*, 2381–2417.

Myers, C. (2014). Digital immortality vs. "The right to be forgotten": A comparison of U.S. and EU laws concerning social media privacy. *Revista Română de Comunicare şi Relaţii Publice, 16*(3), 47–60.

Neal, M. (2014). Respect for human dignity as "substantive basic norm." *International Journal of Law in Context, 10*(1), 26–46. doi:10.1017/S1744552313000359

Nissenbaum, H. (1998). Protecting privacy in an information age: The problem of privacy in public. *Law and Philosophy*, *17*(5), 559–596. doi:10.2307/3505189

Nissenbaum, H. (2009). *Privacy in context: Technology, policy, and the integrity of social life*. Palo Alto, CA: Stanford University Press.

Ojanen, T. (2014, December). Privacy is more than just a seven-letter word: The Court of Justice of the European Union sets constitutional limits on mass surveillance: Court of Justice of the European Union Decision of 8 April 2014 in Joined Cases C-293/12 and C-594/12, digital rights Ireland and Seitlinger and others. *European Constitutional Law Review*, *10*(3), 528–541. doi:10.1017/S1574019614001345

Organization for Economic Cooperation and Development. (1980, September 23). OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data, para. 7, 50(a), 51. Retrieved from http://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm

Palfrey, J., & Gasser, U. (2013). *Born digital: Understanding the first generation of digital natives*. New York, NY: Basic Books.

Pesciotta, D. T. (2012). I'm not dead yet: Katz, Jones, and the Fourth Amendment in the 21st century. *Case Western Reserve Law Review*, *63*, 187.

Photopoulos, C.(2011). *Managing catastrophic loss of sensitive data: A guide for IT and security professionals*. Rockland, MA: Syngress, 2011.

Reding, V. (2012, January). *The EU data protection reform 2012: Making Europe the standard setter for modern data protection rules in the digital age*. Speech at Innovation Conference Digital, Life, Design, Munich, Germany. Retrieved from http://europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/12/26&format=PDF.

Rosen, J. (2012). The right to be forgotten. *Stanford Law Review Online*, *64*(88), 88–92.

Sariyar, M., Schluender, I., Smee, C., & Suhr, S. (2015). Sharing and reuse of sensitive data and samples: Supporting researchers in identifying ethical and legal requirements. *Biopreservation and Biobanking*, *13*(4), 263–270. doi:10.1089/bio.2015.0014

Schwartz, P. M., & Solove, D. J. (2011). The PII problem: Privacy and a new concept of personally identifiable information. *New York University Law Review*, *86*, 1814–1894.

Simonton, D. (1976). The sociopolitical context of philosophical beliefs: A transhistorical causal analysis. *Social Forces*, *54*(3), 513–523. doi:10.2307/2576278

Sisk, E. P. (2016). Technical difficulties: Protecting privacy rights in the digital age. *New England Journal on Criminal and Civil Confinement*, *42*, 101–143.

Soffer, T., & Cohen, A. (2015). Privacy perception of adolescents in a digital world. *Bulletin of Science, Technology & Society, 34*(5/6), 145–158. doi:0270467615578408

Solove, D. J. (2004). *The digital person: Technology and privacy in the information age*. New York, NY: New York University Press.

Taddicken, M. (2014). The "privacy paradox" in the social Web: The impact of privacy concerns, individual characteristics, and the perceived social relevance on different forms of self‐disclosure. *Journal of Computer‐Mediated Communication*, *19*(2), 248–273. doi:10.1111/jcc4.12052

Tene, O. (2011). Privacy: The new generations. *International Data Privacy Law, 1*(1), 15–27. doi:10.1093/idpl/ipq003

Trepte, S., & Reinecke, L. (Eds.). (2011). *Privacy online: Perspectives on privacy and self-disclosure in the social Web*. Heidelberg, Germany: Springer-Verlag.

United Nations. (1948). The universal declaration of human rights. Retrieved from http://www.un.org/en/universal-declaration-human-rights/

United Nations General Assembly. (1990, December 14). Guidelines for the regulation of computerized personal data files. Retrieved from http://www.refworld.org/docid/3ddcafaac.html

U.S. Federal Trade Commission. (2015, February). Consumer sentinel network data book for January to December 2014. Retrieved from https://www.ftc.gov/system/files/documents/reports/consumer-sentinel-network-data-book-january-december-2014/sentinel-cy2014-1.pdf

Volokh, E. (2000). Freedom of speech and information privacy: The troubling implications of a right to stop people from speaking about you. *Stanford Law Review, 52*(5), 1049–1124.

Wacks, R. (1989). *Personal information: Privacy and the law*. Oxford, UK: Clarendon Press.

Zuiderveen, B. F. J., Van, E. M., & Gray, J. (2015). Open data, privacy, and fair information principles: Towards a balancing framework. *Berkeley Technology Law Journal*, *30*(3), 2073–2131. doi:10.15779/Z389S18