https://doi.org/10.48009/2_iis_2009_165-174

DRAFTING INFORMATIONAL PRIVACY LAWS: INFORMATION SCIENCE PERSPECTIVE

Sabah S. Al-Fedaghi, Kuwait University, sabah@alfedaghi.com

ABSTRACT

In this paper, we describe an information science approach to drafting information privacy laws. Information scientists can contribute to the design of these laws not as mere technical experts, but rather as collaborators in design and construction of the legal text. Specifically, this paper concentrates on rules governing the handling of personal identifiable information (PII). First, we develop a systematic definition of PII and then identify five generic acts of handling it. A flow model is utilized to build complete scenarios where this type of information is involved. The results are applied to the proposed USA Personal Data Privacy and Security Act of 2005 in order to build a basic framework for drafting legal texts in the context of information privacy.

Keywords: Legal Aspects, Privacy, Law, Personal Identifiable Information

INTRODUCTION

Laws and policies impose many requirements on handling of information in business practices. From the point of view of an information scientist, information-related legal policies lack precision in regard to rules for information handling. The language in which regulations are stated in legal documents related to information processing is normally very vague, making it difficult to formalize requirements and constraints. Terms such as "collecting," "processing," "disclosing," and so forth are used loosely, without a pattern tying them to processes of information.

Information science style is usually based on design that identifies objects and operations united in sequences of processes. This style can be applied in drafting of information-related regulations. The aim of blending this technique in drafting of information-related laws is to promote the idea that information scientists can contribute to design of these laws not as mere technical experts, but rather as collaborators and partners in construction of the legal text.

We experiment with such an approach to drafting information privacy regulation. The approach involves simply constructing a state transition diagram of possible flow of personal identifiable information, taking into account all possible types of actions performed in processing this information. To focus the proposed methodology, we scrutinize the USA Personal Data Privacy and Security Act of 2005 (henceforth referred to as *PPSA*) through design of its underlying PII handling model and redrafting of some of its parts.

PPSA is a proposed regulation that aims

> To prevent and mitigate identity theft; to ensure privacy; and to enhance criminal penalties, law enforcement assistance, and other protections against security breaches, fraudulent access, and misuse of *personally identifiable information*. [Italics added] [12]

PPSA can be used as an example of the contribution of information science to drafting of legal texts in the information privacy context. The text of PPSA is first scrutinized (section 2) to show its weaknesses from the information scientist's point of view. Section 3 introduces a foundation for redrafting of PPSA that includes definition of personal identifiable information and possible actions that involve this type of information. The definition of PII and its flow model were published previously [1, 2, 3, 4], but not in the context of drafting of laws. Furthermore, some of the materials are presented differently than in the published version.

SCRUTINIZING THE LAW

PPSA is proposed legislation that "would help consumers better protect the privacy of their personal information in the face of recurrent data security breaches across the country" [8].

Key features of the act include the following:

- Provides notice to Americans when they have been harmed, and also addresses the underlying problems of lax security and lack of accountability in dealing with personal data.

- Requires data brokers to let individuals know what information they have about them and, where appropriate, to allow individuals to correct inaccuracies.
- Requires notice to law enforcement, consumers, and

credit reporting agencies when digitized sensitive personal information has been compromised.
- Prohibits the display and sale of Social Security numbers without consent.
- Addresses the government's use of personal data by requiring the General Services Administration to evaluate the privacy and security practices of government contractors potentially handling personal data, and to include penalties in government contracts for failure to protect data privacy and security [8].

In the next three subsections, we scrutinize the text of PPSA as presented in [12]. We discuss some of the issues that have motivated redrafting of the proposed act; however, concentrating on PPSA does not affect general applicability of the approach to design of information-related laws.

Mix of terms

PPSA uses several terms related to *personal identifiable information*. For example, "SEC. 2. FINDINGS" uses the terms personal identifiable information, identity, personally identifiable information, and personal information. The term "personally identifiable information" is defined in "SEC. 3. DEFINITIONS" to mean "any information, or compilation of information, in electronic or digital form serving as a means of identification, as defined by section 1028(d)(7) of title 18, United States Code" [12]. However, "personal identifiable information" and "personal information" have never been explained in PPSA. This mix of terms blurs identification of the most important concept in PPSA: personal (or is it personally?) identifiable information. The Short Title of the act uses the term "Personal Data" (in *Personal Data Privacy and Security Act of 2005*).

A great deal of vagueness may thus be the result in interpretations of PPSA. For example, suppose that the "personal information" is that *John confessed that he is thinking of hurting Mary* (e.g., confidential information in a medical report). According to PPSA, "A data broker shall, upon the request of an individual, clearly and accurately disclose to such individual … all personal electronic records pertaining to that individual…" Does this information belong to John, to Mary, or to both? We will introduce a definition of personal identifiable information that completely clarifies this issue.

In addition, the relationship between personal information and identities is specified indirectly through section 1028 of title 18, United States Code. "SEC. 2. FINDINGS" mentions "the integrity of their [individuals'] personal information *and* identities," so apparently they are separate terms. Implicitly, we can understand that personal information *embeds* identities. PPSA seems to consider "identity theft" a privacy issue. Some "identity theft" cases are privacy intrusion matters. Suppose that identification is stolen for the purpose of accessing a storage area and nothing else. On the other hand, consider the case of stealing a physical key to access the same area. Why is the theft in the first case related to privacy? We propose separating the issue of identity theft from the issue of mishandling of personal identifiable information.

Specific personal identifiable information

Section 1028(d)(7) of title 18, United States Code, mentioned in the definition of "personally identifiable information," specifies that information is issued by some authority:

> (3) the term "identification document" means a document made or issued by or under the authority of the United States Government, a State, political subdivision of a State, a foreign government, political subdivision of a foreign government, an international governmental or an international quasi-governmental organization which, when completed with information concerning a particular individual, is of a type intended or commonly accepted for the purpose of identification of individuals. [12]

Apparently, personally identifiable information is any information that embeds identity and is issued by some organization. So, for example, if some information (e.g., a threat) is circulated about a person whose nickname (i.e., not issued by a formal organization) is Fido, then it is not personal identifiable information, even though it happens that this information identifies him uniquely. Furthermore, even if Fido's picture is attached to the information, it is still not personal identifiable information. We claim that such a definition is incomplete for the purpose of information privacy laws.

A loose bag of acts on personal identifiable information

Close examination of PPSA reveals that uncertainty exists about the type of acts or operations on personally identifiable information that ought to be specified. The following is a list of types of actions to be controlled (italics added).

- "*own*, *use*, or *license* personally identifiable information" (SEC. 2. FINDINGS (5))

- "*collecting*, *transmitting*, or otherwise *providing* personally identifiable information" ((5) DATA BROKER)
- "*acquisition* of and *access* to sensitive personally identifiable information." ((10) SECURITY BREACH)
- "*obtains*, *accesses*, or *transmits*" (SEC. 104. AGGRAVATED FRAUD IN CONNECTION WITH COMPUTERS)
- "*access*, *use*, *compilation*, *distribution*, *processing*, *analyzing*, or *evaluating* personally identifiable information" (SEC. 301. TRANSPARENCY AND ACCURACY OF DATA COLLECTION)
- "*access*, *use*, *compilation*, *distribution*, *processing*, *analysis*, and *evaluation* of any personally identifiable information" (SEC. 303. RELATION TO STATE LAWS)
- "*collecting*, *accessing*, *transmitting*, *using*, *storing*, or *disposing* of personally identifiable information" (SEC. 401. PURPOSE AND APPLICABILITY OF DATA PRIVACY AND SECURITY PROGRAM, and SEC. 421. RIGHT TO NOTICE OF SECURITY BREACH)
- "*access to use* of personal electronic records" (SEC. 402. REQUIREMENTS FOR A PERSONAL DATA PRIVACY AND SECURITY PROGRAM (C))
- "*access*, *disclosure*, *use*, or *alteration* of personally identifiable information" (SEC. 402. REQUIREMENTS FOR A PERSONAL DATA PRIVACY AND SECURITY PROGRAM).
- "*use*, *transmission*, *storage*, and *disposal*" (RISK MANAGEMENT AND CONTROL)
- "*changes* to personally identifiable information systems" (SEC. 402- (e) Periodic Assessment).

It is not clear why it is "collecting, transmitting, or otherwise providing personally identifiable information" in "(5) DATA BROKER," while it is "collecting, accessing, transmitting, using, storing, or disposing of personally identifiable information" in "SEC. 401. PURPOSE AND APPLICABILITY OF DATA PRIVACY AND SECURITY PROGRAM, and SEC. 421. RIGHT TO NOTICE OF SECURITY BREACH." Why not include "using," "disposing," etc. in the first list? Is "alteration" different from "changing"? Is "access to use" different from "use"? What does "evaluation of personal identifiable information" mean?

These lists also seem incomplete. For example, nowhere is "creating" personal identifiable information mentioned, as in the case of data mining (e.g., from collected data, an agent concludes that *John Smith is a terrorist*). One list includes the term "analyzing," but this does not reflect the importance of an operation such as creation of personal identifiable information.

In conclusion, we claim that a more precise notion of actions on personal identifiable information is needed.

DEFINING PERSONAL IDENTIFIABLE INFORMATION

Different types of information pertinent to this paper are shown in Figure 1. So-called personal information is a type of information that includes PII and personal non-identifiable information (NII).

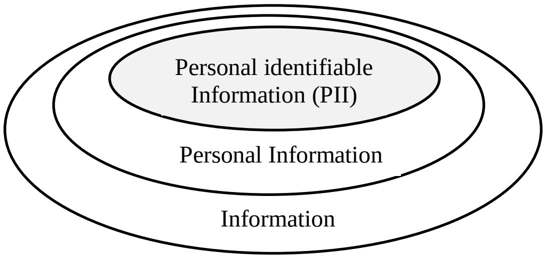

Figure 1. Different types of information.

Personal non-identifiable information is called "personal" because its owner (a person) has an interest in keeping it private, even though it does not embed his/her identity. This information is owned by the person, as in the expression "personal belongings," that is, a personal collection of research papers, songs, etc.

From a security point of view, PII is more sensitive than an "equal amount" of NII (to be discussed later). With regard to policy, PII has a more policy-oriented significance (e.g., the 1996 EU directive) than NII. With regard to technology, there are unique PII-related technologies (e.g., P3P) and techniques (e.g., k-anonymization) that revolve around PII. Additionally, PII possesses an objective definition that provides a means (identities of its *proprietor*) for separating it from other types of information, which facilitates organizing it in a manner not available to other types of information.

It is typically claimed that what makes data "private" or "personal" is either specific legislation, e.g., a company must not disclose information about its employees, or individual agreement, e.g., a customer has agreed to an electronic retailer's privacy policy. However, this line of thought blurs the difference between personal identifiable information and other "personal" information. Personal identifiable information has an "objective" definition in the sense that it is independent of such authorities as legislation or agreement.

PII involves relationships (e.g., possession) with *proprietors* (persons about whom the information communicates something), *non-proprietors* (persons who have other persons' PII), and *non-persons* such as institutions, agencies, or companies. For example, a person may possess PII of another person, or a company may have the PII of someone in its database; however, *proprietorship* of PII is reserved only for its proprietor regardless of who possesses it.

Reference as a base for defining PII

To base personal identifiable information on firmer ground, we turn to establishing some principles related to such information. For us, personal identifiable information is any information that has referent(s) to uniquely identifiable persons [2]. In logic, reference is the relation of a word (logical name) to a thing. Every PII refers to its proprietor(s) in the sense that it "leads to" him/her/them as distinguishable entities in the world. This reference is based on his/her/their unique identifier(s). The relationship between persons and their own PII is called proprietorship [1].

A piece of information is PII if at least one of the objects to which it refers is a singly identifiable person. Any singly identifiable person in the PII is called a proprietor of that information. The proprietor is the person about whom the PII communicates information. If exactly one object exists of this type, the PII is an atomic PII; if more than one singly identifiable person exists, it is a compound PII. An atomic PII is a piece of information about a singly identifiable person. A compound PII is a piece of information about several singly identifiable persons.

Any compound PII is privacy-reducible to a set of atomic PII. For example, *John and Mary are in love* can be privacy-reducible to *John and someone are in love* and *Someone and Mary are in love*. Note that our PII theory is a syntax (structural) based theory. It is obvious that the privacy-reducibility of compound personal identifiable information causes a loss of "semantic equivalence," since the identities of the referents in the original information are separated. Semantic equivalence here means preserving the totality of information, the pieces of atomic information, and their link.

Identifiers and PII

Consider the set of unique *identifiers* of persons. Ontologically, the Aristotelian entity/object is a single, specific existence (a particularity) in the world. For us, the identity of an entity is its *natural descriptors* (e.g., tall, black eyes, male, blood type A, etc.). These descriptors *exist in* the entity/object. Tallness, whiteness, location, etc. exist as aspects of the existence of the entity. We recognize the human entity from its natural descriptors. Some descriptors form *identifiers*. A *natural identifier* is a set of natural descriptors that facilitates recognizing a person *uniquely*. Examples of identifiers include fingerprints, faces, and DNA. No two persons have identical natural identifiers. An *artificial descriptor* is a descriptor mapped to a natural identifier. Attaching the number 123456 to a particular person is an example of an artificial descriptor in the sense that the number is not inherent in the (natural) person. An *artificial identifier* is a set of descriptors mapped to a natural identifier of a person. Date of birth (an artificial descriptor), gender (a natural descriptor), and a 5-digit ZIP code (an artificial descriptor) are three descriptors that form an artificial identifier for 87% of the U.S. population [10]. By implication, no two persons have identical artificial identifiers. If two persons somehow have the same Social Security number, then this Social Security number is not an artificial identifier because it is not mapped uniquely to a natural identifier.

A basic principle in our definition of PII is as follows:

Identifiers of proprietors are PII.

Such definition is reasonable since the mere act of identifying a proprietor is a reference to a unique entity in the information sphere. Every unique identifier of a person is a basic PII in the sense that this identifier cannot be decomposed into more basic PII. The second principle defines PII in general:

Any personal identifier or piece of information that embeds identifiers is personal identifiable information.

Thus, identifiers are the basic PII that cannot be decomposed into more basic PII. Furthermore, every complex PII includes in its structure at least one basic identifier.

Note that here we are not discussing the issue of flexibility or narrowness of PII definitions. This is a matter that can be settled after precise definition of PII. For example, PPSA limits PII by introducing the notion of "sensitive" PII.

Complexity of PII

The *atomic personal identifiable information* is the "unit" of personal identifiable information. It includes one identifier and, in general, non-identifiable information. We assume that at least some of the non-identifiable information is *about* the proprietor. In theory this is not necessary. Suppose that an identifier is amended to a random piece of non-identifiable information (noise). In the PII theory the result is (complex) atomic PII. In general, mixing noise with information preserves information.

The structure of a *complex* PII is constructed from several components:

- Basic PII plus non-PII, i.e., the PII *John S. Smith* and the non-PII *Someone is sick* form the atomic PII (i.e., PII with one proprietor) *John S. Smith is sick*.
- Complex PII forms more complex PII, e.g., *John S. Smith and Mary F. Fox are sick.*

Defining PII as "information identifiable to the individual" does not mean that the information is "especially sensitive, private, or embarrassing. Rather, it describes a relationship between the information and a person, namely that the information—whether sensitive or trivial—is somehow identifiable to an individual" [7]. However, personal identifiable information is more "valuable" than personal non-identifiable information, because it has an intrinsic value as "a human matter," just as privacy is a human trait. Does this mean that the scientific information of how to make a nuclear bomb has less intrinsic moral value than the pinfon *John is left-handed*? No, it means *John is left-handed* has a higher moral value than the non-PII *There exists someone who is left-handed*. It is important to compare equal amounts of information when we decide the status of each type of information [1].

PII and Non-PII

Consider the PII *Alice visited clinic* Y. It is PII because it represents a relationship, that of the proprietor Alice with an object, the clinic. Information about the clinic may or may not be privacy related information. For example, year of opening, number of beds, and other information about the clinic is not privacy related. Thus, such information about the clinic is not related to Alice's PII; however, when the information is that the clinic is an abortion clinic, then Alice's PII is related to this non-identifiable information about the clinic. That is, {*Alice visited clinic Y, Clinic is an abortion clinic*} has privacy significance. The decision about the boundary between a certain PII and its related non-identifiable information is difficult to formalize.

We notice that our analysis of PII and non-PII can help in determining the "amount" of PII. According to Amant [5],

> Recent bills have not adequately addressed these concerns. Under a regulatory regime like H.R. 4127, individuals entitled to notification by database businesses would still experience difficulty determining *how much personal information* had been taken because the legislation's proposed language does not require fact-finding by the database businesses to determine exactly *what personal information has been compromised*. [italics added]

ACTING ON PII

Previous sections have established a reasonably precise picture of the meaning of different types of information, personal identifiable information in particular. In this section we turn to the types of actions that can be performed on PII. While such operations as collecting, accessing, transmitting, using, storing, processing, etc. have been mentioned in many studies about information, a systematic framework for relating such operations in an organized manner has never been developed.

According to Al-Fedaghi [4], information is a *flowthing*. A flowthing refers to types of things that flow, hence, are received, processed, created, released, and communicated. Figure 2 shows the state transition diagram of information flow and includes five states: received, processed, created, disclosed, and communicated.

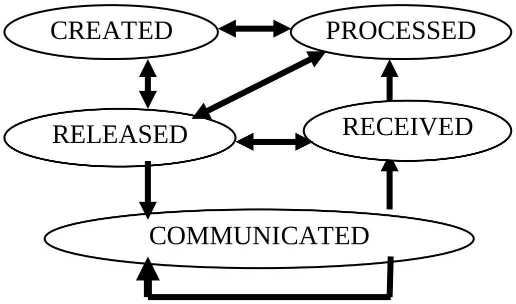

Figure 2. Transition states of information.

These are basic states of information, and information should at any time be in only one of them. Look at any piece of information and it should be in one and only one of these states at a given time. Take, for example, the state of being "stored." "Stored" is not a basic state because received information can be stored *received* information. It is received information that is stored in its received (original) state. Storage indicates "freezing" of information in its current state in the stream of flow. Similarly, *created* information can be *stored created information*. It is created information that is stored in its created state. Similarly, we may have stored processed information. Processing of information changes its form (e.g., translation, compression) but not its content (e.g., not new information).

Processed information cannot be in its received state, however, because "processing" changes its original form. It has *flowed* from one stage to another, analogous to the difference between raw materials and processed materials. As soon as any type of process is applied to raw materials, they are no longer raw materials.

Applying the flow model to PII, the generic operations (may also be called stages of the PII system) that can be performed on PII are thus receiving, processing, creating, disclosing, and communication. Each operation denotes an act on PII, and the arrows in the figure show possible flow of PII among stages of acting on PII. Different rules can be declared for PII in different states. For example, processing of PII may be permitted, but creating of PII is prohibited.

Figure 2 shows a circulation system of PII analogous to the model of circulation of water among various compartments in the environment. New PII is created by proprietors or non-proprietors (e.g., medical diagnostics by physicians). As soon as it enters the system through collection by someone, many possibilities for paths of flow exist. The PII could be used immediately, or stored for a long or short period. It could be processed to change its form or to extract embedded information, or mined to create new information from it. It could be transferred to another agent, or even returned to its proprietor. Meanwhile, it could be duplicated, thus producing copies that circulate independently in the system among different stages. It or any of its copies could be destroyed, anonymized, or encrypted. A piece of PII may have the history: released/communicated by proprietor, collected by agent 1, stored, processed (duplicated), released/communicated to agent 2, collected by agent 2, processed, mined, utilized (in conjunction with other persons' personal information) to create new PII, stored, etc.

Each of the five stages may include substages such as storage and "use." "Use" refers to going outside the circulation system to utilize PII. For example, PII such as address can be used to deliver a customer's purchase. "Customer purchase" is a type of use of PII. In contrast, receiving, processing, creating, disclosing, and communicating are not uses of PII; rather, they are states of PII.

The collecting stage is the information acquisition stage that accepts information from external suppliers and injects it into the circulation system. It includes the possibility of using the collected (raw) personal information; thus, in "use," information exits the system (e.g., customer address used in "product delivery"). This stage also includes the possibility of storing the collected information. At the collecting/receiving stage, we have to consider that the information can be collected from two sources: (1) proprietor, or (2) a third party (non-proprietor).

Processing the PI stage involves acting on (e.g., anonymizing, data mining, summarizing, recording, organizing, adapting or altering, retrieving, consulting, disseminating or otherwise making available, aligning or combining, blocking, erasing, or destroying, translating) PII. Processing of PII is performed on acquired information from the collecting stage or the creating stage (see figure 2). The actual processing occurs when information is modified in form. Data mining is a type of processing that may generate "new information" (flow goes to a creating stage). An example of generation of new information is the categorization of other persons' PII to generate the new PII that *John is a risk*. Other types of processing that do not generate new information, but only change the appearance of PII, include comparing, compressing, translating, and deleting. The "destruction" of PI is a type of process performed "when data no longer serve a purpose, and if it is practicable, it may be necessary to have them destroyed (erased) or given an anonymous form" [10].

The releasing stage involves declaring PII to be information that can be communicated outside the system. For example, passengers in an airport can be in a state of being released, in the sense that they have finished all necessary procedures and are just waiting to board. They are released but not yet in a transported state (information is not in the communication channel). PII disclosure/release is performed on information acquired from a proprietor or from collected, created, or processed information. Disclosure depends on the communicating stage that transfers the information from the disclosing agent to the collecting agent, which can be the same agent, another agent, or the proprietor him/herself.

RE-DRAFTING PPSA

The definitions and flow model given in the previous sections provide a description of PII handling operations that can be used to build rules for information privacy. *Information handling* means performing the five operations of receiving, processing, creating, releasing, and communicating, and secondary operations such as storing, destroying, copying, etc.

According to the definitions and flow model of PII, we can use the terms "personal identifiable information" and "identity" in a precise way. PII is information that refers uniquely to an identifiable person. An identity is a minimal PII. We can classify identities into categories like those created by formal organizations, those created from secondary descriptors, etc.

Utilizing the PII flow

In this section, we rewrite some sections of PPSA that include incomplete and/or unclear actions on PII. We choose to rewrite parts of PPSA instead of starting from scratch, and develop a new text, in order to achieve two goals simultaneously: to uncover ambiguity in PPSA and demonstrate the clarity of the new version.

1. In PPSA, we find:

> (10) SECURITY BREACH-
>
> (A) IN GENERAL- The term "security breach" means compromise of the security, confidentiality, or integrity of computerized data through misrepresentation or actions that result in, or there is a reasonable basis to conclude has resulted in, the unauthorized *acquisition of and access* to sensitive personally identifiable information. [Italics added]

This paragraph can be rewritten:

> (New A) IN GENERAL- The term "security breach" means compromise of the security, confidentiality, or integrity of computerized data through misrepresentation or actions that result in, or there is a reasonable basis to conclude has resulted in, the unauthorized *acquisition, processing, creation, releasing, or communication* of sensitive personally identifiable information.

In this version of (A), "acquisition" is used instead of "receiving" in the flow model. It is more complete than the original, because it is possible that an agent can perform not only authorized acquisition, but also unauthorized processing, creation, releasing, or communication of PII. Unauthorized "access" in (A) implies releasing of PII.

2. In PPSA, we find:

> (B) EXCLUSION- The term "security breach" does not include a good faith acquisition of sensitive personally identifiable information if the sensitive personally identifiable information is not subject to further unauthorized disclosure.

"(B) EXCLUSION" does not recognize that "acquisition" is but one stage in handling of PII. Figure 1 shows a complete picture of other possible stages of handling PII. Is it possible that an agent collects PII in good faith, but then processes it not in good faith? For example, the agent collects PII for the purpose of delivering goods (good faith); nevertheless, the agent processes it for internal use without releasing it to a third party. Also, it is possible that PII is created, released, and communicated internally, without being disclosed to a third party. "Good faith acquisition" ought to be clarified as the first stage in handling of PII, or we have to specify "good faith creation," "good faith release," "good faith communication" of PII. To cover all possibilities, we rewrite PPSA as follows:

> (new B) EXCLUSION- The term "security breach" does not include a good faith *handling* of sensitive personally identifiable information if the sensitive personally identifiable information is not subject to further unauthorized *release and communication*.

3. In PPSA, we find:

> SEC. 301. TRANSPARENCY AND ACCURACY OF DATA COLLECTION.
>
> (b) Disclosures to Individuals-
>
> (1) IN GENERAL- A data broker shall, upon the request of an individual, clearly and accurately disclose to such individual for a reasonable fee all personal electronic records pertaining to that individual maintained for disclosure to third parties in the databases or systems of the data broker at the time of the request.
>
> (2) INFORMATION ON HOW TO CORRECT INACCURACIES- The disclosures required under paragraph (1) shall also include guidance to individuals on the processes and procedures for demonstrating and correcting any inaccuracies.
>
> (c) Creation of an Accuracy Resolution Process- A data broker shall develop and publish on its website timely and fair processes and procedures for responding to claims of inaccuracies,

including procedures for correcting inaccurate information in the personal electronic records it maintains on individuals.

Inaccuracies

In this section of PPSA, incompleteness is caused by a partial view of the PII flow. If inaccurate PII exists, three sources of this inaccuracy can be identified:

(a) inaccuracy in the received PII

(b) inaccuracy in the processed PII (e.g., inaccurate translation)

(c) inaccuracy in created PII

Each of these types may require different regulations for handling of PII. For example, "inaccuracy in the received PII" may involve the outside source that provided such information. "Inaccuracy in the received PII" is completely within the responsibility of the PII handler and, hence, requires "internal" scrutiny. "Studies have shown that at least some of the information in almost all of the dossiers created by [commercial data brokers] is inaccurate and that these errors have adversely impacted individuals" [11]. PPSA concentrates on accuracies in received information, especially when it is available as public record information. There is a need to approach this topic systematically through identification of different types of PII. Each type of information has its own sources of inaccuracies (e.g., unreliable source of received PII, defective translation of processed PII, questionable mining method/decision making that produces created PII, insecure bookkeeping of released PII, error-ridden transmission of transferred PII.)

Notification to proprietors

PPSA mandates notification of the individual whose sensitive PII was breached. A business entity is exempt from the notice requirements if a risk assessment concludes that there is a "de minimis" risk of harm to the individuals whose sensitive PII was at issue in the security breach. Classification of PII assists in classifying the threats that warrant public notification. For example, suppose that internal processing has determined that *Smith is a dangerous person* based on collected data about Smith. Does PPSA require that this *created* information be disclosed to Smith, or does disclosure include only *collected* information? The agent with unauthorized access to this information can make several interpretations, including:

- This business entity has determined that *Smith is a dangerous person*

- This business entity has received the information that *Smith is a dangerous person*

Clearly, a *de minimis* risk of harm is viewed differently since the reliability of the origin of this information (FBI vs. tabloids) may be a factor in possible intruder actions.

This notification issue raises a great deal of concern. For example, a director of information policy warns that "The idea is to increase security. But opening databases to access is not increasing security. The issue is supposed to be security, and they're going to make databases less secure" [9]. In our approach, it is possible to make different rules for releasing different types of PII. For example, it is reasonable to require proprietor's access to his/her PII that a company has collected from outside sources, whereas access to created or processed PII can be limited.

Difference between PII security and PII privacy

Warning that "the idea is to increase security" [9] in a comment on the disclosure policy in PPSA reflects uncertainty about the relationship between PII security and PII privacy. In general, security is defined as "the quality or state of being secure—to be free from danger." In the context of information (including PII), security is a state of well-being of information that maintains confidentiality (e.g., only sender, intended receiver), authenticity (e.g., assurance of the origin), integrity (e.g., trustworthiness of data), and availability (e.g., ability to use) of information. This security definition can be applied to received, processed, created, released, and communicated states of PII, along with substates such as stored PII.

The security of PII is one aspect in maintenance of privacy of PII, as indicated in many laws and privacy guidelines (e.g., OECD privacy guidelines [10]). Privacy of PII goes beyond preserving the security of information to imposing of rules on gathering, releasing, creating, and using PII between consenting partners.

It may be a good approach to separate the issues of "identity theft" from "breach of PII." Identity theft is usually aimed at criminal usage of the identity of the victim. According to Amant [5],

> Beyond actual or potential pecuniary harm, database breaches can have serious repercussions that *have nothing to do with stealing identities* or the associated economic loss, at least when using the traditional definition of economic loss. A security breach that discloses substantial health

> information could lead to embarrassment or reputational harm, with only subtle, largely incalculable effects on future economic well-being … Recent bills have not adequately addressed these concerns. [Italics added]

In addition, according to Amant [5], "Supporters of linking the consumer notification trigger to the risk of identity theft make the valid point that a more sensitive trigger could lead to over-notification."

Using the PII model introduced previously, we can identify two types of identity theft:

1. Identity theft that compromises information privacy through use of identity to gather more PII. This type involves unauthorized seizure of identity (receiving stage) and then processing, creating, releasing, and/or communicating PII.

2. Identity theft used in non-privacy related acts such as to extract money, benefits, etc. This type involves unauthorized seizure of identity but does not proceed to the processing and creation stages, as shown in figure 3.

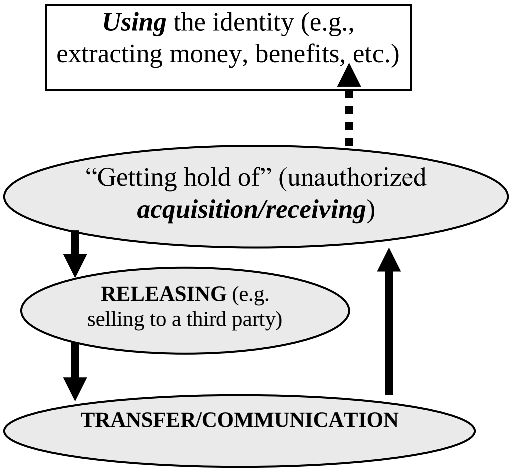
Using the identity (e.g., extracting money, benefits, etc.)
"Getting hold of" (unauthorized ***acquisition/receiving***)
RELEASING (e.g. selling to a third party)
TRANSFER/COMMUNICATION

Figure. 3. In non-privacy identity theft, the PII (identity) flow does not proceed to the stages of processing and creation.

In type 2, the identity is utilized in some act with no interest in obtaining further information about the victim. It is a security breach more than a privacy problem. It involves the stage of "getting hold of" (unauthorized acquisition/ receipt of) identity and using it as shown in figure 3. The dotted arrow in figure 3 indicates a transfer from information flow to another type of flow such as money or actions.

Breaching PII usually aims at the non-identity parts of PII. In the case of spying, the interest is in the person's behavior (e.g., consumption of alcohol); in the case of direct marketing, the interest is in the person's condition or interests (e.g., health, hobbies). Issues such as completeness, accuracy, alteration, integrity, and disposal of information are relevant to the non-identity parts of PII.

We can identify identities that can/cannot be utilized to process PII. Accordingly, a notification policy can be designed. For example, if the stolen identity includes a credit card number, then an immediate act of blocking any *use* of that credit card can be taken. We assume that a credit card number does not lead to other PII the way a Social Security number does.

Amant [5] proposes a provision that would state that notification of security breaches is required unless there is *no* risk of harm. From previous discussion, we can see that the risk is sensitive to two types of harm: information privacy and non-privacy harms. This distinction can contribute greatly to wording of the law.

Compound PII

In PPSA, the issue of compound PII is completely neglected. In SEC. 301. TRANSPARENCY AND ACCURACY OF DATA COLLECTION, (b) Disclosures to Individuals states that

> (1) IN GENERAL- A data broker shall, upon the request of an individual, clearly and accurately *disclose* to such individual for a reasonable fee all personal electronic records pertaining to that individual … [italics added]

Suppose that the information is *Smith, Mary's husband, had a relationship with Alice*. Does this mean that Mary has the right to access this PII? Apparently, PPSA, as stated above, deals with atomic PII that includes PII referring to a single proprietor. Nevertheless, this ought to be specified explicitly. Furthermore, compound PII should be specifically recognized and methods for releasing it specified.

Thus, PPSA can be rewritten:

> (1) IN GENERAL- A data broker shall, upon the request of an individual, clearly and accurately *disclose* to such individual for a reasonable fee proprietary personal electronic records pertaining to that individual … [italics added]

"Proprietary" means PII of which the individual is the proprietor. For example, in *Smith, Mary's* *husband, had a relationship with Alice*, Mary's proprietary information is: *Mary has a husband*.

CONCLUSION

This paper has introduced an information science approach to drafting information privacy laws. Specifically, the paper concentrates on rules to govern the handling of personal identifiable information (PII). The results are applied to the proposed USA Personal Data Privacy and Security Act of 2005. The approach is very promising both for drafting of laws and for designing of relevant systems. Certain approaches in software engineering extract software requirements through converting legal text into semiformal constraints and rules.

REFERENCES

1. Al-Fedaghi, S. (2007). Anatomy of personal information processing: application to the EU privacy directive. *International Journal of Liability and Scientific Enquiry, 1*. http://www.inderscience.com/storage/f715896102124113.pdf
2. Al-Fedaghi, S. (2008). Scrutinizing the rule: Privacy realization in HIPAA. *International Journal of Healthcare Information Systems and Informatics*, *3*(2).
3. Al-Fedaghi, S., & Thalheim, B. (2008). Databases of personal identifiable information, IEEE/ACM/IFIP SITIS'08 *Workshop on Security and Privacy in Telecommunications and Information Systems*, November 30th–December 3rd, Bali, Indonesia.
4. Al-Fedaghi, S. (2008). Software engineering interpretation of information processing regulations, *IEEE 32nd Annual International Computer Software and Applications Conference*, Turku, Finland, July 28–August 1, 2008.
5. Amant, B. St. (2007). The misplaced role of identity theft in triggering public notice of database breaches. *Harvard Journal on Legislation, 44*. http://www.law.harvard.edu/students/orgs/jol/vol44_2/stamant.pdf
6. Breaux, T., & Antón, A. I. (2008, Jan.). Analyzing regulatory rules for privacy and security requirements. *IEEE Transactions on Software Engineering*, *34*(1), 5-20.
7. Kang, J. (1998). Information privacy in cyberspace. Transactions. *Stanford Law Review 50*(1193), 1212-20.
8. Leahy, P. (2005). Specter, Leahy introduce Personal Data Privacy and Security Act of 2005. http://leahy.senate.gov/press/200506/062905a.html
9. McCullagh, D. (2005, June 29). Senators propose sweeping data-security bill. *cnet news*. http://news.cnet.com/Senators-propose-sweeping-data-security-bill/2100-7348_3-5769156.html
10. OECD. (1980). OECD guidelines on the protection of privacy and transborder flows of personal data. http://www.oecd.org/document/18/0,2340,en_2649_34255_1815186_1_1_1_1,00.html
11. Schaufenbuel, B. J. (2008, June). Breaking the Congressional logjam: A new approach to the regulation of commercial data brokers. The John Marshall Law School. http://www.jdsupra.com/documents/3c414df5-6d85-402c-a410-aa6427b78ed3_searchable.pdf
12. Text of S. 1332(2005). [109th]: Personal Data Privacy and Security Act of 2005. http://www.govtrack.us/congress/billtext.xpd?bill=s109-1332
13. Sweeney, L. (1997). Weaving technology and policy together to maintain confidentiality. *Journal of Law, Medicine and Ethics, 25*, 98-110.