Journal of Information Technology Impact
Vol. 11, No. 2, pp. 101-120, 2011

A Flow-Based Model to Assess Privacy Impact

Sabah Al-Fedaghi[1]
Kuwait University
Kuwait

Abdilhadi Jeragh
Kuwait Petroleum International LTD
Kuwait

Abstract

Privacy regulations and laws have been introduced to control handling of information about persons. Compliance with these regulations and laws presents a significant challenge for organizations holding data about identifiable persons. A privacy impact assessment (PIA) is an important instrument for ensuring conformity to regulatory requirements, determining risks, and evaluating privacy protections. This paper scrutinizes recent approaches to PIA and introduces a foundation for such a process. This foundation is used to reformulate basic notions such as definition, handling, and uses of personal identifiable information. As an example case, we apply our methodology to the 2010 US Department of Homeland Security's guidance on PIA.

Keywords: Personal information, identifiable information.

Introduction

Advancements in information and communication technology (ICT) are rapidly transforming society in ways that intensify interactions among citizens, businesses, and government. ICT has witnessed fast developments in storage, processing, and communication of information at unprecedented speed and volume. These developments have allowed for greatly increased volume of collected personal data and the capacity to manipulate information. Governments and companies have been quick in applying ICT to enhance their functions and services. The capacity to assemble this information for commercial and government operations represents a great risk to privacy, especially in areas such as data mining and surveillance.

Such a development is a sensitive issue in modern society. The security of information about persons has unique characteristics in comparison with sensitive data protection. The breach of security of such information can result in direct physical or psychological injury to a person.

This effect has led to introduction of privacy regulations and laws to control the ways in which information about persons is handled. These regulations and laws can apply to governments, businesses, and persons. Compliance presents a significant challenge for private organizations and government agencies holding data about identifiable persons. To mention a few examples, the EU Data Protection Directive (1995) requires each EU member nation to establish a national authority for administering data privacy issues, endowed with investigative and

intervention powers. The Canadian Personal Information Protection and Electronic Documents Act (2000), effective January 1, 2001, created an oversight and enforcement mechanism concerned with methods of handling personal information.

In response to such developments, several techniques have been introduced to manage privacy risks related to the collection, use, and disclosure of personal information, including privacy impact assessment, compliance audits, privacy seals and associated self-regulatory initiatives, and privacy enhancing technologies. This paper concentrates on privacy impact assessment (PIA).

PIA is viewed as a means by which business and government can identify and avoid privacy problems. According to Shroff (2007),

> In Hong Kong, privacy impact assessment is an important part of a policy approach to building trust and confidence in e-business. In Australia the process is recommended as part of any new Public Key Infrastructure system. A number of Canadian governments, federal and provincial, have or are developing policies requiring privacy impact assessment to be undertaken on new projects.

The US Office of Personnel Management (OPM, 2010) asserts,

> A privacy impact assessment (PIA) is one of the most important instruments through which the Office of Personnel Management (OPM) establishes public trust in its operations… The PIA is a vital tool that evaluates possible privacy risks and the mitigation of those risks at the beginning of and throughout the development life cycle of a program or information technology (IT) system.

The US E-Government Act of 2002 declares that to insure "citizen-centered electronic Government" and sufficient protections for the privacy of personal information, all government agencies "developing or procuring information technology that collects, maintains, or disseminates information that is in an identifiable form" are required to develop PIA (E-Government Act, 2002). According to the Privacy Office of the Department of Homeland Security (DHS),

> The Privacy Impact Assessment [PIA] is one of the *most important instruments* through which the Department creates transparency and establishes public trust in its operations…
>
> The PIA is a vital tool that evaluates possible privacy risks and the mitigation of those risks at the beginning of and throughout the development life cycle of a program or system… Privacy considerations must be contemplated systematically throughout the Department… By documenting the procedures and measures through which the Department protects the privacy of individuals, the Department is more transparent and can better carry out its mission. (DHS, 2010)

PIA is widely recognized as an issue in various situations. For example, the 2010 BP Oil Spill Response raised the issue of government use of "Internet-based platforms … [to collect] information posted by individual account users" (OPS-NOC, 2010). This task required PIA. In

2009, the European Commission established working groups to create an RFID Privacy Impact Assessment Framework (EDRI-Gram, 2010).

To give an idea of the materials and issues involved in PIA, questions about the Travel Expense Reporting System (TERS) serve as examples:

> Does this system collect any personal information in identifiable form about individuals?
> What elements of PII are collected and maintained by the system?
> Why is the information being collected?
> What are the sources of the information in the system? (TERS, 2010)

For such an important tool, we claim that current approaches to PIA are not based on a firm conceptual foundation. To substantiate this claim, we analyze a recently developed PIA as a representative sample to show the elements lacking in present PIA specifications. The PIA methodology introduced in this paper is generally applicable to any PIA, and each organization or agency can develop its own version of PIA based on this methodology.

Sample Case: US Department of Homeland Security Guidance on PIA

In June 2010 the US Department of Homeland Security (DHS) issued the latest version of "guidance on Privacy Impact Assessments (PIA)" (DHS, 2010), hereafter DHS PIA. This was the latest report on PIA flow since publication of *Privacy Impact Assessment Guidance* in 2006 and 2007 and the 2004 *Privacy Impact Assessments Made Simple*. According to the DHS PIA definition,

> The PIA demonstrates to the public and to Congress that systems owners and developers have consciously incorporated privacy protections into the development, implementation, and operation of their systems. It is also a tool for individuals working on a program or accessing a system to understand how to best integrate privacy protections while working with PII.

The DHS PIA provides a template for conducting and auditing privacy based on a set of questions divided into eight sections or topics.

This provides an opportunity to scrutinize basic concepts in an actual PIA and to redraft some part to demonstrate the possibility of developing a firm theoretical foundation for PIA. We will analyze some of the questions in the DHS PIA template and remodel it based on our new foundation. Not every piece of the template will be included, because reformulating the main components of the template is sufficient to show the applicability and advantages of the new conceptual foundation without unnecessary repetition.

Problem: Defining Personal Identifiable Information

DHS PIA defines personally identifiable information (PII) as information "that permits the identity of an individual to be directly or indirectly *inferred* [emphasis added], including any other information which is linked or linkable to that individual" (DHS, 2010). So if information *x* leads to identifying an individual, then *x* is PII. This *inference*-based definition seems at times not to make sense; for example, does it refer to a logical inference or to the common sense of

inference? In the absence of any indication of logical treatment, a sense of inference in language is quite vague and subjective. Suppose that Sherlock Holmes inferred that the killer lived close to the train station, and this *led* the police to determine the identity of the killer. Does Holmes's inferred information constitute PII?

The term "permit" in the DHS PIA is also fuzzy. For example, a "super user" password in a computer system *permits* direct inference of (and can be used to distinguish or trace) the identities of all individual users. Does the super user password comprise PII? Information can be in the possession of a manager that *permits* him/her to know the identity of employees in administrative units under his/her control. Is this information PII?

The phrase "Information which is linked or linkable to that individual" also seems fuzzy. Suppose that after searching John's house, the police find a recipe for a certain kind of food in the kitchen. Since the recipe is linked to John, is it PII?

Such a definition is common in many works on privacy. A term similar to PII, "personal data," is defined in EU Directive 95/46/EC as follows:

> Article 2a: 'personal data' shall mean any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity.

In general, "any information *relating* to an identified or identifiable natural person" is at best a fuzzy description. Suppose the identifiable natural person is Albert Einstein; *all* related information includes any information he produced and owned during his lifetime. The definition does not answer what PII is; rather, it talks about what is related to PII.

So, it seems that the issue of what constitutes PII is an unsettled issue. In a 2010 article in *Communications of the ACM*, Narayanan and Shmatikov (2010) claim,

> In today's data protection practices, both in the U.S. and internationally, "personally identifiable information" (PII)—or, as the U.S. Health Insurance Portability and Accountability Act (HIPAA) refers to it, "individually identifiable" information— has become the lapis philosophorum of privacy. Just as medieval alchemists were convinced a (mythical) philosopher's stone can transmute lead into gold, today's privacy practitioners believe that records containing sensitive individual data can be "de-identified" by removing or modifying PII.

Narayanan and Shmatikov (2010) specifically target privacy laws that "account for the possibility of deductive disclosure and… do not lay down a list of informational attributes that constitute PII."

Various definitions of PII have been used in important policies and laws. These definitions have created a great deal of disagreement. Narayanan (2010) declares that "For a concept that is so pervasive in both legal and technological discourse on data privacy, PII is surprisingly difficult to define. […] PII is meaningless, […] The term means next to nothing and must be greatly de-emphasized, if not abandoned, in order to have a meaningful discourse on data privacy."

Definition of PII

Contrary to the previous claims, a reasonable definition has been given and elaborated in many publications, including Al-Fedaghi (2005–2011), Al-Fedaghi and Deabas (2006), Al-Fedaghi and Thalheim (2008), Al-Fedaghi, Fiedler, and Thalheim (2006), Abdul Ghani and Sidek (2009), and Sato et al. (2009). The following discussion of PII and its handling is summarized from these references, with some new explanations and illustrations added.

Definition

The basic concept of personal identifiable information assumes two basic types of entities: *natural persons* and *non-natural persons.* PII is any information for which the *referent* signifies a natural person. The referent is said to be the *proprietor* of PII. *Proprietorship* of PII is different from the concepts of possession and copyrighting. It is also different from the legal concept of ownership.

Without loss of generality, we limit our discussion to linguistic PII. The formal semantics of the word *referent* is very important in this line of thought. The *referent* is recognized by mapping the word (logical name) in relation to the actual object (natural person) in reality. This mapping to a natural person limits possible *extension* to specific human beings. Personal identifiable information is any information that contains a *referent* to uniquely identifiable persons. Every PII *refers to* its proprietor(s) in the sense that it "leads to" him/her/them as distinguishable entities in the world.

Accordingly, there are two types of PII:

1. Atomic PII, where the PII *refers to* a single proprietor.
2. Compound PII, where the PII *refers to* more than one proprietor.

Proprietorship of PII is nontransferable in the absolute sense. Others can possess or (legally) own it, but they are never its proprietors (i.e., it cannot become their proprietary information). Atomic PII of a proprietor is proprietary information of that proprietor, while others (e.g., other individuals, companies) can only possess it. Compound PII is proprietary information of its referents: all donors of pieces of atomic PII embedded in the compound PII.

Compound PII can be reduced to a set of atomic PIIs (Al-Fedaghi, 2005). For example, *John and Mary are in love* can be converted to {*John and someone are in love, Someone and Mary are in love*}. Thus, it is possible to separate the privacy view in compound PII.

Construction of PII

Consider the set of unique *identifiers* of persons. In this paper, the identity of an entity comprises its *natural descriptors* (e.g., tall, brown 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 (minimal) set of natural descriptors that facilitate recognizing a person *uniquely.* Examples of identifiers include fingerprints, faces, and DNA. In general, 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 it is not recognizable in the (natural) person. *An artificial identifier* is a (minimal) set of descriptors mapped to a natural identifier of a person. 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 the 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. 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 the concern here is not the issue of flexibility or narrowness of PII definitions. This is a matter that can be settled after a precise definition encompassing all types of PII is developed.

Ambiguity

The *reference* (to reality) concept in defining of PII is an important aspect. Technically, identification is not necessary to identify a referent. For example, the statement *John F. Kennedy is a very busy airport* includes identification but does not include a natural person referent; therefore, it is not PII. It is "equivalent" to standing beside the airport and stating, *This is a very busy airport*.

Anwar (2009) criticized the definition of PII given above:

> [It] includes observation, reputation, or even public information in the realm of personal information, and thereby, may introduce more ambiguity. For example, information referring to John in his professional capacity as mayor, for example, should not be considered as his personal information.

This is analogous to criticizing the standard definition of integer set because it is infinite. The point is that this definition is inclusive/exclusive with regard to membership. It provides a specification of the membership base that is used to define other subsets such as public PII, sensitive PII, and so forth.

Sensitivity

The sensitivity of information is one of the most important factors in determining the individual's perception of privacy (Lederer, Beckmann, & Dey, 2003). In general, the notion of sensitivity is a particularly difficult concept. 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" (Kang, 1998). Clearly, much of PII, as defined in the previous section, is insignificant in terms of privacy. Insignificance does not imply lack of value. An insignificant amount of gold not worth the effort to mine is not worthless. Even though no criterion precisely divides significant from insignificant types of PII, it seems that, in most cases, the difference between them is apparent. Many works in the area of privacy have no difficulty identifying (significant) privacy in domains such as health information and financial information. "Significance" here refers to the threshold of an intrinsic value of PII.

Problem: Handling of PII

In section 1.0, DHS PIA demands to "list all statutory and regulatory authority for operating the project, including the authority to collect the information listed in question 2.1." Going to section 2.1, we read:

> Identify (1) the categories of individuals for whom information is collected, and (2) for each category, list all information, including PII that is *collected and stored* by the project…. If the project or system creates new information … describe how this is done and the purpose of that information… If the project receives information from another system, …, describe the system from which the information originates" [italics added].

Suppose that a category of individuals is related to release and transfer of PII; section 2.1 of DHS PIA mentions explicitly only "PII that is *collected and stored.*" The implication here is that types of handling such as release and transfer of PII are not identified. Why not be complete and include these? Notice that released PII is information within the system, analogous to products included in the inventory of a factory waiting to be shipped out.

Later, DHS PIA adopts the known Fair Information Practice Principles, including

Principle of Purpose Specification: Explain how the collection ties with the purpose of the underlying mission of the organization and its enabling authority.

But, does this mean only the *collection* of PII? Why not apply "Explain how" to the creation, processing, release, and transfer of PII? Later DHS PIA requires the reader to "List each use of the information collected or maintained." Why mention only *collected* OR *maintained*? Is "collected" synonymous with "maintained" in this context?

Also in DHS PIA, we read that "to define the scope of the information requested and/or collected, as well as reasons for its collection," then,

"*2.1 Identify the information the project collects, uses, disseminates, or maintains.*"

We observe that "collect," "use," "disseminate," and "maintain" are arbitrary operations performed by projects. For example, if a project creates PII by data mining, do we need to identify such information? Does creation of PII establish a type of collection or use, disseminate, or maintain information? Do these four operations cover all possible ways of handling of PII? Why, for example, not identify storing, destroying, copying, etc.? Are these covered by

maintaining of PII? Suppose that an agency does not collect, use, disseminate, or maintain PII. Suppose that for some reason (e.g., mistaken address) it *receives* PII; does it return it to the sender? Is there any obligation to protect it?

We argue that different ways of handling PII are not based on any systematic method. Such a claim will become clearer next, when we demonstrate such a method.

Handling of PII

The flowthing model (FM) is a method for specifying possible states of things that flow (Al-Fedaghi, 2005–2011). Flowthings, e.g., PII, are things that can be received (arrive and are accepted), processed, created, released, and transferred. A flow system (flowsystem – see Figure 1) represents stages at which flowthings are in various exclusive states. States of PII in this sense refer to their condition, as in states of matter: solid, liquid, and gas.

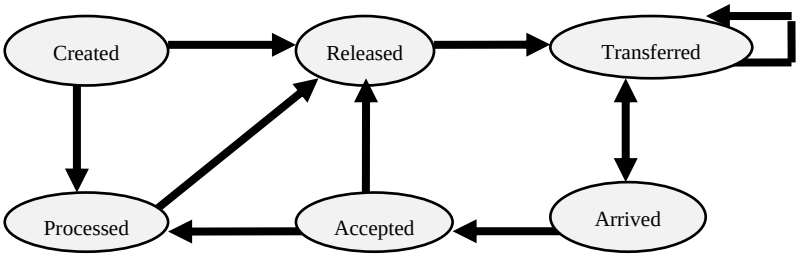

Figure 1. PII flowsystem (assuming released flowthing is not returned).

Accordingly, operations on PII are limited to six operations. This makes rules for handling PII very specific and easier to describe in privacy laws.

Suppose that the information *John is a security risk* flows into (circle 1 in Figure 2) DHS. *John is a security risk* is in the transferred state. It has not yet been delivered to DHS. In a manual system it is still in the possession of the postman, even though it is inside DHS. When it *arrives*, it can be rejected and sent back (circle 2), or it can move to the accepted state, thus entering the DHS system. Assuming no copying, *John is a security risk* is in one and only one state: transferred, arrived, or accepted. Upon acceptance it may flow to be released (circle 4) and transferred (circle 7) to another department (circle 8). Or, it may flow (circle 5) to be processed, to be changed in form (e.g., translated). In this case, it may be released (circle 6) and transferred (circle 7) to another department (circle 8). Again, in this transformation *John is a security risk* in only one state at any given moment.

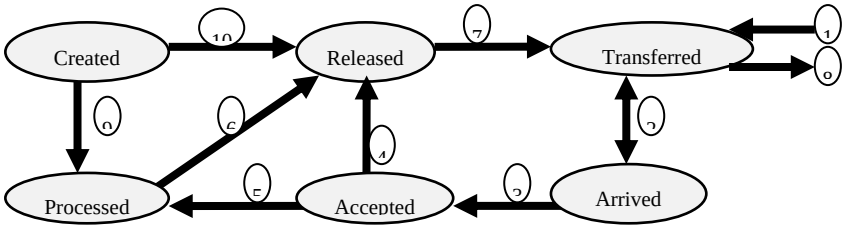

Figure 2. PII flowsystem (assuming Released flowthings do not return).

On the other hand, *John is a security risk* could be created in DHS, to be released (circle 10), transferred (circle 7), and flow to the outside (circle 8). The PII that *John is a security risk* has moved from the state of *created* PII, to the state of *released* PII, then to the state of being *transferred* to the outside world, three stages through which PII has moved in sequence. Note that DHS may decide to *release* it, then let it stay in this condition until time to transmit. Regardless how short the time, there is a state of release (decision to disclose) between creation and transfer (decision to transmit) to the outside. OR, *John is a security risk* could be further processed (circle 9), and so forth.

It is possible that as DHS creates the information that *John is a security risk*, it also receives (arrival and acceptance) *John is a security risk* from an outside source. These are two different instances (flowthings) of a piece of information. It is not possible to have the same instance in the *received* and *created* states simultaneously. It is analogous to having liquid water and also the same amount of frozen water; both are identical instances of water, but in different states.

The FM not only sums up the mutually exclusive generic stages of the life cycle of PII but also describes the transformation between these stages. *Accepted* PII cannot go (flow) directly to a *transferred* state without going through the *released* stage (regardless how short the released state in time). One can release documents by putting them in the "out tray" to be physically transferred hours or days later.

Transfer is that part of the flowsystem that interfaces with the channel. As an illustration, consider the home system shown in Figure 3 (left). Transfer is the entry section of the house, leading from the gate to the door. Upon arrival at the door, the flowthing may not be accepted, hence the bidirectional arrow between Transfer and Arrival in Figure 1. If accepted, the flowsystem enters the house. Release is the counterpart of Arrival. The flowthing stands by the door waiting to be transferred to the street. If the door is jammed, the flowthing can stand in the released state for a while.

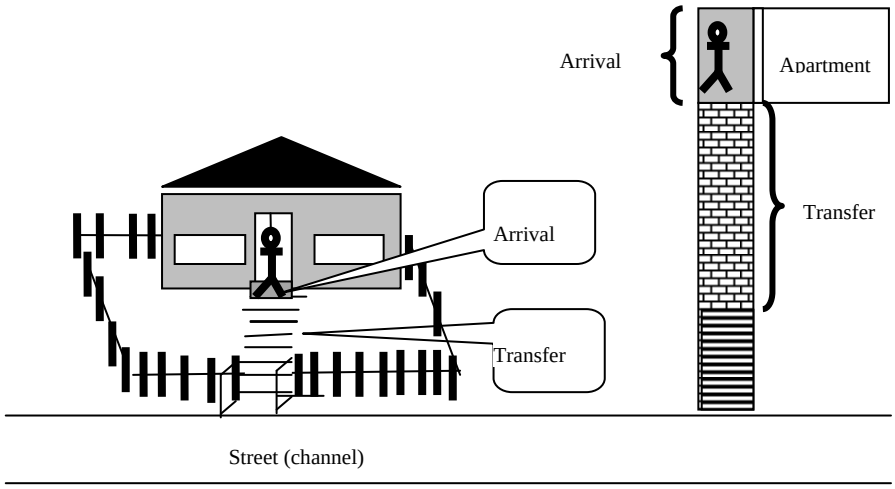

Figure 3. Illustration of Release state

The flowsystem assumes that a flowthing cannot turn around, and the flowthing remains in Release state. Instead of the house in Figure 3 (left), the model is more like that of some New York City apartments, with its own elevator opening to the street (Figure 3, right). If the down button is pushed, changing the state from released to transfer, the elevator cannot turn around before reaching the street.

This conceptualization is necessary for conceiving of flowsystems. In computer communication, when a packet reaches the port (part of the transfer module), say, of a router, it flows from the port to the input buffer, without the possibility of turning around to return to the communication channel. In a program, when transferring input to store, there is no possibility of reversing the direction of flow in the middle of transfer. Moving from one state to the next state in FM is determined by the flowsystem.

Acceptance means entering the system. Each system has an interface with the channel where flowthings arrive; nevertheless, Arrival does not guarantee acceptance. A Created flowthing means it is generated internally as when a packet is "manufactured" to be transmitted. Creation is another source of flowthings that flow in the system along with flowthings arriving from other systems. If a flowthing has arrived and been accepted by a system, it is then not possible to be in the state of being created internally.

The triggering system

Triggering in FM activates events and flows. For example, processing of a customer's order in the orders flowsystem may trigger creation of an invoice in the invoices flowsystem. Thus, one function of triggering is to "connect" different types of flows, analogous to the way in which creation of an electrical signal triggers a water heater to start heating water. A basic principle in FM is not mixing flowthings, but to keep them in special flowsystems. Another function of triggering is to provide a mechanism to model the effect of flow. For example, when processing logical statements, the effect is to create new statements. Thus, the processed statements do not flow to the creation stage; however, the effect of processing is creation.

Figure 4 is a diagram of various types of triggering. Consider a flowthing reaching the Arrival stage. This may trigger (see A in the figure) the flowsystem of notification to create a notice and send it to management. Management may trigger the arrived stage to let the thing flow to the accepted stage or to reject it. The decision of acceptance or rejection may also be made upon Arrival without referring to management. It is also possible that knowledge of the arrival of a new thing triggers Accept, Process, Create, or Release to "finish" things already being made or in process, in order to allow the newly arrived thing to progress in the system.

Spheres

In FM, the environment or context of information is called a sphere. Spheres can have subspheres, and they can be part of an information system, an organization, a human person, etc. Consider the following example given in DHS PIA:

> Example: The *system* will generate a response that there is a possible match to the terrorist screening database. This possible match will be maintained in the system with the

information previously provided by the individual. A trained analyst will review the possible match and make a determination as to whether or not the individual is on the list. This determination will also be maintained in the system.

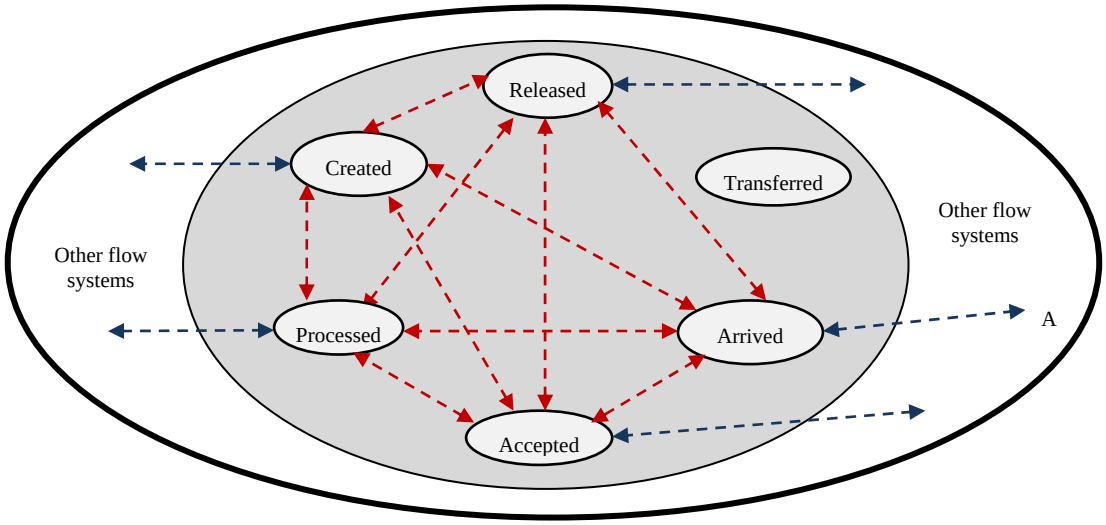

Figure 4. Triggering

There are three spheres in this example: the system, the individual, and the analyst. Figure 5 depicts these spheres and their flowthings. The individual's sphere includes one flowsystem of PII. *Receive* is used to designate both *Arrive* and *Accept* stages. The individual creates (circle 1) his/her PII and this PII flows (circle 2) to the system. The system processes PII and triggers (circle 3) creation of a match/mismatch decision. Notice that a decision is a flowthing that can be created, released, transferred, arrived, accepted, and processed. Both the system decision (circle 4) and PII (circle 5) are sent to the analyst. The analyst processes PII and makes the decision where these processes affect (trigger - circle 6) each other and lead to creation of a confirm/unconfirm decision (circle 7) that flows to the system (circle 8).

In the DHS PIA description of the example, it is not clear what the system does after receiving the analyst's decision, but it is possible to complete the conceptual picture by making the system trigger (alert) authority (fourth sphere) in case of a matching decision. As will be further shown in this paper, the FM description presents a more complete conceptual account of PII-related events than that provided in the DHS PIA written description.

FM-based specifications

FM can be used as a foundation to write different specifications for handling of PII. In this section we select some issues raised in our sample case, DHS PIA, to demonstrate that in addition to defining PII based on the notion of referent, the flow-based approach provides comprehensiveness and depth in describing and understanding operations and constraints related to PII.

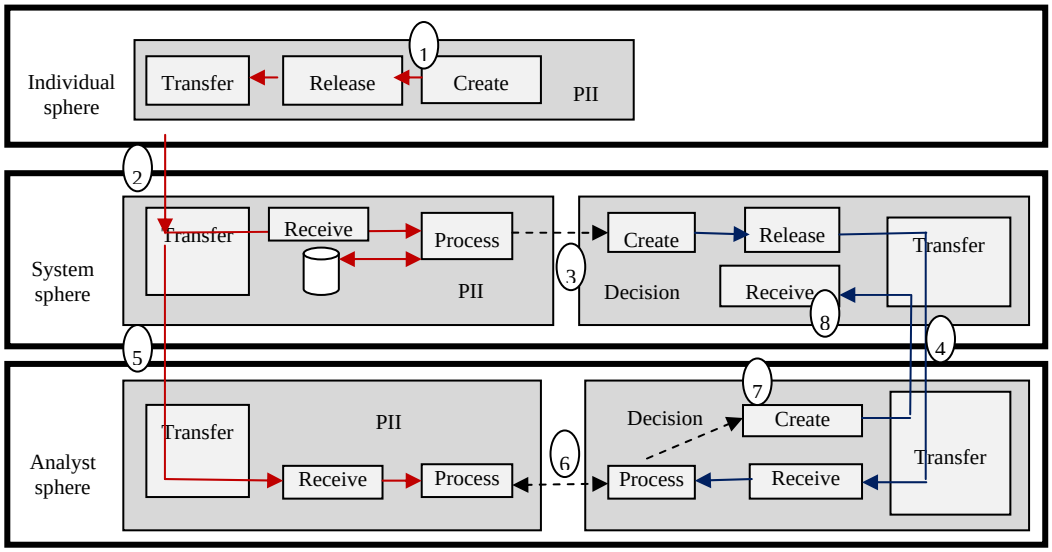

Figure 5. Example involving three spheres

Operations on PII

As discussed previously, in DHS PIA "to define the scope of the information requested and/or collected, as well as reasons for its collection," one must "*2.1 Identify the information the project collects, uses, disseminates, or maintains.*" An alternative version of this is the following:

Identify the arriving, accepted, processed, created, released, and transferred PII in a project.

This version covers all possible types of handling of PII.

Sources of PII

DHS PIA 2.2 deals with the question, "What are the sources of the information and how is the information collected for the project?" This question aims at identifying the sources of PII, which are of two types, internal and external, as shown in Figure 6. There are four sources of PII with respect to a sphere, as follows:

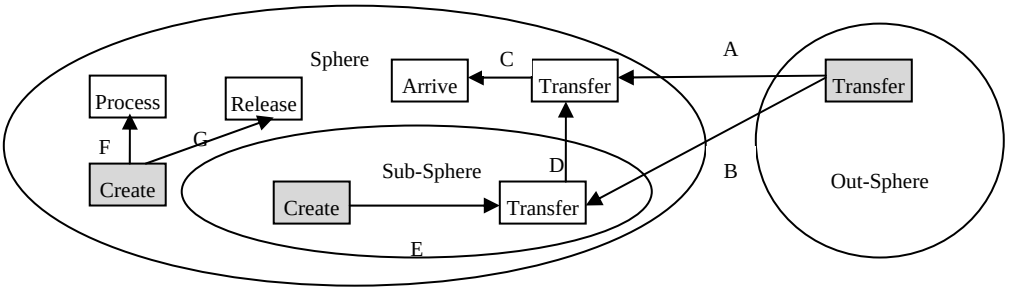

Figure 6. Sources of PII for a sphere

1. PII that is created by the sphere itself, and may flow internally (F and G)
2. PII that is created by a subsphere and transferred to the sphere (E and D)
3. PII that flows from outside (A)
4. PII that flows from outside to a subsphere then to the sphere (B and D)

In such a framework, "sources" of PII are systematically categorized; for example, rules about importing PII from outsiders may be stricter than constraints on PII flow by internal sources.

We notice here that mechanically drawing the flow map identifies sources of PII; thus the first part of *DHS PIA* 2.2, which seeks to answer, "What are the sources of the information and how is the information collected for the project?" is unnecessary if such a map is developed.

On the other hand, identifying sources of PII provides only half the picture. The question *Which constituent of the PII system triggers the sources to "turn on" the flow of PII?* is as important as that of the sources of PII. For example, suppose a hospital releases health PII to a third party with the permission of the patient; then, in such a flow, the source of the flow is the hospital, while the patient's permission is the trigger.

One aspect of the importance of triggering is accountability, with the question, *Who are the requester and the sender?* as significant as the source (origin) of the transaction. With an "opt-in" policy, the company is the source, but the triggering (the thing that "turns on" the flow of PII) is the customer's permission. With opt-out, the source and the trigger are the company itself, with permission of the customer assumed.

The point here is that identifying triggers of flows is as important as recognizing the sources of PII in the context of PIA. Focusing only on sources is analogous to analyzing the economy in terms of supply while ignoring demand. Incorporating the triggering events in Figure 6 requires specification of the flows in the figure. Here we can assume that the sphere, say, that of hospital management, triggers a subsphere (finance department) to prepare (create) the final account statement of a patient. As shown in Figure 7, the finance department triggers an outsider (e.g., insurance company) to release information that is incorporated into the prepared statement that flows to management.

Triggering creates complementary flows to draw a complete picture of activities. In this example, it represents the starting points of flows. In computer programming terminology, it marks the main program and subsequent calls for subroutines.

In FM, such a conceptual map as shown in Figure 7 answers the second part of DHS *PIA* 2.2, *how is the information collected for the project?* The picture is understood by managerial and technical persons; thus, it can provide a tool for communication among them.

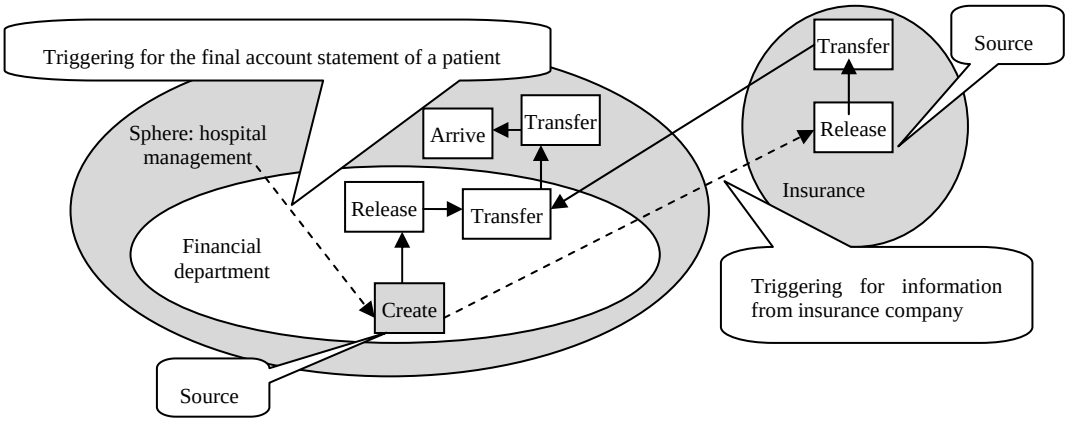

Figure 7. Hospital management activity triggers its financial

Public PII

DHS PIA 2.3 asks, then requires the following: "Does the project use information from commercial sources or publicly available data? If so, explain why this information is used." Here it would be best to develop a deeper understanding of the notion of "publicly available data."

Some information is said to be "documented" in the sense that even though individuals commonly seek to conceal it, nevertheless, it is a matter of public record (e.g., newspaper articles, court records, etc.) and can be known without prying into others' private lives. It is claimed that once personal information becomes public record information, the notion of privacy intrusion is irrelevant with respect to this information (Parent, 1983).

In addition some authors do not consider "public information" about an individual as "private" as non-public information or, in our terminology, PII. Rosen (2001), for example, argues that " when private information is taken out of context, the social judgments that result are more damaging to the individual, and more likely to lead to cognitive errors on the part of society, than the social judgments that result when public information is taken out of context."

The meaning of "information from commercial sources" mentioned in *DHS PIA* 2.3 can be assumed to refer to *legally* collected PII by commercial sources, such as PII for such purposes as insurance and credit cards; however, the question, *Does the project use information from commercial sources or publicly available data?* seems of little value. The issue here is whether the PII in the possession and use of the project is (legally or otherwise) questionable or objectionable.

FM provides generic types of *uses*: possession (by arrival and/or acceptance), processing, creation, releasing, and transferring, as shown in Figure 8. It is not difficult to develop a table that specifies the classification of PII and its use according to this figure.

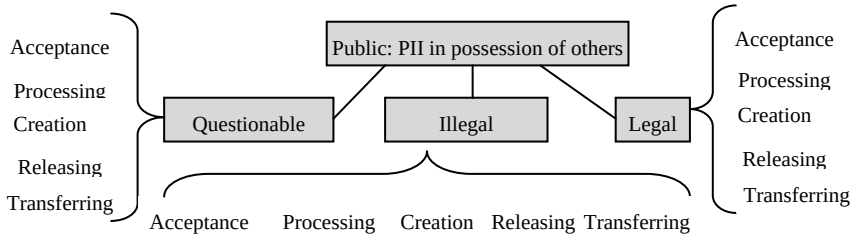

Figure 8. Classification of PII and issues related to PII handling

Consider, for example, *received* information in the *possession* of the organization by arrival or acceptance. For example, PII is in the possession of DHS because it has arrived, but it has not been accepted. Or it has been rejected, but a copy of it is still in the possession of DHS. It is owned by DHS, but it has never been processed, created, released, or transferred. This class of PII can be subcategorized as legal, questionable, or illegal.

We can see that FM furnishes a systematic method of PII classification and subclassification that is applicable to different types of qualifications, making the PIA process more manageable and precise.

Accuracy and misinformation

DHS PIA 4.1 demands discussing "how accuracy of the data is ensured," but what is "accuracy" in the PII context? The classical meaning of accuracy of a measurement is its closeness to its true value. In statistics, data accuracy is generally articulated in terms of a confidence interval. How is this related to PII? It seems that the issue here concerns misinformation.

The concept of Personal Identifiable Misinformation (PIM) is hardly recognized in an explicit manner. PIM is false PII. Since the definition of PII in FM is built upon the correspondence theory, where PII refers to at least one singly identifiable individual, the introduction of PIM expands this reliance on mapping (coupling) to reality (*extension*) to include *truthfulness* along with mapping to the proprietor. *Truth* makes the membership problem (e.g., deciding/identifying PIM) far more difficult. PIM requires the existence of a singly identifiable individual, and that individual *satisfies* PIM. The PII *John is angry* is PIM if John is not angry, assuming John is a singly identifiable individual.

Consider the PII *Mary is honest*. If the statement *Mary is honest* is true, then *Mary is dishonest* is PIM. PIM is a privacy-related matter as much as true PII. The tautology *Mary is honest or dishonest* can be counted as (insensitive) PII; nevertheless, in certain situations, people are very sensitive even about the mere act of mentioning such traits. It may be argued that PIM has all the significant aspects associated with true PII. It can be "sensitive" since as a result of revealing PIM, the proprietor may feel hurt, vulnerable to harm, etc. All characteristics of PII such as identification, sensitivity, secrecy, and publicity can be applied to PIM. Even though it is misinformation, the proprietor may block others from possessing it for fear of being ridiculed, etc.

True PII embeds *double truth*. It is *true* in identifying a person (agreement with referent-part), and it is *true* as an assertion (i.e., agreement with its discourse). Truth means "uncoveredness" and "unhiddenness" that is "wrested away"; thus, individuals "get snatched out of their diddenness… The … uncoveredness of anything is always, as it were, a kind of robbery" (Heidegger, 1962). PIM is *true* in identifying the referent but it is *false* as an assertion.

We conclude that in the context of PIA, the issue of PIM is more important than the vague notion that *DHS PIA* 4.1 calls "accuracy." Developing a mechanism to identify and treat PIM in any organization should be part of its privacy awareness.

Uses and role of PII

DHS PIA section 3, under the title "Uses of the Information," raises questions about *uses* of PII. The problem is that the notion of "use" is not clear. Consider the first question:

3.1 Describe how and why the project uses the information.

We have already mentioned that FM provides generic types of *uses*: possession (by arrival and/or acceptance), processing, creation, releasing, and transferring, as shown in Figure 8. DHS PIA gives the following example of use of information: A project needs to collect name, date of birth, and passport information because that information provides the best matching capabilities against the terrorist screening database.

Simply put, PII is used in a matching process against a database. Here *use* approximately means input. That is, PII is input to a matching process against a database. Consequently, use may refer to the role of PII in the system. If this is true, FM represents a complete conceptualization of the role of PII.

In the DHS PIA example of terrorist screening, there are two types of PII, as follows:

1. PII1 that includes name, date of birth, and passport information
2. PII2 that includes terrorist's PII

The system may also handle non-PII information that describes general characteristics of a terrorist. Figure 9 shows the role of PII in such a scenario. There are two flowsystems feeding the information flowsystem that handles the matching process to produce match/non-match. PII1 and PII2 flowsystems receive PII from sources that are not mentioned; they may process, create additional PII, and send PII to the main system (of the project).

The main system processes incoming PII in combination with other non-PII information to produce a match/no-match decision. This specification raises two privacy issues: the source of both types of PII, and the logic and data (non-PII) of the main process that single out a "hit." It magnifies and focuses on such privacy constraints instead of on the DHS PIA's general description of a "use" of information: A project needs to collect PII "because that information provides the best matching capabilities against the terrorist screening database."

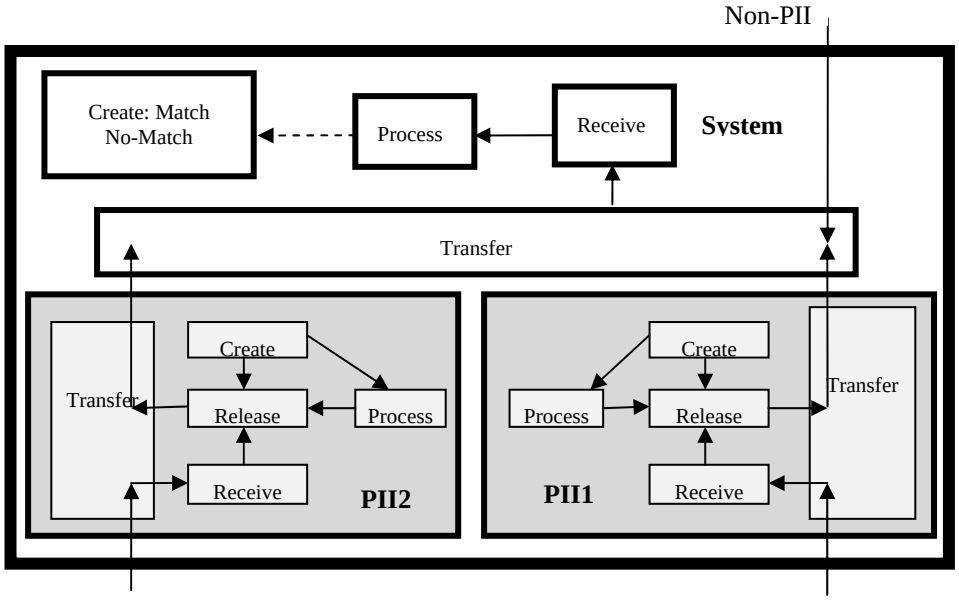

Figure 9. Description of the role of PII

Conclusion

Compliance with privacy regulations and laws presents a significant challenge for organizations holding data about identifiable persons. Privacy impact assessment is one of the most important instruments used to ensure conformity to regulatory requirements, determine risks, and evaluate privacy protections. This paper scrutinized the 2010 US Department of Homeland Security's guidance on privacy PIA as a representative case of current approaches to PIA. Three main concepts are introduced as a base for building PIA:

1. Personal identifiable information, PII, is defined as information that contains a *referent* to uniquely identifiable persons. Every PII *refers* to its proprietor(s) in the sense that it "leads to" him/her/them as distinguishable entities in the world.
2. PII is classified into atomic (one referent) or compound (multi-referent), and it is possible to reduce compound PII to atomic PII.
3. A flow-based model is introduced as a foundation for the specification of PII handling. It is represented as a flow system with six exclusive stages (operations): arrival, acceptance, processing, creation, release, and transfer.

This foundation is utilized in reformulating and modeling in defining PII, handling PII, sources of PII, public PII, accuracy and misinformation in PII, and uses of PII. The paper has demonstrated that this methodology can provide a foundation for PIA. Further research would explore applying it to other privacy related topics such as drafting of laws, and tools such as privacy enhanced information systems. Several issues such as handling of compound PII need further exploration.

References

Abdul Ghani, N., & Sidek, Z. M. (2009). Controlling and disclosing your personal Information. Transactions on Information Science & Applications 6(3).

Al-Fedaghi, S. (2005). How to calculate the information privacy. Proceedings of the Third Annual Conference on Privacy, Security and Trust (PST 2005), October 12-14, 2005, St. Andrews, New Brunswick, Canada. http://www.lib.unb.ca/Texts/PST/2005/

Al-Fedaghi, S. (2005). Privacy as a base for confidentiality. Paper presented at the Fourth Workshop on the Economics of Information Security, Harvard University, Cambridge, MA. http://infosecon.net/workshop/pdf/4.pdf

Al-Fedaghi, S. (2005). The right to be let alone and private information. Proceedings of the 7th International Conference on Enterprise Information Systems, Miami, USA, 2005; also published in C.-S. Chen, J. Filipe, I. Seruca, & J. Cordeiro (Eds.), Enterprise Information Systems VII, Springer). http://www.springerlink.com/content/p2j60k45u8614183/

Al-Fedaghi, S. (2006). Crossing privacy, information, and ethics. Paper presented at 17th International Conference, Information Resources Management Association, Washington, DC.

Al-Fedaghi, S. (2007). Personal information ethics. In M. Quigley (Ed.), Encyclopedia of Information Ethics and Security. Hershey, PA: Information Science Publishing.

Al-Fedaghi, S. (2007). How sensitive is your personal information? The 22nd ACM Symposium on Applied Computing (ACM SAC 2007), Seoul, Korea.

Al-Fedaghi, S. (2009). The ethics of information: What is valued most. *The Open Ethics Journal* 3. Retrieved from http://www.bentham.org/open/toj/articles/V003/118TOJ.pdf

Al-Fedaghi, S. (2011). Toward a unifying view of personal identifiable information. Paper presented at *4th International Conference on Computers, Privacy, and Data Protection*, Brussels, Belgium.

Al-Fedaghi, S., & Al-Shiridah, G. A. (2010). Conceptual foundation for personal information management. *International Journal of Electronics, Information and Systems*, *12*(2).

Al-Fedaghi, S., & Deabas, M. (2006). Systematic approach to personal information impact assessment. Paper presented at *International Conference on the Digital Information Industry - Information Privacy Key Issues and Best Practices* (DII 2006), Seoul, Korea.

Al-Fedaghi, S., Fiedler, G., & Thalheim, B. (2006). Privacy enhanced information systems. Proceedings of *the 15th European-Japanese Conference on Information Modeling and Knowledge Bases*: Tallinn, Estonia, 2005. (Also published in Y, Kiyoki, J. Henno, H. Jaakkola, & H. Kangassalo (Eds.), Information Modelling and Knowledge Bases XVII, Frontiers in Artificial Intelligence and Applications 136, 2006).

Al-Fedaghi, S., & Thalheim, B. (2008). Databases of personal identifiable information. Paper presented at *IEEE/ACM/IFIP SITIS'08 Workshop on Security and Privacy in Telecommunications and Information Systems*, Bali, Indonesia.

Anwar, M. (2008). An identity- and trust-based computational model for privacy. PhD thesis, Department of Computer Science, University of Saskatchewan, Saskatoon, Canada. http://library2.usask.ca/theses/available/etd-01022009-184502/unrestricted/final-v5-thesis.pdf

DHS. (2010, June). "Privacy impact assessments official guidance". United States Department of Homeland Security, Privacy Office. Retrieved from http://www.dhs.gov/xlibrary/assets/privacy/privacy_pia_guidance_june2010.pdf

EDRI-Gram Newsletter. (2010). Industry proposed RFID privacy impact assessment framework, Number 8.10, 19 May.

E-Government Act (2002). Section 208. Retrieved from http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=107_cong_public_laws&docid=f:publ347.107

EU Data Protection Directive. (1995). The 95/46/EC Act of the European Parliament and of the Council of 24 October 1995. Retrieved from http://www.cdt.org/privacy/eudirective/EU_Directive_.html

Heidegger, M. (1962). *Being and time*. London: ScM Press Ltd.

Kang, J. (1998). Information privacy in cyberspace transactions. 50 *Stanford Law Review*, 1193:1212-20.

Lederer, S., Beckmann, C., & Dey, A. (2003). Managing personal information disclosure in ubiquitous computing environments. IRB-TR-03-015. Retrieved from http://www.intel-research.net/Publications/Berkeley/070920030922_139.pdf

Narayanan, A. (2010, June 21). Myths and fallacies of personally identifiable information. Retrieved from http://33bits.org/2010/06/21/myths-and-fallacies-of-personally-identifiable-information/

Narayanan, A., & Shmatikov, V. (2010). Myths and fallacies of personally identifiable information. *Communications of the ACM* 53.

OPS-NOC (2010, April). BP oil spill response: Social media event monitoring initiative. Office of Operations Coordination and Planning and National Operations Center. Retrieved from http://www.dhs.gov/xlibrary/assets/privacy/privacy_pia_ops_bpoilspill.pdf

Parent, W. A. (1983). Privacy, morality, and the law. *Philosophy and Public Affairs*, *12*(4).

Privacy Commissioner and Department of Justice. (2000, Aug.). Privacy impact assessment for justice information systems. A working paper jointly produced by the Office of the Information and Privacy Commissioner/Ontario and the United States Department of Justice, Office of Justice Programs. Retrieved from http://www.ipc.on.ca/english/Resources/Discussion-Papers/Discussion-Papers-Summary/?id=326

Rosen, J. (2001, Spring). Out of context: The purposes of privacy. In G. Kateb, J. Rosen, & F. Schauer *Part V: Invasions of Privacy: Violations of Boundaries*. Social Research. Retrieved from http://www.findarticles.com/cf_dls/m2267/1_68/75658589/p3/article.jhtml?term=

Sato, K., Izumi, S., & Kato, Y. (2009, Aug.). Privacy-based personal and group information modeling in Semantic Web. Paper presented at *13th IASTED International Conference on Internet and Multimedia Systems and Applications*, Honolulu, Hawaii. Retrieved from http://hirose.sendai-nct.ac.jp/~kaoru/pub/Kana09-2.pdf

Shroff, M. (2007, June). "Privacy impact assessment handbook". Office of the Privacy Commissioner. Retrieved from http://privacy.org.nz/assets/Files/Brochures-and-pamphlets-and-pubs/48638065.pdf

TERS. (2010, Sept.). Travel Expense Reporting System TERS Privacy Impact Assessment (PIA), version 1.2. Retrieved from http://www.occ.treas.gov/policies/ters-pia-2010.pdf

US Office of Personnel Management, OPM. (2010). Retrieved from https://www.opm.gov/privacy/PIAs/PIAGuide.pdf

---

[1] Dr. Sabah Al-Fedaghi is an Associate Professor in the Computer Engineering Department, Kuwait University. His research interests include database systems, natural language processing, information systems, information privacy & security, and information ethics. Email: sabah@alfedaghi.com.

[2] Mr. Abdilhadi Jeragh is an IT Systems Support Engineer at the Kuwait Petroleum International LTD. Email: abdilhadi@jeragh.com.