J. Software Engineering & Applications, **2009, 2: 67-76**
doi:10.4236/jsea.2009.22011 Published Online July 2009 (www.SciRP.org/journal/jsea)

Interpretation of Information Processing Regulations

Sabah Al-Fedaghi

Computer Engineering Department, Kuwait University, Kuwait City, Kuwait.
Email: sabah@eng.kuniv.edu.kw

Received February 2nd, 2009; revised April 14th, 2009; accepted April 15th, 2009.

ABSTRACT

Laws and policies impose many information handling requirements on business practices. Compliance with such regulations requires identification of conflicting interpretations of regulatory conditions. Current software engineering methods extract software requirements by converting legal text into semiformal constraints and rules. In this paper we complement these methods with a state-based model that includes all possibilities of information flow. We show that such a model provides a foundation for the interpretation process.

***Keywords*:** *Software Requirement, Laws, Regulation, Privacy, Personal Identifiable Information*

1. Introduction

Laws, regulations, and policies such as the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule and the Telecommunications Act of 1996 impose many requirements on business practices for handling information. In 2006, 161 billion gigabytes of digital information were created, captured, and replicated [1]. It is estimated "that today, 20% of the digital universe is subject to compliance rules and standards, and about 30% is potentially subject to security applications" [1]. In 2005, more than 20,000 regulations were passed related to creation, storage, access, maintenance, and retention of information [2].

Compliance with these laws, regulations, and policies requires identification of conflicting interpretations of regulatory requirements. The information system needs to be aligned with legal and regulatory requirements in order to be in compliance.

Statements of regulations in legal documents relevant to information processing contain a great deal of natural language ambiguity that makes it difficult to formalize requirements and constraints in software systems. The basic problem can be viewed as how to extract software requirements from regulations.

Researchers have introduced different methods for converting legal language into semiformal specifications; nevertheless, the approaches to interpreting legal text lack compatibility with the software-engineering style of problem solving. A need exists for an underlying information-processing model of the different information processing systems, similar to the classical communication model of sender, receiver, and message. Terms such as "collecting," "processing," "disclosing," and so forth are used loosely, without a pattern tying them together as actions based on information. We will demonstrate such aspects in an example after introducing our model.

So, what is the software-engineering style of problem solving that ought to be applied to interpretation of regulations to meet software requirements? It involves simply constructing an information flow model, and taking into account all possible types of actions utilized in processing information. While it is not practical to take into account every possible interpretation, we propose a state-based model that includes a limited number of possibilities for software responses to all categories of information handling.

2. Related Work

The software engineering field is rich with work related to software requirements for domain and systems descriptions. Methods have been proposed to extract requirements from policies and regulations using formal models [3], semantic parameterization [4], and ontology [5]. Several publications deal with the problem of extracting goals from natural language documents and Internet privacy policies [6,7]. Breaux and Antón [4,8] developed a method to trace the words in regulations to semantic primitives. Giorgini et al. [9] described a framework that enables modeling of actors and goals and their relationships. May et al. [3] introduced a methodology to extract formal models from regulations and applied it to the HIPAA Privacy Rule.

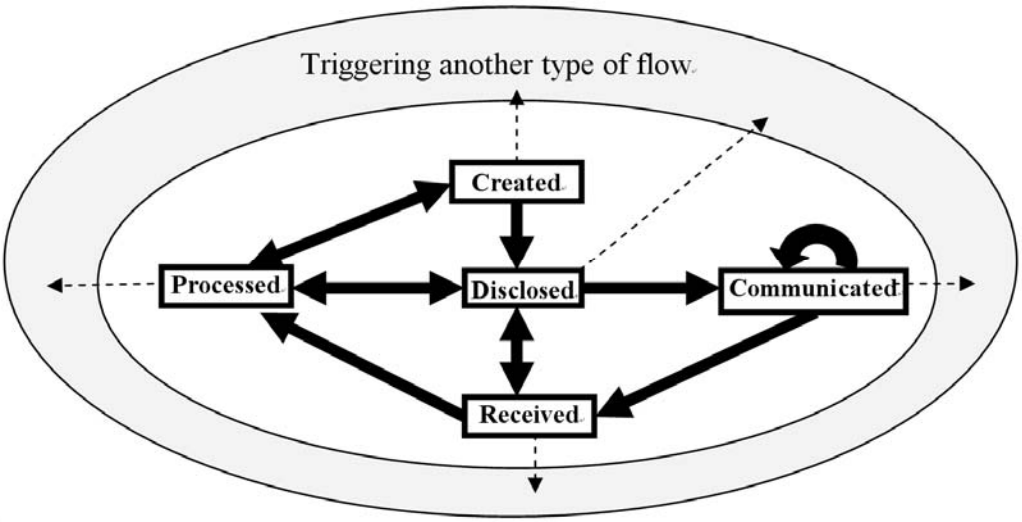

Figure 1. Information states in FM with the possibility of triggering another type of flow

We will concentrate on the recent methodology of Breaux and Antón [10] "to extract access rights and obligations directly from regulation texts . . . [and] present results from applying this methodology to the entire regulation text of the U.S. Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule." In our case, we will discuss a very limited portion of such a commendable venture.

Information processing models have evolved from the classic 1949 Shannon-Weaver communication theory of transference of electrical signals from sender to receiver. Nevertheless, the main aspect of Shannon and Weaver's model is communication, while information appears as a secondary component of the model. The "transmitted materials" in this communication can be information, data, meaningless signals, energy, or, with proper perspective, other entities.

We are interested more in information *properties* than in the communication act itself. Such an approach concentrates solely on information, while communication aspects (as important as they are) become secondary features of the flow of information. This paper introduces an information flow model that includes five stages: collection, processing, creation, disclosure, and communication. Such a model provides a better framework on which to base interpretation of information handling processes.

3. Model of Information Flow

The flow model (FM) has been proposed and used in several applications. In FM, the flow of flowThings indicates movement inside and between spheres. The sphere is the environment of the flow and includes five stages that may be sub-spheres, each with its own five-stage schema. The stages may be named differently; for example, in an information sphere, a stage may be called *communication*, while in action flow, the same stage is called *transferring*.

To illustrate the notion of flowThing, we will assume that the "thing that flows" is information. We use the term *information* to refer to information and misinformation, as in common usage where a reporting statement can be true or false. An information sphere denotes the information environment. The lifecycle of information is a sequence of states in its lifecycle, as follows: 1) *Received*, 2) *Processed* (in a way that changes its form, but not content), 3) *Released*, 4) *Communicated* (to another sphere), and 5) *Created* (i.e., a new piece of information). These five states of information form the main stages of the stream of flow, as illustrated in Figure 1. Each stage may include sub-stages, such as storage and usage.

The states shown in Figure 1 are exclusive in the sense that if information is in one state, then it is not in any of the other four states. Consider a piece of information, σ, possessed by a hospital; σ is thus in one of the following states:

1) σ has just been collected from some source (patient, friend, sent by some agency) and stored in the hospital record, waiting to be used. It is *collected* (row) information that has not yet been processed by the hospital.

2) σ has been processed in some way, converted to another form (e.g., digital), translated, compressed, etc. Also, it may be stored in the hospital information system as *processed* data waiting for some use.

3) σ has actually been created in the hospital as the result of doctor's diagnoses, lab tests, etc. Thus, σ is in the possession of the hospital as *created* data to be used.

4) σ is being released from the hospital infosphere. It is designated *disclosed* information ready for transfer. Analogous to a factory environment, σ represents materials ready to be shipped outside the factory. It may actu-

ally be stored for some period waiting to be transported; nevertheless, its designation as "for export" keeps it in such a state.

5) σ is in a transfer state, being transferred between two infospheres. It has left the disclosure state and will enter the collection state, where it will become collected information in the new infosphere.

It is not possible for processed information to directly become collected information in the same infosphere. Processed information can become collected information in another infosphere by first becoming disclosed information, and then transferred information, in order to arrive at the other environment.

We use this model, called the information flow model (IFM), to classify information in generic theoretical categories that can be applied in any infosphere, including laws and regulations. According to Kurt Lewin, "There is nothing quite as practical as a good theory."

Consider the following extract from the Safety Officer's Briefing Book of the United States Air Force Auxiliary [11]:

Water *vapor* is an invisible *gas*, similar to nitrogen and oxygen, the two gases which make up 98% of our atmosphere. We see clouds and fog when the *gaseous* water *vapor* is cooled sufficiently to allow a change of state from *gas* to *liquid*. We see snow, sleet, and hail when the *liquid* water is further cooled to a *solid* state. And when water changes directly from *gas* to *solid* it becomes frost, through a process called "sublimation." [Italics added.]

Imagine this to be a statement of regulations with no model of water circulation among its states of liquid (water), gas (vapor, fog), and solid (ice). We could probably manage to write safety regulations using terms such as "clouds" and "frost," as in the following rewrite of the previous quote:

Water is an invisible material, similar to nitrogen and oxygen, the two gases which make up 98% of our atmosphere. We see clouds and fog when the water is cooled sufficiently. We see snow, sleet, and hail when the water is further cooled. And when water changes, it becomes frost, through a process called "sublimation."

A great deal of conceptual clarity is lost without specification of the three natural states of water in a model of its circulation among those states. Similarly, information handling regulations not based on the information flow model are simply vague descriptions of the proper way to handle information.

As another example, consider applying the IFM to HIPAA Privacy Rule [12]. According to the Privacy Rule, "A covered entity may *use or disclose*, without an individual's authorization, the psychotherapy notes . . ." [13]. We show that, in the context of IFM, *use* and *disclose* may be interpreted in several ways.

Psychotherapy notes are defined in the Rule as:

> . . . notes recorded (in any medium) by a health care provider who is a mental health professional documenting or analyzing . . . that are *separated from the rest of the individual's medical record*. Psychotherapy notes exclude medication prescription and monitoring, counseling session start and stop times, the modalities and frequencies of treatment furnished, results of clinical tests, and any summary of the following items: diagnosis, functional status, the treatment plan, symptoms, prognosis, and progress to date.

A psychotherapist collects information from four sources:

1) The original medical record

2) Information created by the psychotherapist during a session

3) New information revealed by the patient during a session

4) Created information related to the session (e.g., start and stop times: *John's session is at* 4 *pm*, *John arrived late for the session*).

These clear categorizations of information can be supplemented with the type of descriptions given by the rule (e.g., summary of this type or that type). Each category may require different constraints on its use and disclosure. For example, releasing information of type (2) may need the approval of the psychotherapist who wrote it, while the decision to release information of category (1) has nothing to do with the psychotherapist.

The point here is that the IFM allows more refined interpretation of natural language description involving operations on different types of information, such as disclosure. The example above includes at least three types of disclosures: disclosure of collected information, disclosure of created information, and disclosure of processed information. Each type may need different disclosure constraints [14].

4. Personal Identifying Information

This paper focuses on the HIPAA Privacy Rule; for that purpose we regard personal identifiable information (PII) as the main type of information and object of study [12]. The "circulation of PII" can be modeled in IFM as information flow.

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 agreements (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 "private" or "personal" information. Personal identifiable information has an "objective" definition in the sense that it is independent of such authorities as legislation or agreement.

4.1 Definition of Personal Identifiable Information

In the information sphere, information is classified as personal identifiable information (PII) and non-identifiable information (NII). Personal identifiable information is information *about* singly identifiable persons, called *proprietors*. PII is information that has *referents* who are natural persons. Two types of PII can be identified:

1) Atomic personal information, where the information refers to a single proprietor, e.g., *John is 25 years old*. "Referent" here implies an identifiable (natural) person.

2) Compound personal information, where the information refers to more than one proprietor, e.g., *Mary donated her kidney to Alice*.

Any compound PII is privacy reducible to a set of atomic PIIs [12].

4.2 Elaborated IFM Model for PII

"Handling information" involves observing the progress of information from its arrival at the infosphere through the various information stages until it exits, or disappears from, the "information circulation system." "Pieces of information" circulating in the IFM (Figure 2) are envisioned here as "informational objects" that have an existence within the information realm. This type of information, the so-called "meme," is "a hypothetical unit of cultural transmission conceived not as an inert object but as a quasi-organic entity endowed with the capacity of self-replication" [4].

The flow model makes a piece of information visible as soon as it enters the circulation system of IFM. In most cases, the piece of information then moves repeatedly among the stages of the model. IFM includes types of acts on information (labelled 1 through 14 in Figure 2) that transform information from one stage or sub-stage to another.

Creation stage: In Figure 2, the creation stage includes two attached sub-stages: *Storage* and *Actions*. New information is created (e.g., data mining generates new information). The created information is utilized in some way (called sub-stage actions in Figure 2; e.g., decision making), stored, or immediately moved on to the

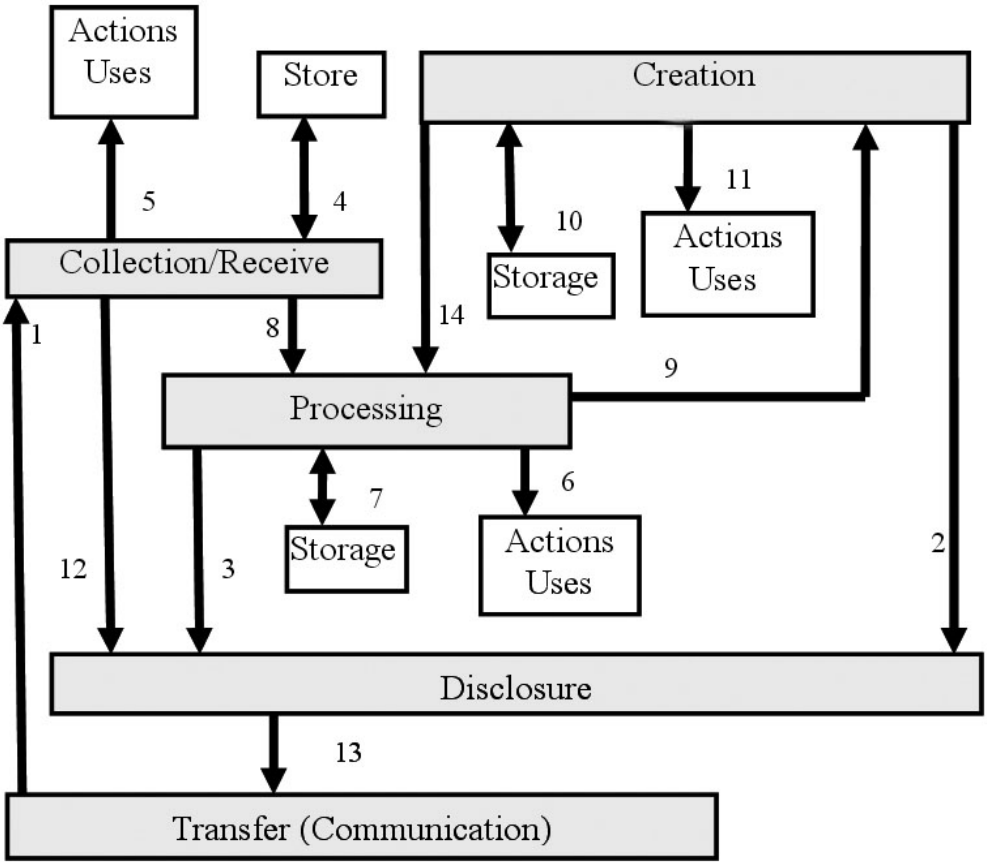

Figure 2. The Information Flow Model (IFM)

stages of processing or departure (disclosure and transfer). In IFM, "action" refers to any utilization of information.

For example, suppose a physician makes a new diagnosis of a disease (creates new information), such as *John Smith has AIDS*. Such new information may be generated by examining (mining) some data. *John Smith has AIDS* is then communicated to an expert (disclosed and transferred), or processed further (cycled through processing-mining-creation-processing) to verify the medical judgment. The physician *creates* (act 9) new information and also *uses* the information in the patient's file for "treatment" (physical action). Another example of *uses* in this context is "home delivery of medicine to a patient" (physical action) that *uses* the address information.

Collection stage: The collection stage is the information acquisition stage, when information is accepted from external suppliers and fed into the IFM circulation system. This stage includes the possibility of *using* the arriving (raw) information; the sub-stage, *Actions*, in IFM is therefore a way for information to exit the system (i.e., arriving information generates the action of physical treatment, without further propagation in IFM). It also includes the possibility of storing the collected information.

Processing stage: The processing stage involves acting on information (e.g., anonymization, data mining, summarizing). Processing is performed on acquired information from the collection stage or the creation stage; see Figure 2. In actual processing, information is modified in form or content.

IFM distinguishes between two types of processing: ordinary processing, which does not produce new or inferred information, and creation of new information (e.g., by mining). An example of creation of new information is the categorization of diverse information used to reach a decision, e.g., *John is a risk*. Other types of processing that do not generate new information, but only change the appearance of information, include comparing, compressing, translating, and deleting.

Disclosure and transfer/communication stages: The disclosure stage involves releasing information to those outside the infosphere. It relies on the transfer stage to carry information from the current infosphere to the collection stage in another infosphere. When information is in the transfer state, it is flowing (e.g., through a channel) between two infospheres.

5. Application to Extraction Constraints

Healthcare regulations, such as HIPAA, are often difficult to grasp and interpret. Interpretation of HIPAA regulations is of utmost importance because of the costs associated with any misinterpretation. According to the U.S. Department of Health and Human Services., the healthcare industry will spend $17.6 billion over the next few years to bring their systems and procedures into compliance with HIPAA [15].

Consequently, we limit this paper to applying the IFM model to the problem of extracting software requirements from HIPPA regulations; however, this focus does not mean the model cannot be applied to any other area with information-handling regulations. Additionally, our treatment is obviously incomplete in the sense that it is not a rewriting of the entire regulation, since it aims mainly to present a general methodology for interpreting informational privacy regulations.

We will concentrate specifically on methodology recently introduced by Breaux and Antón [10] for extracting access rights and obligations directly from the text of regulations. According to Breaux and Antón [10], "'rules' are often precursors to software requirements that must undergo considerable refinement and analysis before they are implementable." Breaux and Antón [10] then present "a methodology to extract access rights and obligations directly from regulation texts . . . to support the software engineering effort to derive security requirements from regulations." The methodology will then "identify and infer six types of data access constraints, handle complex cross-references, resolve ambiguities, and assign required priorities between access rights and obligations to avoid unlawful information disclosures."

The following excerpt from Privacy Rule §164.510 (b)(1)(i) will be used to illustrate use of IFM to develop a complete picture of situations specified by the Rule.

A CE [*Covered Entity*] *may, in accordance with paragraphs* (*b*)(2) *or* (3) *of this section, disclose to a family member, other relative, or a close personal friend of the individual, or any other person identified by the individual, the PHI* [*Protected Health Information*] *directly relevant to such person's involvement with the individual's care or payment related to the individual's healthcare*.

We discuss some of the constraints extracted by Breaux and Antón [10] to exemplify ambiguities uncovered by the definitions of PII and IFM.

Constraint 1: The PHI is directly relevant to the person's involvement in the individual's care.

Comments: We should explicitly declare whether the atomic PII is collected PII, processed PII, or created PII. For example, the proprietor may agree to release his or her telephone number (collected information), but not the results of his or her lab examination (created PII), or his or her doctor's diagnosis (created PII).

Notice that the point here is not *constraints* on the scope of information (e.g., address and phone number); it

is, rather, the *right* of the proprietor to be presented with all alternative actions to which he or she can consent, and the *obligations* of the stakeholder to provide the proprietor with complete information for making informed consent. The choices here are as follows:

- Disclosing only PII collected by the stakeholder from other sources
- Disclosing PII processed by the stakeholder, such as implied PII, translated PII, anonymized PII, etc.
- Disclosing created PII by the stakeholder, such as lab results, doctors' diagnoses, etc.

These actions deal with categories of PII and not specific PIIs; thus, it is practical to list this small number of categories of which the proprietor approves disclosure. Such classification of PII is also suitable for software design and implementation in a decision-making system.

In general, the Privacy Rule permits uses and disclosures for "treatment, payment and health care operations," as well as certain other disclosures, without the individual's prior written authorization. What we propose is that such permission be categorized by the system according to the type of PII, for example,

- *Collected* PII uses and disclosures for "treatment, payment and health care operations" has a different access policy than
- *Created* PII uses and disclosures for "treatment, payment and health care operations."

For example, "uses and disclosures" of created PII may need the approval of its creator (e.g., the physician).

Constraint 2: PHI is directly relevant to the person's involvement in payment related to that person's healthcare.

Comment: PII "directly relevant to the person's involvement in payment related to the individual's healthcare" may also be collected PII, processed PII, or created PII. For example, disclosing such information may not include a hospital's own evaluation of the proprietor's financial reliability. It is similar to releasing "information I have" (collected PII), but not "information I have inferred" (created PII).

Constraint 3: The use is to carry out treatment, payment, or healthcare operations.

Comments: The "uses" of collected PII, processed PII, and created PII are different. In particular, created information carries different types of responsibilities from those of collected PII that is then disclosed. The disclosure that (*According to our diagnoses*) *John Smith is an AIDS patient* (collected PII) has a different weight than the disclosure that *John Smith is an ABC employee* (collected PII) even though both disclosures are for the same use.

We conclude that IFM can enhance such a methodology for extracting constraints from regulation texts. IFM simply reveals all paths of information circulation; thus, we can take into account all possible interpretations in the text. Next, we apply IFM to the process called "semantic parameterization," proposed by Breaux and Antón [10].

6. Enhanced Semantic Parameterization

According to Breaux and Antón [10], semantic parameterization is a process to "support engineers who map natural language domain descriptions to models expressed in first-order predicate logic for the purpose of performing automated reasoning and analysis" [16]. It provides:

1) A reference system that systematically detects and resolves such ambiguities.

2) Automated support for placing natural language-like inquiries across collections of requirements that answer who, what, where, when, how, and why questions.

3) A means to formalize and compare different stakeholder viewpoints.

Breaux and Antón [10] consider the following "properties" of information to be access–related activities:

a) The *subject* is the actor who performs an action on an object in the activity. In the IFM, the subject is one of the entities that acts on PII.

b) The *action* is the verb that affects information, such as access, use, disclose, etc. The IFM limits the number of acts on PII that can be applied here.

c) The *modality* modifies the action by denoting the action as a right, obligation, or refrainment. Such a "modality" can be added after the backbone of the information handling processes is designed, as described next.

d) The *object* is limited to information, including the name or date of birth of a patient, or an accounting of disclosures. In IFM, information is limited to PII and includes all acts and uses of information defined by the model in Figure 2.

e) The *target* is the recipient in a transaction, such as the recipient of a disclosure. The target in IFM is one of the agents in PII transactions. There may be several agents, as illustrated in Figure 3. It is also possible that both sides of the transaction belong to the same organization.

In the IFM, we can identify all entities that may be involved in handling PII in order to accomplish the same thing that Breaux and Antón [10] extract from legal wording. IFM gives a complete picture that includes all entities and processes needed to produce a design description of the software, and no more. For example, a policy for access-related activities can be specified according to the type of entity (e.g., a *collector*—a clerk who fills patient data cannot access *processed* information).

Thus, the software engineer can design a system that takes into consideration all cases of information flow. The entities that may be involved in handling PII are several, including proprietor, possessor, collector, processor, different types of users, different types of storekeepers, (data) miner, creator, releaser, and transferor.

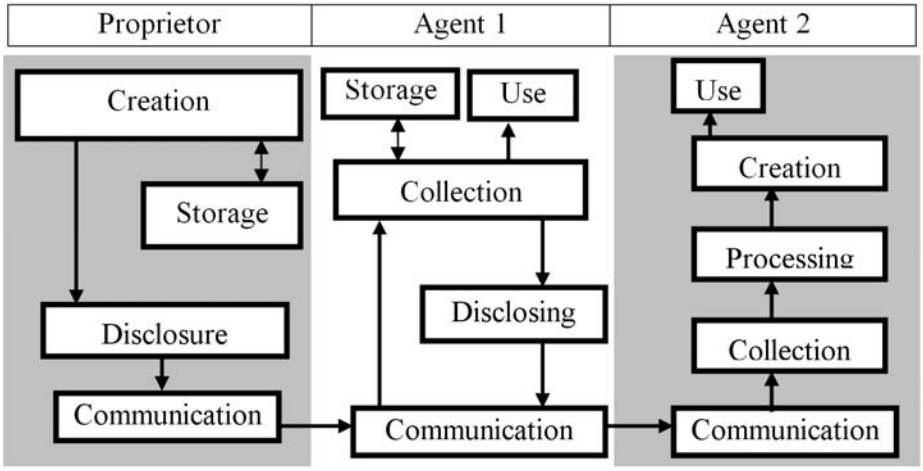

Figure 3. PII flow involving a proprietor and two agents

Consider the following example from Breaux and Antón [10]:

Excerpt from Privacy Rule §164.522(a)(1):

(*i*) *A CE must permit an individual to request that the CE restrict*:

(*A*) *Uses or disclosures of PHI about the individual to carry out treatment, payment or healthcare operations*; *and*

(*B*) *Disclosures permitted under §*164.510(*b*)

(*ii*) *A CE is not required to agree to a restriction.*

(*iii*) *A CE that agrees to a restriction under paragraph* (*a*)(*1*)(*i*) *of this section may not use or disclose PHI in violation of such restriction, except that, if the individual who requested the restriction is in and the restricted PHI is needed to provide emergency treatment, the CE may use the restricted PHI, or may disclose such information to a HCP, to provide such treatment to the individual.*

Applying the Breaux and Antón [10] method to this excerpt yields constraints and rules, some of which are as follows:

- The *use* is to carry out treatment, payment, or healthcare operations.

- The *disclosure* is to carry out treatment, payment, or healthcare operations.

- The CE agrees to a restriction.

- The individual requested the restriction.

- The individual is in need of emergency treatment.

- The PHI is needed to provide emergency treatment.

- A CE must permit an individual to request a restriction.

- A CE is not required to agree to a restriction.

A software engineer then applies *patterns* to identify rights, obligations, and constraints. For example, "the *basic activity pattern* describes a subject who performs an action on an object and *modality* distinguishes the activity as a right, obligation or refrainment. Each rule uses these two patterns to ensure that the statement has precisely one subject, action, object and modality" [10].

The IFM introduces an intermediate stage of analysis that provides a foundation for the legal text. The excerpt from Privacy Rule §164.522(a)(1) specifies relationships between a CE and a proprietor: use and disclosure, as described next.

6.1 The Use Relationship

Figure 4 shows the *Use* relationship embedded in the excerpt. The CE is a possessor of PII about the proprietor. The CE may be a collector, processor, creator, storekeeper, or (data) miner of this PII, or all or some of these roles. In all cases, the relationship between the CE and proprietor leads to some *use* (actions in IFM). The *uses* are of two types:

- Use to "carry out treatment, payment, or healthcare operations"

- Use for "need of emergency treatment" (dotted lines in Figure 4).

The software engineer can represent all possible cases in this situation. Thus, he or she can put constraints and rules on the *use* of PII generated by the CE that are different from constraints and rules on *use* of PII collected from outsiders, including the proprietor him/herself. If PII, e.g., a bank account number, is given by the proprietor, then obviously there are no restrictions on the use of such PII to "carry out payment." It would make no sense for a patient to give his or her bank account number and then request restrictions on CE use of such information to "carry out payment;" therefore, the software engineer can design a system such that it traces the source of PII, hence focusing the rules accordingly. The IFM magnifies all types of relationships between entities such that ambiguity in the legal text can be easily spotted.

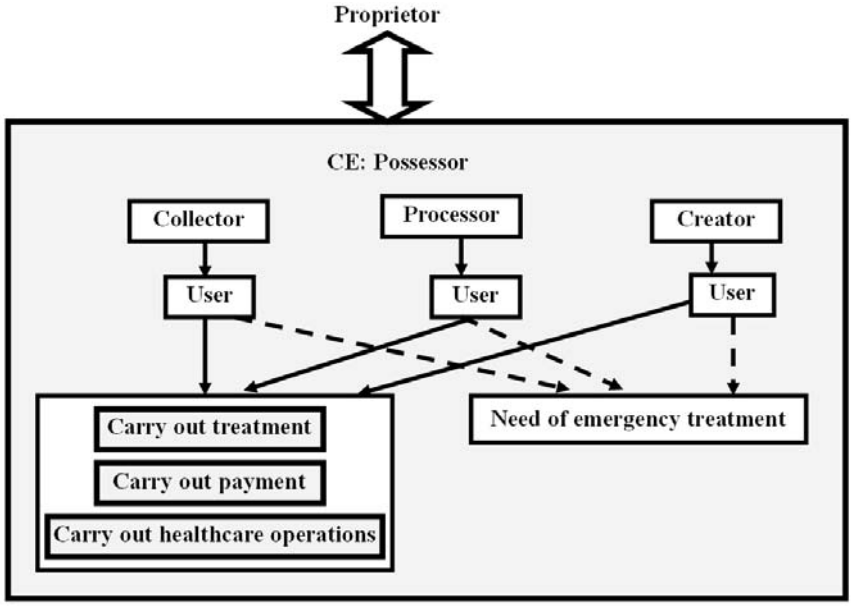

Figure 4. A relationship between a proprietor and CE involving uses of PII

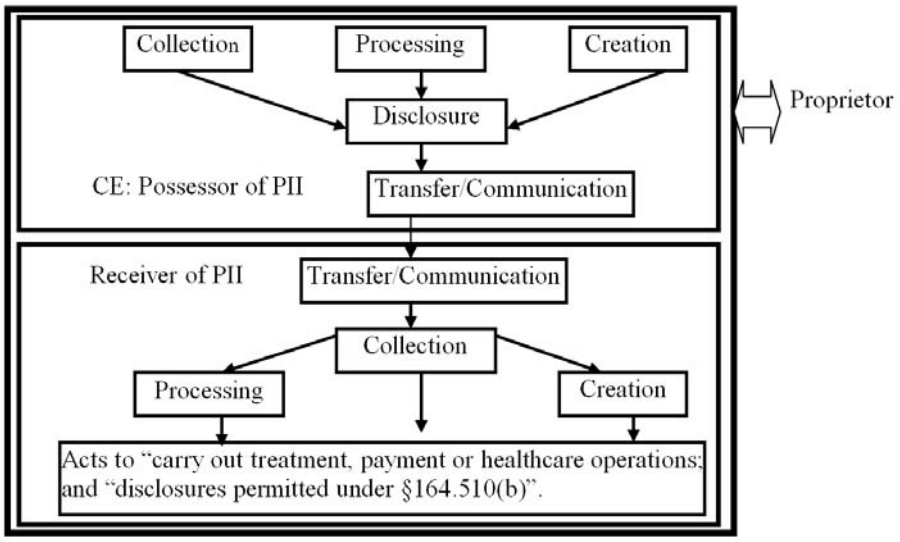

Figure 5. Relationship between proprietor and a situation involving CE disclosure of PII

6.2 The Disclosure Relationship

Now consider the disclosure relationship between the CE and the proprietor embedded in Privacy Rule excerpt §164.522(a)(1) given previously. The first question that comes in mind is, what type of PII is disclosed? Is it collected, processed, or created PII by the CE? Privacy Rule §164.522(a)(1) lumps these types together.

Since the CE is the party who discloses PII, then the user is the party who receives this information from the CE. Thus, the relationship involves three entities: the CE, the proprietor, and the receiver of PII, who uses it to "carry out treatment, payment or healthcare operations" and "disclosures permitted under §164.510(b)." Figure 5 represents this relationship.

Figure 5 depicts a proprietor (wide arrow) in a situation that includes the CE disclosing PII to a collecting agent, who uses it to "carry out treatment, payment or healthcare operations" and "disclosures permitted under §164.510(b)." Figure 6 represents the disclosure condition as described by the Privacy Rule. In comparison with Figure 5, the Privacy Rule is very brief and thus may create ambiguity in disclosure possibilities.

6.3 Assigning Roles to Users

Previous sections demonstrate that the Privacy Rule's

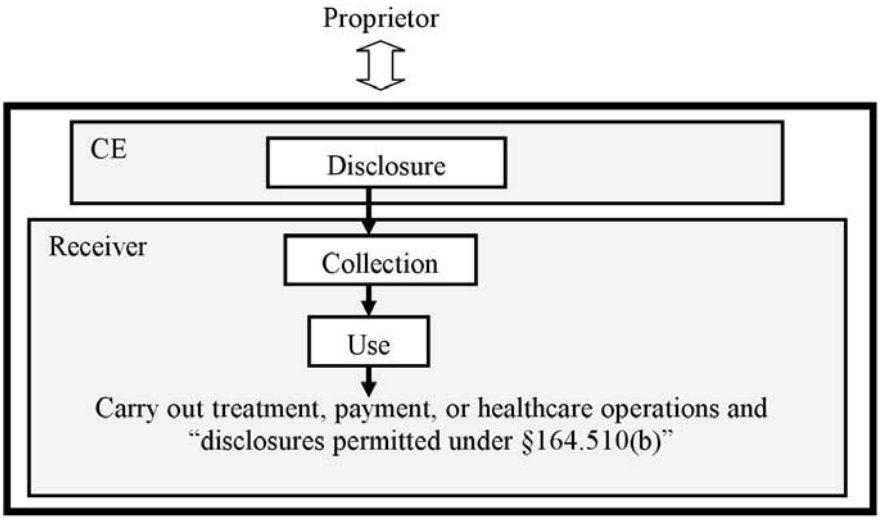

Figure 6. The Privacy Rule is very brief in describing the disclosure condition in comparison with the IFM

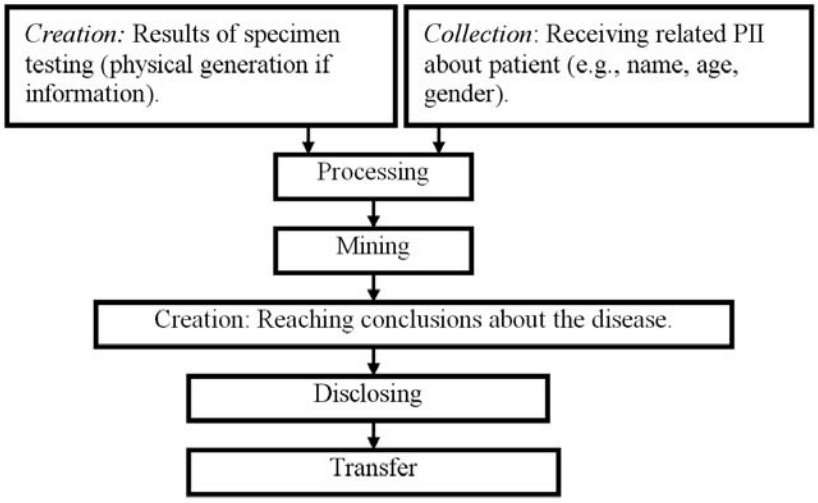

Figure 7. Laboratory as a user of PII

brevity hides types of uses and disclosures that may require different constraints and access rules. The notion of *purpose* that Breaux and Antón [10] use to indicate the goal of an activity remains to be discussed. In Role-Based Access Control (RBAC) systems [17], stakeholders (users) are permitted or denied access to information based on their *roles*. Roles are assigned to users, whereas purposes (we call them "uses") are assigned to data. In IFM, instead of a mapping between roles and purposes, a sub-graph of IFM is assigned to the users. For example, Figure 7 shows a sub-graph assigned to the laboratory as a user of PII in the hospital system. The hospital system discloses PII to the lab as part of a request for specimen testing. The lab possesses PII from two sources: collected PII from the hospital system, and created PII in the lab itself (the top two boxes in Figure 7). This PII is processed; mined and final conclusions are generated, then disclosed back to the hospital system.

The whole sub-graph (which probably also includes "Storage" sub-stages) represents the role of the lab. The sub-graph replaces the purpose, "resting specimen in laboratory." Access permission/denial of lab workers is decided according to their IFM (sub-graph). It is possible that inside the lab's IFM, several IFMs exist for the superintendent of the lab, lab technician, etc.

In such a system, where the infospheres of different PII possessors and users are very clear, it is possible to implement disclosure and access requirements in the fullest detail within a framework that can be understood even by laymen. The proprietor can request restrictions on disclosers to inside or outside handlers of PII in the possession of the CE. The software engineer can design the PII flow diagram as the network engineer draws the communication network. The drawing can be comple-

mented with all types of requirements at different points of PII flow. Complacence level can be claimed accordingly.

7. Conclusions

In this paper we have proposed complementing current software engineering methods to extract requirements from legal text, with a model that includes all possibilities of information flow. We have shown through examples that such a model provides a foundation for the interpretation process. The IFM is very general, such that it can be applied in all kinds of work in this area. We have concentrated on the analysis of Breaux and Antón [10] in order to reach some depth in illustrating the application of the IFM.

A great deal of work is needed to integrate the IFM into a workable product that generates the software requirement. The most productive approach is to incorporate the model into an existing methodology, rather than developing an IFM-based scheme from scratch. On the other hand, as a future work, theoretical analysis can be developed for the methodology to formally characterize its features.

REFERENCES

[1] D. Reinsel, C. Chute, W. Schlichting, J. McArthur, I. Xheneti, A. Toncheva, and A. Manfrediz, "A for- ecast of worldwide information growth through 2010." An IDC White Paper, 2007. http://www.emc.com/about/destination/digital_universe/pdf/Expanding_Digital_Universe_IDC_WhitePaper_022507.pdf

[2] Nexsan Technologies Inc, White paper on enabling information lifecycle management, 2005. http://www.meganet1.com/pdf/Enabling%20Information%20Lifecycle%20management.pdf

[3] M. J. May, C. A. Gunter, and I. Lee, "Privacy APIs: Access control techniques to analyze and verify legal privacy policies," 19th IEEE Workshop Computer Security Foundations, pp. 85–97, 2006.

[4] T. D. Breaux and A. I. Antón, "Deriving semantic models from privacy policies," 6th IEEE International Workshop on Policies for Distributed Systems and Networks, pp. 67–76, 2005.

[5] S-W. Lee, R. Gandhi, D. Muthurajan, D. Yavagal, and G-J. Ahn, "Building problem domain ontology from security requirements in regulatory documents," International Workshop on Software Engineering for Secure Systems, Shanghai, China, pp. 43–50, 2006.

[6] A. I. Antón, J. B. Earp, Q. He, W. Stufflebeam, D. Bolchini, and C. Jensen, "Financial privacy policies and the need for standardization," IEEE Security and Privacy, Vol. 2, No. 2, pp. 36–45, 2004.

[7] A. I. Antón, "Goal-based requirements analysis," 2nd IEEE International Conference on Requirements Engineering, pp. 136–144, 1996.

[8] T. D. Breaux and A. I. Antón, "Analyzing goal semantics for rights, permissions and obligations," 13th IEEE International Conference on Requirements Engineering, pp. 177–186, 2005.

[9] P. Giorgini, F. Massacci, J. Mylopoulos, and N. Zannone, "Modeling security requirements through ownership, permission and delegation," 13th IEEE International Conference on Requirements Engineering, pp. 167–176, 2005.

[10] T. Breaux and A. I. Antón, "Analyzing regulatory rules for privacy and security requirements," IEEE Transactions on Software Engineering, Vol. 34, No. 1, pp. 5–20, January 2008.

[11] D. Tindal, "Safety officer's briefing book," Civil Air Patrol, United States Air Force Auxiliary, February 1 2000. http://www.iawg.cap.gov/archives/ iawgsafety-manual.pdf.

[12] S. Al-Fedaghi, "Scrutinizing the rule: Privacy realization in HIPAA," International Journal of Healthcare Information Systems and Informatics (IJHISI), Vol. 3, No. 2, 2008.

[13] HHS, "Summary of the HIPAA privacy rule," U.S. Department of Health & Human Services, 2003. http://www.hhs.gov/ocr/privacysummary.pdf.

[14] S. Al-Fedaghi, "Software engineering interpretation of information processing regulations", IEEE 32nd Annual International Computer Software and Applications Conference (IEEE COMPSAC 2008), Turku, Finland, July 28–August 1, 2008.

[15] Office for Civil Rights, US Department of Health and Human Services, "Medical privacy: National standards to protect the privacy of personal health information," 2000 http://www.hhs.gov/ocr/hipaa/finalreg.html.

[16] T. D. Breaux and A. I. Antón, "Semantic parameterization: A conceptual modeling process for domain descriptions," North Carolina State University Computer Science Technical Report TR-2006-35, October 2006.

[17] R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman, "Role-based access control models," IEEE Computer, Vol. 29, No. 2, pp. 38–47, 1996.

