# Towards Regulatory Compliance: Extracting Rights and Obligations to Align Requirements with Regulations

Travis D. Breaux, Matthew W. Vail and Annie I. Antón
*Department of Computer Science*
*North Carolina State University*
*{tdbreaux, mwvail, aianton}@eos.ncsu.edu*

## Abstract

*In the United States, federal and state regulations prescribe stakeholder rights and obligations that must be satisfied by the requirements for software systems. These regulations are typically wrought with ambiguities, making the process of deriving system requirements ad hoc and error prone. In highly regulated domains such as healthcare, there is a need for more comprehensive standards that can be used to assure that system requirements conform to regulations. To address this need, we expound upon a process called Semantic Parameterization previously used to derive rights and obligations from privacy goals. In this work, we apply the process to the Privacy Rule from the U.S. Health Insurance Portability and Accountability Act (HIPAA). We present our methodology for extracting and prioritizing rights and obligations from regulations and show how semantic models can be used to clarify ambiguities through focused elicitation and to balance rights with obligations. The results of our analysis can aid requirements engineers, standards organizations, compliance officers, and stakeholders in assuring systems conform to policy and satisfy requirements.*

## 1. Introduction

Healthcare information systems are becoming ubiquitous and thus increasingly subject to attack, misuse and abuse. Specifications and designs often neglect security and privacy concerns [1]. Moreover, regulations such as the HIPAA [13] – as well as security and privacy policies – are difficult for users to understand and complex for software engineers to use as a guide for designing and implementing systems. Mechanisms are needed to help analysts disambiguate regulations so that they may be clearly specified as software requirements, in order to ensure that these systems conform to policy.

Policies assign different stakeholders various rights, such as the right to refrain from disclosing information, the right to use information freely for a specified purpose, etc. A *right* is a relational claim legitimately ascribed to a *right bearer* with respect to an implicit or explicit other, the *counterparty*. To say that a party has a right is a way of talking about the counterparty's implied obligation. For example, if a healthcare patient has a right to access their health records, then their physician's office has an obligation to provide access. An *obligation* is a duty bound to an *obligated party* that must be complied with, often accompanied with a penalty for non-compliance. Obligations, like goals [2, 21], can be operationalized into requirements. Reformulating rights in terms of the implied rights and obligations of counterparties will enable requirements engineers to specify requirements in terms of stakeholder commitments and question their legitimacy to achieve regulatory compliance.

It is estimated that healthcare organizations will spend $17.6 billion over the next few years to bring their systems and procedures into compliance with HIPAA [22]. Existing guidelines and standards not only fail to provide specific solutions, but also make compliance a significant challenge. According to a 2005 Ernst & Young survey of executives in over 1,300 international organizations, compliance with regulations and policy surpassed worms and viruses as the primary driver of information security policy in 2005 [9]. The consequence of not complying with regulations is now forefront for those responsible for assuring that software systems containing sensitive information remain secure and protected.

Current work is mostly focused on retooling and developing systems to comply with the HIPAA by capturing and processing the meta-data that must now be maintained and accounted for. From a developer's perspective, the first challenge is to interpret and define system requirements, given HIPAA's legal language. The need for such interpretation is not unique to HIPAA, but is particularly difficult in this case because collaborating organizations are performing their own HIPAA interpretations. The costs of error are great. Serious civil and criminal penalties are associated with misinterpretations, and the risk of such

misinterpretations increases with the interoperation of separate systems. A survey of different requirements frameworks reveals that none are well suited to deal with the relationships and obligations specified in the HIPAA [1].

To address this problem, we show how a process called Semantic Parameterization [3, 4] was combined with an extended methodology and applied to the HIPAA Privacy Rule [17] to derive rights and obligations. We provide strategies to identify and resolve ambiguities, express exceptions, and balance rights with obligations. The remainder of this paper is organized as follows. Section 2 discusses related work and defines the key terminology and notation. Section 3 provides an overview of the relevant aspects of the semantic parameterization process as it relates to the analysis of regulatory texts. Section 4 discusses our efforts to validate our extensions to semantic parameterization within the context of the HIPAA. Finally, Section 5 summarizes the contributions and limitations of this work and our plans for future work.

## 2. Background and Terminology

### 2.1 Related Work

Several researchers have sought to derive logical models from regulations and law without a specific focus on requirements engineering [6, 19, 26, 25]. These approaches rely on domain experts with logic programming skills to model regulations. We acknowledge the value in formal methods and ultimately seek to interface with approaches that include consistency and model checking; in our approach, however, we prioritize developing a standard process to extract and represent formal models. Despite this difference, these two approaches do complement each other with regards to the challenges in modeling regulations as we now discuss.

Bench-Capon et al. and Kowalski illustrate how to use negation to handle exceptions in regulation statements [6, 18]. In addition to negation, we allow some rights and obligations that are exceptions to have priority over other rights and obligations (see Section 4.2). Bench-Capon et al. also discuss the difficulty with interpreting cross-references in regulations [6]. Cross-references can refer to sets of performances (e.g., disclosures, notifications) from entire sub-sections, requiring an analyst to extract relevant material across multiple contexts. For this reason, we prefer prioritization to avoid the potential pitfalls of cross-context interpretation.

Kerrigan and Law describe a system that models environmental regulations using first-order predicate logic [19]. The system computes weighted relatedness scores between sections of regulation text using ontologies. We recognize a need to compare regulations at the finer statement-level to compare individual rights, obligations and constraints. To accomplish this aim, we show how queries over semantic models from our prior work [4] are used to partially order individual rights and obligations based on their level of refinement (see Section 4.3).

Normative theory investigates the relationships between permissions (rights) and obligations. Horty has shown that Deontic logic and reasoning must distinguish between what an agent “ought to do” and “ought to be” [15, 16]. Boella and van der Torre considers the relationship between permissions and obligations [7]: notably, are permissions simply the absence of obligatory restrictions and do permission presuppose balanced obligations? In Section 4.5, we begin systematically answering these questions in the context of specific regulations by showing how an approach to balance rights and obligations provides a means to investigate *implied* rights and obligations.

In requirements engineering, Darke and Shanks provide a conceptual framework for discussing viewpoints with regards to requirements coverage [8]. Frameworks have been proposed to facilitate elicitation [27], integrate viewpoints [23] and identify inconsistencies [10]. In our work, multiple viewpoints from different stakeholders and counterparties are embedded in rights and obligations that can be balanced to expose implied rights and obligations for other stakeholders (see Section 4.5).

Giorgini et al. present Secure Tropos (ST), a formal framework for modeling security requirements applied to Italian privacy legislation [12]. ST distinguishes between permissions (at-most) and obligations (at-least) in the context of delegation. ST employs Datalog to provide model checking capabilities necessary to find inconsistencies. Our work complements their framework by providing a standard process to extract permissions (rights) and obligations from regulations and law.

### 2.2 Terminology and Notation

We define the following key terms:

- A *stakeholder* is an entity afforded rights and/or obligations by the HIPAA Privacy Rule.
- A *right* is an action that a stakeholder is conditionally permitted to perform.
- An *obligation* is an action that a stakeholder is conditionally required to perform.
- A *delegation* is a right or obligation that a policy or stakeholder assigns to another stakeholder.
- A *rule statement* is the regulation text that includes the right or obligation and any constraints.
- A *constraint phrase* is the part of a rule statement that describes a single pre-condition.

- A *normative phrase* contains words that indicate what "ought to be" as rights or obligations.

*Notation Used*: the semantic models that appear in Sections 4.4 and 4.5 are expressed in the Knowledge Transformation Language (KTL) from prior work [4]. Symbols in the language have only one part-of-speech form: nouns, adjectives, verbs and adverbs; articles and prepositions are not allowed. Symbols prefixed by question marks are treated as wildcards. Symbols prefixed by exclamation marks are negated. Expressions support two types of anti-reflexive and asymmetric relations: 1) *delta* relations `x[y]` or `y=x`, which both read "x is y" and 2) *alpha* relations `x{y}` or `y:x` which read "x has y" and "y of x," respectively.

## 3. Methodology for Analyzing Regulations

In this study, we adopt a process called Semantic Parameterization, in which rights and obligations from regulation texts are restated into restricted natural language statements (RNLS), to describe discrete activities [4]. RNLS(s) are mapped into semantic models that are amenable to formal analysis [3, 4]. Semantic Parameterization was developed using Grounded Theory, in which theory that is systematically obtained from a dataset is valid for that dataset [11]. To date, we have applied Semantic Parameterization to three datasets: (1) the 100 most frequently-occurring semi-structured goals mined from over 100 privacy policies [3, 4]; (2) a pilot study analyzing unstructured text from a fact sheet [14] that summarizes the HIPAA Privacy Rule [17], yielding 15 rights and 19 obligations [5]; and (3) this case study in which we analyzed the unstructured regulation text from the Privacy Rule to yield 42 rights and 79 obligations. After each study, we generalize the parameterization process by reconciling the existing theory to overcome new limitations.

As in previous studies [3, 4], we employ a two-phase methodology. In the first phase, we extract rule statements using a relaxed form of Semantic Parameterization. The relaxed form uses only two RNLS patterns to separate the right or obligation phrase(s) from relevant constraint phrase(s). Constraint phrase(s) restrict the scope of actors and objects that already appear in the right or obligation phrase. Separated constraint phrase(s) are used to construct pre-conditions in the form of logical expressions. In the second phase, we derive semantic models using the activity pattern [4] from right, obligation and constraint statements, as necessary. Because the second phase is described in prior work [3, 4], its details are beyond the scope of this paper. We now illustrate the first phase by example.

The relaxed form of Semantic Parameterization uses only two RNLS patterns: (1) activities that distinguish subjects and objects [3, 4]; and 2) activities following condition keywords (*if*, *unless*, *except*) identified in an earlier pilot study [5]. Consider the unrestricted natural language statement $UNLS_1$ summarized from §164.522(a)(1)(iii) in the HIPAA Privacy Rule and parameterized as RNLS(s):

**$UNLS_1$**: A covered entity that agrees to a restriction ***may not*** use or disclose protected health information, ***except if*** the individual who requested the restriction is in need of emergency treatment.

**$RNLS_1$**: The covered entity who (**$RNLS_2$**) ***may not*** disclose protected health information, ***except if*** (**$RNLS_3$**).
**$RNLS_2$**: The covered entity agrees to a restriction.
**$RNLS_3$**: The individual who (**$RNLS_4$**) needs emergency treatment.
**$RNLS_4$**: The individual requests the restriction.

To extract rights and obligations, we first identify the normative phrase that defines what stakeholders are *permitted* or *required* to do. In $UNLS_1$, the normative phrase "may not" indicates an obligation and will appear in the first RNLS and obligation statement ($RNLS_1$). By extracting rights and obligations exclusively using normative phrases, the analyst's attention remains focused on *expressed* as opposed to *implied* rights and obligations. As we show in Section 4.5, implied rights and obligations are systematically obtained by balancing semantic models derived from expressed rights and obligations.

Each RNLS is restricted to one discrete state or activity and permits only one verb; however, $UNLS_1$ has two verbs, *use* and *disclose*, and describes two obligations *may not use* and *may not disclose*. In general, if the statement has a logical disjunction in the subject or verb phrases, we separate the statement into pair-wise distinct rights or obligations. English conjunctions (and, or) are not always equivalent to logical-and and logical-or. The analyst must decide whether the entities in an English conjunction are dependent (logical-and) or independent (logical-or) of one another when restating them in an RNLS.

RNLS(s) are separate constraints on a right or obligation if they distinguish subjects and objects (e.g., $RNLS_2$, $RNLS_4$) or follow condition words (e.g., $RNLS_3$). Consider $RNLS_1$, which distinguishes the *covered entity* using the nested $RNLS_2$ whereas $RNLS_3$ distinguishes the *individual* using the nested $RNLS_4$. In addition, the condition keywords *except if* in $RNLS_1$ are followed by the nested $RNLS_3$. We do not classify other RNLS(s) such as transitive verbs followed by verb phrases, instruments or purposes [3, 4] as separate constraints. Rather, these phrases remain nested in the

right, obligation or constraint statement for the purpose of this study.

The content in the regulation texts is highly segregated and indexed by stakeholder and process. To maintain traceability, an index from the original sub-section in the regulation text is mapped to each right, obligation and constraint statements. Finally, the constraints are organized into a logical expression based on a simple heuristic: two constraints are in a disjunction only if they are independent when invoking the rule, otherwise they are in a conjunction. The four RNLS(s) are indexed and organized as follows:

**Constraints:**

*A.* The covered entity agrees to a restriction. 164.522 (a)(1)(iii)

*B.* The individual needs emergency medical treatment. 164.522 (a)(1)(iii)

*C.* The individual requests a restriction. 164.522 (a)(1)(iii)

**Obligation**:

1. The covered entity **may not** disclose protected health information. 164.522 (a)(1)(iii) [$A \wedge \neg B \wedge C$]

The constraints *A*, *B*, and *C* are extracted from $RNLS_2$, $RNLS_3$ and $RNLS_4$, respectively. In this case, the obligation and each constraint are from the same sub-section and are indexed accordingly. The obligation is labeled with constraints *A*, *B*, and *C* in a logical conjunction in square brackets. Because $RNLS_3$ was part of an exception (e.g., *except if*), we negate the corresponding constraint *B* in the conjunction. We further discuss integrating exceptions in Section 4.2.

Regulation texts tend to be laden with cross-references to different sections within the given regulation. Analysts must follow each cross-reference to evaluate the impact of the referenced content on each right or obligation statement. Requirements engineers can use the indices to help maintain traceability as constraints are incorporated from different sections.

## 4. Analyzing Healthcare Regulations

To address the need for methodologies that aid in designing trustworthy healthcare systems that conform to policy and regulations, we applied our methodology to the HIPAA Privacy Rule [17]. The Rule is a regulation governing healthcare privacy in the U.S., including the provision and enforcement of privacy policies. Based on our discussions with CSOs, CISOs, and CPOs, organizations tend to assign priority to complying with those regulations most likely to interface with their beneficiaries, consumers and the public. For this reason, we focused on the following four sections in the Rule:

**§164.520**: Notice of privacy practices for protected health information.

**§164.522:** Rights to request privacy protection for protected health information.

**§164.524:** Access of individuals to protected health information.

**§164.526:** Amendment of protected health information.

The Rule is comprised of two parts, numbered 160 and 164, and contains a total of 33,500 words. The four sections we analyzed contain a total of 5,978 words or 17.8% of the Rule. Each part is sub-divided into subparts and each subpart is divided into sections. Sections §160.103 and §164.503 contain definitions necessary to distinguish the entities governed by the Rule. From the definitions, stakeholders can be organized into a class hierarchy. The class hierarchy in Figure 1 shows the stakeholders from the four Rule sections we analyzed. The arrows in the hierarchy represent sub-class relationships; for example, group health plans (GHP) and health maintenance organizations (HMO) are both types of health plans (HP), and an HP is a type of covered entity (CE).

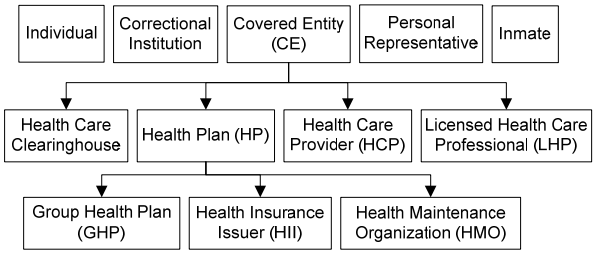


**Figure 1: Stakeholder Class Hierarchy**

In addition to §160.500 on Applicability, the definitions and corresponding class hierarchy are important when deciding which sections of the Rule apply to which stakeholders.

Each Rule section is divided into sub-sections that contain standards and implementation specifications, often with a separate emphasis on unique stakeholders. All of the sub-sections we analyzed were deemed to contain potential system requirements. The sub-sections are also sub-divided in ways that balance between conciseness and readability. Below, we summarize text from §164.520(c)(2) and (c)(3) as a model standard we simply index as (a) to illustrate the effect of applying the methodology to the regulation text. The normative phrase (must) and condition words (if, unless) are **bold**, the constraint phrases are underlined and the obligation phrases are *italicized*.

(a) Standard: *The covered entity **must** provide the individual notice.*

(1) *A covered entity* who has a direct treatment relationship with an individual ***must*** …
  (A) *Provide notice* no later than the first service delivery;
  (B) **If** the covered entity maintains a physical delivery site:
    *i. Have the notice available for individuals to take.*
    *ii. Post the notice in a clear and prominent location.*

(2) For the purposes of paragraph (a)(1), *a covered entity* who delivers service electronically, ***must*** *provide electronic notice* **unless** the individual requests to receive a paper notice.

Applying our methodology, we derive constraints A–E and obligations 1–4, below. The *italicized* phrases in obligations 2, 4, and 5 are ambiguities resolved using the regulation text.

**Constraints:**

*A.* The CE has a direct treatment relationship with the individual. (a)(1)
*B.* The notice is provided no later than the first service delivery. (a)(1)(A)
*C.* The CE maintains a physical delivery site. (a)(1)(B)
*D.* The CE delivers service electronically. (a)(2)
*E.* The individual requests to receive a paper notice. (a)(2)

**Obligations:**

*1.* The CE must provide notice to the individual (a).
*2.* The CE must provide notice *to the individual.* (a)(1)(A) [A ∧ B]
*3.* The CE must have the notice available for individuals to take. (a)(1)(B)(i) [A ∧ C]
*4.* The CE must post the notice in a clear and prominent location *for the individual to read*. (a)(1)(B)(ii) [A ∧ C]
*5.* The CE must provide electronic notice *to the individual.* (a)(2) [A ∧ B ∧ D ∧ ¬E]

Right, obligation and constraint statements are often distributed across several sub-sections in the Rule. For example, the subject *CE* is specified in sub-section (a)(1); however, the obligation phrases *provide notice*, *have notice available*, and *post the notice* each appear separately in sub-sections (a)(1)(A), (a)(1)(B)(*i*), and (a)(1)(B)(*ii*), respectively. The constraint *A* appearing in sub-section (a)(1) is applied across each of these obligations as well as the obligation in sub-section (a)(2) due to a cross-reference back to (a)(1). Cross-references to other sections in the Rule pose the greatest challenge to analysts, since each section is written from a different viewpoint; this makes the relevance of constraints from other sections subtle and uncertain.

The remainder of Section 4 is organized as follows: Section 4.1 provides summary results from applying the methodology; Section 4.2 presents two types of exceptions found in our analysis and strategies to address them; Section 4.3 illustrates the need to compare rights and obligations and our approach to address this need; Section 4.4 shows how semantic models are used to identify ambiguities in rights and obligations; and Section 4.5 shows how semantic models are used to balance rights with obligations.

## 4.1. Analysis Results from HIPAA

We identified 46 rights and 79 obligations in §164.520–§164.526. Table 1 summarizes the total number of rights ***(R)***, obligations ***(O)***, constraints ***(C)*** and cross-references ***(CR)*** identified per section. From a requirements engineering perspective, each constraint in the scope of the system must be satisfied by corresponding functionality in system design. Increasing the number of such constraints potentially increases the complexity of governed systems.

**Table 1: Number of Rights, Obligations, and Constraints in HIPAA §164.520–§164.526**

| ***Section*** | ***R*** | ***O*** | ***C*** | ***CR*** |
|---|---|---|---|---|
| 164.520 | 9 | 17 | 54 | 37 |
| 164.522 | 7 | 19 | 19 | 9 |
| 164.524 | 20 | 26 | 67 | 29 |
| 164.526 | 10 | 18 | 42 | 23 |

Table 2 summarizes the total number of unique normative phrases ***(N)*** we identified as well as the modality. In the table, *anti-rights* refer to activities the regulation explicitly exempts from stakeholder rights, but does not require the stakeholder to avoid (e.g., does not have a right to*).* Similarly, *anti-obligations* are activities that the regulation explicitly exempts from stakeholder obligations, but does not require the stakeholder to avoid (e.g., is not required to). We distinguish both anti-rights and anti-obligations from activities that are disallowed, such as obligations using the phrase "may not." Also in Table 2, normative phrases with an asterisk (*) indicate rights and obligations assigned through delegation. In delegation, a stakeholder is permitted or required to assign other stakeholders specific rights and obligations.

**Table 2: Normative Phrases in HIPAA Sections §164.520–§164.526**

| ***Phrase*** | ***N*** | ***Modality*** |
|---|---|---|
| does not have a right to | 1 | Anti-Right |
| has a right to | 7 | Right |
| is not required to | 3 | Anti-Obligation |
| may | 16 | Right |
| may deny* | 3 | Right |

| | | |
|---|---|---|
| may not | 2 | Obligation |
| may not require* | 1 | Obligation |
| may require* | 4 | Right |
| must | 39 | Obligation |
| must deny* | 1 | Obligation |
| must permit* | 13 | Obligation |
| must request* | 1 | Obligation |
| retains the right to | 1 | Right |

## 4.2. Prioritizing and Reconciling Exceptions

When extracting rights and obligations, the analyst must occasionally address an exception to a right or obligation. We identified 12 exceptions, each of which follows a special condition keyword (e.g., *except, if not, unless*). We address each exception in one of two ways: 1) increasing the priority of rights or obligations that were exceptions; or 2) by applying DeMorgan's Law to the constraints in an exception. We now illustrate both of these approaches by example.

In the first approach to exceptions, we define priority such that rights with higher priorities will exclusively overrule rights of lower priority; the same holds true for prioritized obligations. Consider rights $R_{0.1}$ and $R_{0.4}$ from §164.520:

**$R_{0.1}$:** An individual **has the right to** adequate notice from the CE of the uses and disclosures of PHI (a)(1).

**$R_{0.4}$:** An inmate **does not have a right to** notice (a)(3) from the CE of the uses and disclosures of PHI (a)(1).

In Section (a)(3) of the Rule, the phrase "exception for inmates" precedes the original rule statement used to derive $R_{0.4}$. In addition, the reference "under this section" denotes that the exception applies to all of §164.520. As a consequence, we increase the priority of $R_{0.4}$ over $R_{0.1}$ and any other rights from this section to ensure that inmates do not inadvertently receive rights the law did not intend them to have. In addition to the above example, four other exceptions were addressed by increasing the priority of the rights or obligations in the exception.

The second type of exception applies to constraints for a single right or obligation. In two cases, we identified exceptions that apply to a conjunction and disjunction of constraints. Consider the $UNLS_2$ summarized from §164.524, Section (a)(1):

**$UNLS_2$:** The individual **has a right to** access their PHI that is maintained in a designated record set (DRS), **except** for: (i) Psychotherapy notes; or (ii) PHI compiled for a legal proceeding; …

For a single constraint we only need to negate the exception and *logically-and* it to the constraint set for the right or obligation. However, given a set of constraints, the negation must be distributed over the set while observing DeMorgan's Law before we *logically-and* the result to the constraint set. Consider the constraints and right derived from $UNLS_2$:

**Constraints:**

*A.* The PHI is maintained in a DRS. (a)(1).
*B.* The PHI is psychotherapy notes. (a)(1)(i)
*C.* The PHI is compiled for a legal proceeding. (a)(1)(ii)

**Right:**

*1.* The individual may access their PHI. (a)(1) [$A \wedge \neg B \wedge \neg C$]

From the context of the exception, we conclude the exception is $\neg(B \vee C)$ and, after applying DeMorgan's Law, $(\neg B \wedge \neg C)$. We identified five exceptions where single constraints were only negated to integrate them.

## 4.3. Comparing Rights and Obligations

The Privacy Rule groups rights and obligations with a shared context together in the same section. While these rights and obligations are related, they may describe different actors, means of implementation, or reasons for which the regulation is intended. Consider, for example, the extracted obligations $O_{0.4}$, $O_{0.7}$ and $O_{0.8}$ from §164.520, below:

**$O_{0.4}$:** The GHP **is not required** to provide notice to any person (a)(2)(*iii*).

**$O_{0.7}$:** The CE **must** provide the notice to any person or individual (c).

**$O_{0.8}$:** The HP **must** provide the notice (c)(1)(*i*) to any person or individual (c).

Obligations $O_{0.4}$, $O_{0.7}$ and $O_{0.8}$ describe the obligations of the GHP, CE and HP, respectively, to provide notice to an individual. Recall from the stakeholder class hierarchy (Figure 1) that these actors have class relationships, such that a GHP is a type of HP and an HP is a type of CE. From these obligations alone, it is possible that an HP exists who must satisfy two obligations: first as an HP ($O_{0.8}$) and secondly as a CE ($O_{0.7}$). Also, a potential conflict exists since a GHP may not need to satisfy either obligation $O_{0.7}$ or $O_{0.8}$.

Fortunately, the semantic models derived from rights and obligations are comparable using queries [4]. Queries establish a partial order on semantic models sufficient to group rights or obligations that affect a single stakeholder at different levels of abstraction. To demonstrate, we modified our query algorithm to accommodate the stakeholder class hierarchy. Furthermore, we introduce *person* as a super-class for individual and inmate. In addition to obligations $O_{0.4}$, $O_{0.7}$ and $O_{0.8}$, consider the following obligations from §164.520 in the Rule:

**$O_{0.2}$:** The GHP **must** provide notice to any person (a)(2)(*ii*)(B).

**$O_{0.10}$:** The HCP **must** (c)(2) provide notice (c)(2)(*i*) to the individual (c).
**$O_{0.13}$:** The CE **must** provide electronic notice to the individual (c)(3)(*i*).
**$O_{0.14}$:** The CE **must** provide a paper copy of the notice to the individual (c)(3)(*ii*).
**$O_{0.15}$:** The HCP **must** automatically provide electronic notice to the individual (c)(3)(*iii*).

Each obligation has the same general requirement: the stakeholder must provide notice to a person. Applying the modified query algorithm generated the hierarchy illustrated in Figure 2:

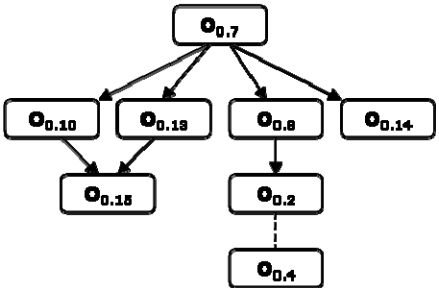


**Figure 2: Obligation Class Hierarchy**

Obligation $O_{0.7}$ is the most abstract obligation in the hierarchy – we call this the *least-constrained form* (LCF). The LCF can be generated from the minimally required model attributes for a right or obligation: the *subject*, *action* and *object* (required by the activity pattern) and the normative phrase (required by the methodology). Obligation $O_{0.15}$ combines attributes from $O_{0.10}$ and $O_{0.13}$: specifically, the notice must be *electronic* and provisioned by the *HCP*. Compare this refinement to obligation $O_{0.14}$, which governs the provision of paper notices. The obligation $O_{0.4}$ is actually an exception to $O_{0.2}$, denoted by the dotted line, which defines the situations when a GHP is not required to provide notice to individuals. Queries provide limited support to detect conflicts such as the one between obligations $O_{0.2}$ and $O_{0.4}$. Only those conflicts resulting from negation on type-similar model symbols (e.g., *right* and *not right*; *CE* and *not HCP)* are both detectable. Constraints are comparable in the same manner using queries.

### 4.4. Resolving Ambiguities

We consider two types of ambiguities in the Privacy Rule: (1) English conjunctions and (2) under-specifications.

English conjunctions are potentially (and frequently) ambiguous. Consider the following obligation from §164.522 in the Rule:

**$O_{2.5}$:** The CE must document in writing **or** electronically any request **and** agreement to a restriction on the use **or** disclosure of PHI (a)(3).

In $O_{2.5}$, there are three (bolded) English conjunctions. We chose to interpret these as *exclusive-or*, *logical-and*, and *logical-or*, in the order in which they appear. To derive this interpretation, we assume the CE does not need to document both "in writing" and "electronically": this alternative would be *logical-or*, not *exclusive-or*, as we propose. Furthermore, we assume the CE is only required to document requests for restrictions to which they agree, and not requests to which they disagree. The alternative interpretation is *logical-or*, which would increase the quantity of documentation produced by corresponding requirements. This illustrates how alternate interpretations are biased by the stakeholder goals. For example, the covered entity needs to reduce cost and workload, the auditor needs information to review practices, and individuals need to add and remove restrictions.

Due to under-specifications in the Rule, there is a risk of losing important contextual information distributed across multiple statements while extracting rights and obligations. Analysts can leverage the correspondence between RNLS(s) and semantic models to identify missing information or clarify ambiguities. We highlight two frequent types of ambiguity observed in this study: missing co-requisite attributes and verb phrases masquerading as nouns. We discuss each of these in detail using right $R_{4.4}$ from §164.524 below:

**$R_{4.4}$**: The individual **has a right to** have a denial reviewed. 164.524(a)(3)

The activity pattern is frequently used in semantic models [4] and has three co-requisite attributes: *subject*, *action* and *object*. If the subject or object is well-defined within the broader context of the regulation text, some phrases will omit one or both. For example, the subject of phrases describing the *purpose* is typically omitted. In Figure 3, the semantic model for $R_{4.4}$ is shown using the activity pattern to describe the right of the individual (Line 1).

```
1   activity [ right : individual & R.4.4 ] {
2       subject = ?someone
3       action = review
4       object = denial
5   }
```

**Figure 3: Ambiguous Activity Pattern**

The broader context from which $R_{4.4}$ was extracted states the reviewer of denials is an LHP designated by the CE; however, because the methodology is statement-driven, $R_{4.4}$ has been extracted without a proper *subject* attribute (Line 2). The activity pattern requires a *subject*, therefore this ambiguity can be automatically identified using queries to check for

missing subjects, actions and objects [4]. In addition to the activity pattern, specific verbs have co-requisite attributes. For example, the verb *disclose* has a *target* or *recipient* of the disclosure and the verb *use* has a *purpose* for which an *object* is used [4].

Another frequent ambiguity occurs when verb phrases masquerade as nouns. Specific activities that are elaborated in verb phrases can be summarized by a single noun; English gerunds are one example, often observed in purposes. For example, the *denial* in $R_{4.4}$ modeled on Line 4 in Figure 3 can be elaborated by the verb phrase "to deny…" The broader context reveals that the *denial* refers to the "CE denies access" where *access* describes the "individual's access to PHI." Figure 4 shows how to expand these nouns into separate activities: the *denial* (Line 4) and the *access* (Line 7). In addition, we elicit the required *subject*, *action*, and *object* attributes for each new activity.

```
1   activity [ right : individual & R.4.4 ] {
2       subject = LHP
3       action = review
4       object = denial {
5           subject = CE
6           action = deny
7           object = access {
8               subject = individual
9               action = access
10              object = PHI
11          }
12      }
13  }
```

**Figure 4: Verb Phrases Masquerade as Nouns**

By maintaining a separate list of these nouns and their verb counterparts, these ambiguities are automatically identified and can be partially resolved whenever the nouns are assigned to the *object* or *purpose* attribute. We developed an algorithm to automatically identify these ambiguities; however, the values assigned to the *subjects* and *objects* must be manually acquired from the regulation text.

Using the query algorithm, we developed a tool that automatically identified the following missing co-requisite attributes in semantic models from all four sections: 28 subjects, 8 actions, and 21 objects, in addition to 27 sources and 37 targets that are required for some verbs.

### 4.5. Balancing Rights and Obligations

Rights and obligations that govern interactions between stakeholders may be balanced to derive *implied* rights and obligations. Deriving implied rights and obligations is important to increase requirements coverage, since obligations derived from rights – either implied or expressed – may be operationalized as requirements. We discuss balancing rights and obligations for interactions including *delegations*, where a stakeholder assigns a right or obligation to another stakeholder, and *provisions*, where a stakeholder provides another stakeholder some object.

We identified 28 rights and obligations that allow stakeholders to delegate rights and obligations to other stakeholders. In the semantic models for delegations, the *actions* are from a restricted set of transitive-verbs (e.g. *restrict*, *require*, *deny*, and *permit*) and the *object* of the delegation is an activity (implied right or obligation) performed by the *subject* of that activity. For example, consider right $R_{6.3}$ from §164.526, below, that describes a right of the CE to delegate an obligation to the individual.

**$R_{6.3}$:** The CE **may require** the individual to make written requests for amendment to their PHI.

The corresponding semantic model in Figure 5 illustrates the right in Lines 1–15. Balancing this right requires extracting the implied obligation (Lines 4–14) and modeling that activity separately in Lines 15–24. Implied rights and obligations are rarely unconditional. For example, before the individual is obligated, the CE must first obligate the individual by invoking their right to do so, which further requires the CE to satisfy constraints on their right to obligate the individual.

```
1   activity [ R.6.3 & right ] {
2       subject = CE
3       action = require
4       object = activity {
5           subject = individual
6           action = request
7           object = activity {
8               subject = CE
9               action = amend
10              object = PHI : individual
11          }
12          instrument = writing
13          target = CE
14      }
15  }

15  activity [ R.6.3.B & obligation ] {
16      subject = individual
17      action = request
18      object = activity {
19          subject = CE
20          action = amend
21          object = PHI : individual
22      }
23      instrument = writing
24  }
```

**Figure 5: Example Right Balanced with an Obligation**

We identified 41 rights and obligations that describe provisions, a class of stakeholder interactions where one stakeholder provides an object and another stakeholder receives that object. Semantic models for

these activities also use specific actions such as *disclose*, *inform*, *notify*, *provide*, and *receive*. In addition to the subject attribute of the activity, a complementary co-requisite attribute must be specified as either the *source* (the provider) or *target* (the receiver). Consider the right $R_{0.2}$ from §164.520:

**$R_{0.2}$:** The individual **has a right to** receive notice from the GHP. 164.522(a)(1)

In Figure 6, the right $R_{0.2}$ is modeled in Lines 1–6 and balanced by the right $R_{0.2.B}$ modeled in Lines 7–12. The action in $R_{0.2}$ is *receive* and has the co-requisite attribute *source* (the provider of the notice). Some actions in provisions will have a binary opposite where the subject of the activity assumes the value of the co-requisite attribute. For example, in the balancing of $R_{0.2.B}$, the action *provide* is a binary opposite to *receive*. As a result, the *subject* (Line 2) and *source* (Line 5) from $R_{0.2}$ are mapped onto the *target* (Line 11) and the *subject* (Line 8) in $R_{0.2.B}$, respectively.

```
1   activity [ R.0.2 & right ] {
2       subject = individual
3       action = receive
4       object = notice
5       source = GHP
6   }

7   activity [ R.0.2.B & obligation ] {
8       subject = GHP
9       action = provide
10      object = notice
11      target = individual
12  }
```

**Figure 6: Balancing Direct Provisions**

Finally, we observed that when the purpose of an obligation is a provision, the purpose may be balanced by an implied right. For example, consider the obligation $O_{0.12}$ from §164.520, below:

**$O_{0.12}$:** The HCP **must** (c)(2) post the notice for an individual to read. (c)(2)(ii)(B)

In Figure 7, obligation $O_{0.12}$ is modeled in Lines 1–10. The *purpose* (Line 5) is an activity that describes an implied right of the individual. We balance $O_{0.12}$ with the obligation $O_{0.12.B}$ modeled in Lines 11–15.

```
1   activity [ O.0.12 & obligation ] {
2       subject = HCP
3       action = post
4       object = notice
5       purpose = activity {
6           subject = individual
7           action = read
8           object = notice
9       }
10  }

11  activity [ O.0.12.B & right ] {
12      subject = individual
13      action = read
14      object = notice
15  }
```

**Figure 7: Balancing Indirect Provisions**

In summary, we show three ways to balance expressed rights and obligations with implied rights and obligations: delegations, direct provisions, and indirect provisions. Identifying implied rights and obligations helps ensure requirements coverage is more complete under the law.

## 5. Discussion and Future Work

We now outline the limitations in our methodology, some observations and discuss future work.

More work is required to consider the role of constraints in identifying conflicts between rights and obligations. In this work, we only identify trivial conflicts by observing negation and type-similar values in semantic models. More sophisticated approaches are required to consider process conflicts in which the cause of a conflict is implicit in a chain of events. In requirements engineering, Lamsweerde et al. model conflicts at the goal-level [20]. We believe decomposing constraints using semantic models will provide insight into model symmetries and queries that may assist with requirements partitioning [24] or identifying more sophisticated conflicts.

Another matter concerning constraints that was not addressed in depth is the distinction between who provides information to satisfy constraints from who validates such information. For example, the following RNLS is a constraint from §164.520, Section (a)(2)(i): *The individual is enrolled in a group health plan (GHP)*. To satisfy this constraint, an individual who intends to receive services from a healthcare provider (HCP) might *provide* contact information for their GHP, who would in turn validate the individual's enrollment for the HCP. Using our methodology, one can iterate through constraints to distinguish between who provides and who validates such information.

The phrase heuristics we identified are limited to this dataset and may be insufficient or inconsistent when analyzing other regulations and policies. Only three of the 14 phrase heuristics identified in our pilot study [5] were observed in this study; largely due to the summary nature of that dataset. As for inconsistencies, the English word "may" means either "is permitted to…" or "is expected to…" in which only the first form indicates a right or permission.

Finally, we observed that certain events appear in the pre-conditions of rights and obligations and other events are by-products of invoking rights and obligations. We are presently developing a process to systematically identify and link these events to states

containing rights and obligations. We believe the state-transition diagrams produced by this process can be used in risk analysis and compliance monitoring.

## 6. Conclusion

In conclusion, we presented the results of applying the Semantic Parameterization process in an extended methodology to extract rights and obligations from regulation text in healthcare. We present our approach to handle exceptions, compare rights and obligations and identify ambiguities. We also show how to balance rights and obligations to identify implied rights and obligations necessary to ensure requirements coverage and consider multiple viewpoints.

## Acknowledgements

We thank the members of ThePrivacyPlace.org for their helpful comments. This work was funded by NSF grant *ITR: Encoding Rights, Permissions and Obligations: Privacy Policy Specification and Compliance* (NSF #032-5269).

## References

[1] A.I. Antón, J.B. Earp, C. Potts and T.A. Alspaugh. "The Role of Policy and Privacy Values in Requirements Engineering," *IEEE 5th Int'l Symp. on Reqs. Engr. (RE'01)*, Toronto, Canada, pp. 138-145, 27-31, 2001.

[2] A.I. Antón, C. Potts, "The Use of Goals to Surface Requirements for Evolving Systems." *Int'l Conf. on Soft. Engr.*, Kyoto, Japan, pp. 157-166, 1998.

[3] T.D. Breaux, A.I. Antón, "Deriving Semantic Models from Privacy Policies." *IEEE 6th Workshop on Policies for Distributed Systems and Networks*, Stockholm, Sweden, pp. 67-76, 2005.

[4] T.D. Breaux, A.I. Antón, "Analyzing Goal Semantics for Rights, Permissions, and Obligations." *IEEE 13th Req'ts. Engr. Conf.*, Paris, France, pp. 177-186, 2005.

[5] T.D. Breaux, A.I. Antón, "Mining Rule Semantics to Understand Legislative Compliance." *ACM Workshop on Privacy in Electronic Society*, Alexandria, Virginia, USA, pp. 51-54, 2005.

[6] T.J.M. Bench-Capon, G.O. Robinson, T.W. Routen, M.J. Sergot, "Logic Programming For Large Scale Applications in Law: A Formalization of Supplementary Benefit Legislation" *1st Int'l Conf. on AI and Law*, Boston, MA, USA, pp. 190-198, 1987.

[7] G. Boella, L. van der Torre, "Permissions and Obligations in Hierarchical Normative Systems." *9th Int'l Conf. on AI and Law*, Scotland, UK, pp. 109-118, 2003.

[8] P. Darke and G. Shanks. "Stakeholder Viewpoints in Requirements Definition: A Framework for Understanding Viewpoint Development Approaches." Requirements Engineering, 1(2) pp. 88–105, 1996.

[9] Ernst & Young, *Global Information Security Survey 2005: Report on the Widening Gap*, 2005.

[10] S. Easterbrook, M. Chechik. "A Framework for Multi-valued Reasoning Over Inconsistent Viewpoints." *23rd Int'l Conf. on Soft. Engr.* Toronto, Ontario, Canada, pp. 411-420, 2001.

[11] B.C. Glaser and A.L. Strauss, *The Discovery of Grounded Theory*, Aldine Publishing Co., 1967.

[12] P. Giorgini, F. Massacci, J. Mylopoulos, N. Zannone. "Modeling Security Requirements Through Ownership, Permission and Delegation." *IEEE 13th Req'ts. Engr. Conf.*, Paris, France, pp. 167-176, 2005.

[13] *Health Insurance Portability and Accountability Act*, USC H.R. 3103-168, April 2000.

[14] *Fact Sheet: Protecting the Privacy of Patients' Health Information*, U.S. Department of Health and Human Services, Washington D.C., April 14, 2003.

[15] J.F. Horty. Deontic logic as founded in nonmonotonic logic. *Annals of Mathematics and Artificial Intelligence*, v. 9, pp. 69--91, 1993.

[16] John F. Horty. *Agency and Deontic Logic*, Oxford University Press, 2001

[17] "Standards for Privacy of Individually Identifiable Health Information." 45 CFR Part 160, Part 164 Subpart E. In Federal Register, vol. 68, no. 34, February 20, 2003, pp. 8334 – 8381

[18] R. Kowalski, "The Treatment of Negation in Logic Programs for Representing Legislation." *2nd Int'l Conf. on AI and Law*, Vancouver, BC, Canada, pp.11-15, 1989.

[19] S. Kerrigan, K.H. Law, "Logic-based Regulation Compliance-Assistance." *9th Int'l Conf. on Artificial Intelligence and Law*, Scotland, UK, pp. 126-135, 2003.

[20] A. van Lamsweerde, R. Darimont, E. Letier, "Managing Conflicts in Goal-Driven Requirements Engineering." IEEE Trans. on Soft. Engr. 24(11), pp. 908-926, 1998.

[21] E. Letier, A. van Lamsweerde, "Deriving Operational Software Specifications from System Goals." In Proc. 10th ACM Symp. on Foundations of Soft. Engr., Charleston, SC, USA, pp. 119-128, 2002.

[22] Medical Privacy - National Standards to Protect the Privacy of Personal Health Information. Office for Civil Rights, US Department of Health and Human Services. 2000. http://www.hhs.gov/ocr/hipaa/finalreg.html.

[23] B. Nuseibeh, J. Kramer, A. Finkelstein. "Expressing the Relationships between Multiple Viewpoints in Requirements." In Proc. *15th Int'l Conf. Soft. Engr.* Baltimore, MD, USA, pp. 187-196, 1993.

[24] W.N. Robinson, S.D. Pawlowski, V. Volkov, "Requirements Interaction Management." ACM Comp. Surv. 35(2), pp.132-190, 2003.

[25] M.J. Sergot, A.S. Kamble, K.K. Bajaj, "Indian Central Civil Service Pension Rules: A Case Study in Logic Programming Applied to Regulations." *3rd Int'l Conf. on AI and Law*, Oxford, England, pp. 118-127, 1991.

[26] D.M. Sherman, "A Prolog Model of the Income Tax Act of Canada." *1st Int'l Conf. on AI and Law*, Boston, MA, USA, pp. 127-136, 1987.

[27] I. Sommerville, P. Sawyer, "Viewpoints: Principles, Problems and a Practical Approach to Requirements Engineering." Annals of Soft. Engr. 3(0), pp. 101-130, 1997.