North Carolina State University
Computer Science Technical Report # TR-2011-4

A Legal Cross-References Taxonomy for Identifying Conflicting Software Requirements

Jeremy C. Maxwell[1,2], Annie I. Antón[1], Peter Swire[3]
[1]Department of Computer Science, North Carolina State University, Raleigh, NC, USA
[2]Allscripts Healthcare Solutions, Raleigh, NC, USA
[3]Moritz College of Law, Ohio State University, Columbus, Ohio, USA
{jcmaxwe3, aianton}@ncsu.edu, swire.1@osu.edu

***Abstract*—Companies must ensure their software complies with relevant laws and regulations to avoid the risk of costly penalties, lost reputation, and brand damage resulting from noncompliance. Laws and regulations contain internal cross-references to portions of the same legal text, as well as cross-references to external legal texts. These cross-references introduce ambiguities, exceptions, as well as other challenges to regulatory compliance. Requirements engineers need guidance as to how to address cross-references in order to comply with the requirements of the law. Herein, we analyze each external cross-reference within the U.S. Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule to determine whether a cross-reference either: introduces a conflicting requirement, a conflicting definition, and/or refines an existing requirement. Herein, we propose a legal cross-reference taxonomy to aid requirements engineers in classifying cross-references as they specify compliance requirements. Analyzing cross-references enables us to address conflicting requirements that may otherwise thwart legal compliance. We identify five sets of conflicting compliance requirements and recommend strategies for resolving these conflicts.**

Keywords-Healthcare IT, Conflicting Requirements, Regulatory Compliance, Requirements Engineering

I. INTRODUCTION

Software developers must ensure that the software they develop complies with relevant laws and regulations. Compliance with regulations, lost reputation, and brand damage resulting from privacy and security breaches are increasingly driving information security and privacy policy decisions [14]. The costs of noncompliance are significant. For example, the ChoicePoint data breach cost the company over 27 million dollars, in addition to having government audits for 20 years [29]. One of these audits revealed further data breaches, resulting in $275,000 in additional fines [22].

Despite the high cost of noncompliance, developing legally compliant software is challenging. Legal texts contain ambiguities [6, 28]. Requirements engineers need to understand domain-specific definitions and vocabulary before they can interpret and extract compliance requirements [28]. Cross-references between different portions of a legal text can be ambiguous and force engineers to analyze the law in a non-sequential manner [5, 6], and cross-references to external legal texts increase the number of documents engineers must analyze in order to obtain compliance requirements [28].

Researchers are providing engineers with techniques and tools for specifying and managing software requirements for legally compliant systems [5, 9, 15, 23, 26, 27, 31, 36]. However, these tools and techniques do not take into account cross-references to other legal texts. These cross-references to external texts are important to analyze, because they may introduce conflicts or refine existing requirements.

The purpose of our research is to develop techniques that aid requirements engineers in identifying compliance requirements that appear to conflict so these conflicts may subsequently be resolved. Herein, we demonstrate techniques that requirements engineers can use to resolve important categories of apparent conflicts prior to meetings with legal domain experts. In our work, "conflict" refers to requirements that differ and may contradict each other. Some conflicts may be resolved with the techniques discussed in this paper; other conflicts will require consultation with subject matter experts.

To assist engineers in classifying legal requirements and identifying conflicts, we develop a taxonomy of legal cross references through a case study of the U.S. Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule[1]. The HIPAA Privacy Rule regulates the use of protected health information (PHI) by certain organizations, called covered entities. Covered entities include doctor offices, hospitals, and health insurers. In our study, we identified and classified 108 cross-references within the Privacy Rule. We examined each of these referenced legal texts, which yielded an additional 69 cross-references among these texts, resulting in 177 total cross-references. Among these cross-references, we identified five sets of conflicting compliance requirements and recommend strategies for resolving these kinds of conflicts. Because noncompliance is costly, it is imperative for requirements engineers to identify and examine cross-references to resolve any potential conflicts.

The remainder of the paper is organized as follows: Section II reviews related work and provides relevant legal context; Section III describes our research design; Section IV presents our results; Section V provides a discussion of our findings; Section VI discusses threats to the validity of our study; and Section VII summarizes the paper.

II. RELATED WORK & BACKGROUND

In this section, we describe related work and provide a legal background.

[1] 45 CFR Parts 160, 162, and 164

A. Related Work

Researchers note that cross-references are challenging for legal compliance [2, 5, 6, 7, 18, 27, 28]. Specifically, cross-references: can be ambiguous about which legal text takes precedence [5, 18, 27]; decrease understanding of legal texts [2]; add additional priorities and exceptions to compliance requirements [28]; may have a differing context from the citing text [7]; and may cite portions of the legal text out of sequence, causing "engineers to skip around the regulation text" [6].

Massey et al. use cross-references, along with other factors, to prioritize compliance requirements, but do not analyze the cross-referenced texts [23]. They classify cross-references as internal (a reference between different portions of the same legal text) or external (a reference between portions of different legal texts) [23]. However, using this simple distinction introduces ambiguity, as the boundaries between legal texts may not immediately be clear to an engineer. Thus, in Section III, we further classify cross-references into one of four patterns.

Requirements engineering research has focused on internal cross-references [5, 24, 27] rather than external cross-references. External cross-references are more challenging to legal compliance than internal cross-references, because different legal texts are likely to have differing context, definitions, and priorities. In our previous work, we model the HIPAA Privacy Rule, Part E, using production rules [24, 26]. We obtain additional preconditions for production rules from internal cross-references. Breaux uses natural language patterns to identify internal cross-references in the HIPAA Privacy Rule and codify mappings between the respective compliance requirements [5]. He then extracts priorities, exceptions, and refinements to compliance requirements from the identified cross-references [5]. External cross-references are outside the scope of Breaux's study [5]. May et al. use Promela to express the HIPAA Privacy Rule, including internal cross-references [27]. However, they do not analyze external cross-references [27]; instead, they use environmental flags to signal whether or not an external cross-reference is satisfied [27].

Van Engers and Boekenoogen use scenarios and the Unified Modeling Language (UML) to detect errors in the law and improve legal text quality [33]. They obtain scenarios by interviewing legal domain experts and model sequence of events using decision trees [33]. They do not, however, capture important contextual information about the scenario such as the scenario's goal, actors involved, and resources needed. After obtaining these scenarios, Van Engers et al. analyze a draft of a bill that, at the time, was going through the Dutch law-making process [33]. They identify and describe four inconsistencies in the law, namely, incorrect cross-references, ambiguous references, gaps in the law, and irrelevant portions of the law [33]. Using our legal cross-references taxonomy, we can identify all of the inconsistencies identified by Van Engers and Boekenoogen. Van Engers and Boekenoogen's methodology requires discussions with legal domain experts to specify the scenarios at the beginning of the process [33].

Cholvy checks regulation consistency using SOL-resolution by modeling the regulation text using a first order language [8]. Using these formal modeling tools, Cholvy identifies potential dilemmas, i.e. a regulation both obligates and forbids an actor from performing an action [8]. Herein, we develop a taxonomy that requirements engineers can use to identify conflicts in the law without having to formally model the regulation. Our taxonomy also aids engineers in identifying refinements to existing compliance requirements and areas of the law that are not applicable to software systems, meaning they can be ignored.

Hamdaqa and Hamou-Lhadj present a classification scheme for legal cross-references outline a tool-supported, automated process for extracting cross-references and generating cross-reference graphs [18]. They classify cross-references into two groups, amendments and assertions [18]. Amendments track evolution in the law. Assertion cross-references are further classified using three subtypes: definition cross-references provide a definition; specification cross-references provide more information about the legal text; and compliance cross-references conform the cited text with the citing text [18]. Their definition and specification subtypes align with the definition and refine classifications in our taxonomy, respectively. Our taxonomy extends their classification scheme by identifying exception, incorrect, unrelated, and general cross-references, as well.

Requirements engineers have used compliance and traceability links in compliance research. Ghanavati et al. use compliance links to trace goals, softgoals, tasks and actors to the law [15]. They use traceability links to connect portions of a Goal Requirements Language (GRL) business model with a GRL model of the law [15]. They tag each of the links with either a quantitative value or a qualitative category representing the degree to which a business satisfies the compliance requirements [15]. Their traceability links can be used to link goal models of the law [15]. Berenbach et al. use just in time tracing (JIT) to identify: (1) regulatory requirements; (2) system requirements that satisfy said requirements; and (3) sections of the law that require further analysis [3]. Cleland-Huang et al. use automated techniques to identify traceability links between HIPAA and 10 sets of requirements specifications [9]. Herein, we present research that can be used to complement existing requirements engineering methods to identify requirements conflicts in cross-references.

Zhang and Koppaka create legal citation networks based on the citations found in case law [37]. Their tool can be used to identify and track legal issues as they evolve [37]. Their tool, however, is designed to be used by legal domain experts and is designed to support case law [37]. In contrast, our analysis techniques are designed to be used by requirements engineers to examine the impact of cross-references to legal texts on software requirements.

Requirements researchers have examined conflicts in software requirements [4, 12, 30, 32, 34]. Robinson and Fickas describe how to detect and resolve requirements conflicts using a tool-supported approach [30]. Boehm and In use the WinWin model for negotiating resolutions to conflicts among quality attributes [4]. Van Lamsweerde et al.

use KAOS to identify and resolve conflicts among software goals [34]. Easterbrook and Nuseibeh use the ViewPoints Framework to handle inconsistencies as a requirements specification evolves [12]. Emmerich et al. examine standards such as ISO and built a prototype policy checker engine in DOORS [13]. Thurimella and Bruegge examine conflicts among the requirements of various product lines [32]. To the best of our knowledge, no researchers have examined conflicts introduced by legal cross-references.

B. *Legal Background*

Cross-references are citations from one portion of a legal text to another portion of that text or to another text. The referencing text is the legal text that contains the cross-reference and the referenced text is the legal text that is cited. Laws in the U.S. are codified in several places at the federal, state, and local levels, but there is no comprehensive legal code [10]. For instance, the complex legal structure of federalism governs when federal or state law takes precedence [19]. At both the federal and state level, the complex legal structure of separation of powers governs how power is allocated among the three branches of government, with statutory law developed by legislative bodies, administrative law issued by executive agencies, and judicial branch decisions that become case law [10]. Herein, we identify cross-references in the HIPAA Privacy Rule to the U.S. Code, the Code of Federal Regulations, and to Executive Orders issued by the President.

The U.S. Code is a compilation of legislative law passed by the U.S. Congress [10]. The U.S. Code is divided by subject into 50 titles. For example, Title 42 relates to Public Health and Wellness, whereas Title 22 relates to International Relations. Citations to the U.S. Code are formatted as (Title-Number) U.S.C. (Section Number). For example, 22 U.S.C. 2709(a)(3) is a citation to Title 22, section 2709, subsection (a), paragraph (3). Sometimes, cross-references will cite a statutory law by its commonly used name, instead of using the U.S. Code title and section number. The U.S. Code contains a table of "Acts Cited by Popular Name" that can be used to determine the title and section numbers for these laws [10]. For example, the Privacy Rule cites the Foreign Service Act. Using the "Acts Cited by Popular Name" table, we determine that the Foreign Service Act is codified at 22 U.S.C. 3901 et seq.

The Code of Federal Regulations (CFR) is a compilation of regulations published by executive branch agencies [10] such as the Department of Health & Human Services (HHS) or the Food and Drug Administration. The CFR is divided by subject into 50 titles but not using the same subject divisions as the U.S. Code. The CFR is cited similar to the U.S. Code. For example, 42 CFR 493.3(a)(2) is a citation to Title 42, section 493.3, subsection (a), paragraph (2).

An Executive Order is an "exercise of presidential authority related to government business" with sequential numbering in the order they are issued [10]. For example, §164.512(k)(2) of the Privacy Rule cites Executive Order 12333, which relates to intelligence activities.

We used several sources to look up legal texts. For U.S. Code citations, we used Cornell University Law School's U.S. Code Collection[2]. We used the Popular Name Tool maintained by the U.S. Office of the Law Revision Counsel[3] to lookup laws in the "Acts Cited by Popular Name" table. For citations to the Code of Federal Regulations, we used the e-CFR[4] maintained by the U.S. Government Printing Office. For Executive Orders, we used two resources: the American Presidency Project[5] hosted by the University of California, Santa Barbara and The Codification of Presidential Proclamations and Executive Orders[6] at the U.S. National Archives. The later resource only contains Executive Orders issued between April 13, 1945 and January 20, 1989, requiring us to use the American Presidency Project for Executive Orders issued outside this date range.

III. RESEARCH DESIGN

In this section, we describe our case study design.

A. *Research Question*

We seek to answer the following research question in our case study:

> **RQ**: What specific challenges do cross-references present for compliance in requirements engineering?

B. *Units of Analysis*

In our case study, the unit of analysis is cross-references. When specifying compliance requirements for software systems, engineers must begin with relevant legal texts [28]. For example, HIPAA governs healthcare systems and Electronic Health Records (EHRs) whereas the Gramm-Leach-Bliley Act (GLBA) governs financial systems. As engineers analyze relevant legal texts, they often encounter cross-references to portions of legal texts that sometimes go unanalyzed for a variety of reasons. Thus, our selection criteria for examining a cross-reference in our study is:

> Does the cross-reference require engineers to analyze a portion of a legal text that would otherwise be unanalyzed?

Figure 1 displays the types of cross-references encountered in our case study. The white rectangles are legal texts—a named legal document. Shaded rounded rectangles are portions of legal texts—discrete legal citations—that are under analysis. Circles represent legal statements and arrows represent cross-references. In Figure 1, the Pattern-A cross-reference and the Pattern B cross-reference from (4) to (5) in Pattern-B are internal cross-references [23]. As discussed in Section II, because prior work has examined internal cross-references [5, 24, 27], we do not examine Pattern-A or Pattern-B cross-references in this case study. Instead, herein, we examine the Pattern-C and Pattern-D cross-references (see Figure 1). Pattern-C represents an external cross-

[2] http://www.law.cornell.edu/uscode/
[3] http://uscode.house.gov/popularnames/popularnames.htm
[4] http://ecfr.gpoaccess.gov/cgi/t/text/text-idx?c=ecfr&tpl=%2Findex.tpl
[5] http://www.presidency.ucsb.edu/index.php
[6] http://www.archives.gov/federal-register/codification/numeric-executive-orders.html

reference—a reference between portions of different legal texts—as classified by Massey et al. [23]. In Pattern-D, the cross-reference points to another legal text portion; in this case, requirements engineers have typically not analyzed the legal statement (see (9) in Figure 1—a cross-reference from a legal statement in the HIPAA Privacy Rule to a legal statement in the HIPAA General Administration Requirements.

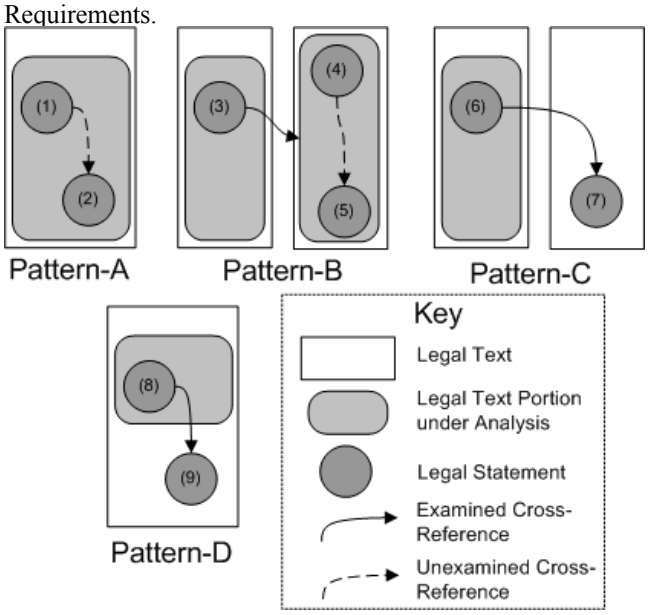

Figure 1. Possible Cross-References

C. Research Methodology

The HIPAA Privacy Rule text, available from the U.S. HHS website[7], served as our source material. Namely, we examine both §164.500-532 and §160.103, which define terms used throughout HIPAA. Given this text, we scanned the entire Privacy Rule to identify the 108 Pattern-C and Pattern-D cross-references within it. For each identified cross-reference, we analyzed the text indicated by the reference to identify new cross-references within those additionally referenced texts. Due to time constraints, we limited our analysis to those cross-references that represent a "distance" of no more than two cross-references away from the Privacy Rule. Within the referenced texts, we identified an additional 69 Pattern-C and Pattern-D cross-references, resulting in 177 total examined cross-references.

Figure 2 graphically represents the Pattern-C and Pattern-D cross-references in HIPAA Privacy Rule. The directional arrows in Figure 2 reflect that at least one cross-reference exists from the referencing legal text to the referenced legal text (both represented as rounded boxes). We use grounded theory analysis [16, 17] to classify cross-references and the impact they have on compliance requirements. In grounded theory analysis, theory is developed from the systematic study of a data set [16, 17]. The developed theory is "grounded" in the data, in that it is applicable only to the given data set [16, 17]. Future studies will allow us to make claims about the generalizability of our results. Grounded theory contrasts with the traditional scientific method, where hypotheses are formulated then tested through experiments. Researchers have previously used grounded theory analysis for requirements engineering research [11, 21] and when analyzing legal and policy requirements [1, 6, 7].

We performed two passes through the HIPAA Privacy Rule. In the first pass, we scanned it and identified Pattern-C and Pattern-D cross-references. In the second pass, we used open coding—tagging each unit of analysis with a descriptive categorization—to classify each cross-reference's effect on compliance requirements. We then followed each cross-reference, and performed the same procedure on the target legal text. Upon classifying each cross-reference, we compiled the classifications into the taxonomy presented below.

IV. Results

Each Pattern-C or Pattern-D cross-reference within the HIPAA Privacy Rule either: (a) introduces a conflicting requirement or definition; (b) refines an existing requirement; or (c) falls outside the software system's scope. Analyzing these cross-references facilitates refinement early in the software development process by enabling requirements engineers to address conflicting requirements that may otherwise thwart legal compliance, and ensures that engineers do not overlook important compliance requirements.

As a result of our case study, we developed a legal cross-reference classification taxonomy (Table I). Requirements engineers can use this taxonomy to classify the effect that a legal cross-reference has on existing compliance requirements. The taxonomy was developed in a descriptive fashion, and is now being proposed as a prescriptive taxonomy (for HIPAA) that will be further validated in future studies in other domains. Our taxonomy complements previous requirements engineering research; before we begin our cross-reference analysis, we assume that compliance requirements have been specified using one of the techniques described in Section II for specifying compliance requirements from legal texts [5, 9, 15, 23, 26, 27, 31, 36].

The six cross-reference types are: constraint, exception, definition, unrelated, incorrect, and general. *Constraint cross-references* add additional constraints on existing compliance requirements. *Exception cross-references* introduce an exception condition to an existing compliance requirement. *Definition cross-references* introduce a definition or term. *Unrelated cross- references* are those in

TABLE I. Legal Cross-Reference Taxonomy

Classification
Constraint
Exception
Definition
Unrelated
Incorrect
General

[7] http://www.hhs.gov/ocr/privacy/hipaa/administrative/privacyrule/adminsimpregtext.pdf

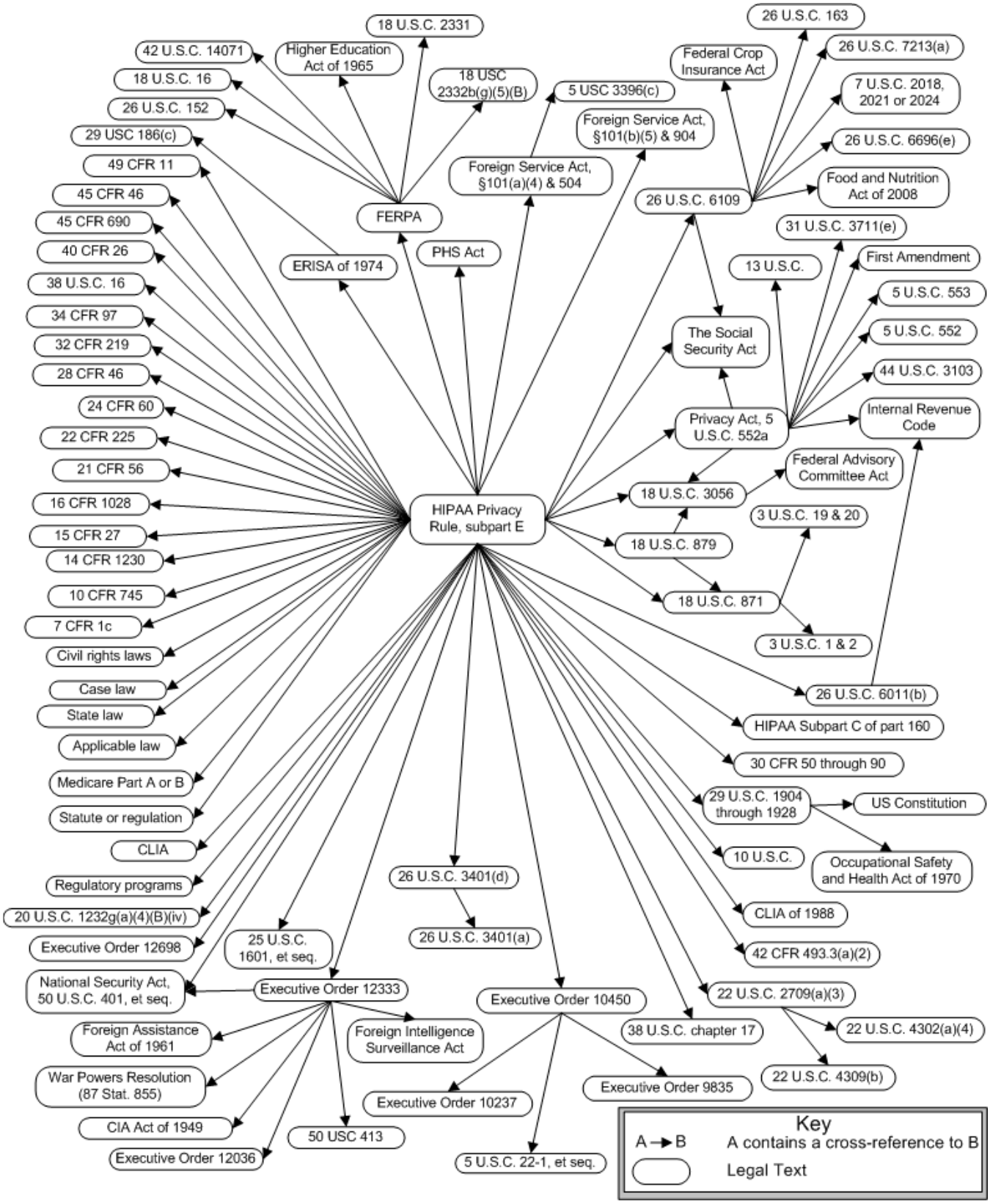

Figure 2. Unique Pattern-C and Pattern-D Cross-References in the HIPAA Privacy Rule

which the referencing or referenced legal texts do not yield requirements for software systems. *Incorrect cross-references* are references that cite an incorrect portion of a legal text. *General cross-references* do not cite a specific legal text, rather, they are citations to "applicable law". In the remainder of this section, we describe each cross-reference type in detail.

A. Constraint Cross-References

Requirements are often refined by disambiguating them. In our study, cross-references refine existing requirements because they add additional constraints. As advocated by Breaux and Antón for internal cross-references [6], we copy constraints from the referenced text into the compliance requirement. For example, §164.512(k)(3) of the HIPAA Privacy Rule states: "A covered entity may disclose PHI to authorized federal officials for the provision of protective services to the President or other persons authorized by 18 U.S.C. 3056." This paragraph contains a cross-reference to 18 U.S.C. 3056, a law that describes the authority and duties of the U.S. Secret Service. Among other duties, the Secret Service is tasked with protecting individuals such as the President, the Vice-President, their immediate families, former Presidents and their families, visiting heads of state, and Presidential candidates. The cross-reference refines the compliance requirement expressed by §164.512(k)(3). To perform this refinement, we copy the list of people the Secret Service is charged to protect into the requirement expressed by §164.512(k)(3). After refinement, the compliance requirement reads "A covered entity may disclose PHI to authorized federal officials for the provision of protective services to the President, the Vice-President, their immediate families, former Presidents and their families, visiting heads of state, and Presidential candidates."

B. Exception Cross-References

Some cross-references introduce exception conditions. For example, in §164.524(a)(1)(iii)(A), individuals are given the right to inspect and obtain a copy of their PHI, except for health information that is covered by the Clinical Laboratory Improvement Amendments (CLIA) of 1988 [8]. When exceptions are encountered, requirements engineers must create a requirement expressing the exceptional case [25]. In the given example, we create a requirement stating that a covered entity may withhold information covered by CLIA from the individual.

C. Definition Cross-References

Legal texts use cross-references to cite definitions from other laws in much the same way as a programmer imports object and function definitions from language libraries. For example, HIPAA does not redefine the definition of "medical care"; instead it cites the medical care definition used in the Public Health Services Act[9]. When we encounter a definitional cross-reference, we add the definition to the list of terms defined in the referencing legal text.

Terms spread across multiple legal texts can have differing and sometimes contradictory definitions [28]. For example, the Privacy Rule cross-references the Privacy Act of 1974 at §164.524(a)(2)(iv). In HIPAA, an individual is defined as the "person who is the subject of PHI" (§160.103), whereas in the Privacy Act of 1974[10], an individual is defined as a "citizen of the United States or an alien lawfully admitted for permanent residence" (§522a(a)(2)). These definitions differ; the HIPAA Privacy Rule protects the privacy of groups that the Privacy Act does not, for example, visitors to the U.S. Requirements engineers must resolve these differing definitions or consult with legal domain experts to determine how to proceed.

D. Unrelated Cross-References

Cross-references can introduce referential ambiguity—portions of a cross-referenced text might not be applicable to software systems [5, 28]. In our study, we identify and set aside those cross-references that are unrelated to software systems. To determine which cross-references are unrelated, we ask the questions "Does the *referencing* legal text paragraphs introduce requirements for software systems?" and "Does the *referenced* legal text paragraphs introduce requirements for software systems?"

Some cross-references occur in portions of a *referencing* legal text that are outside the scope of a software system. For example, §164.512(i)(1)(i)(A) of the Privacy Rule states:

> A covered entity may use or disclose PHI for research, regardless of the source of funding for that research, provided that the covered entity obtains documentation that an authorization or waiver, in whole or in part, of the individual authorization required by §164.508 for use or disclosure of PHI has been approved by either: an institutional review board (IRB), established in accordance with 7 CFR 1c.107, 10 CFR 745.107, 14 CFR 1230.107, 15 CFR 27.107, 16 CFR 1028.107, 21 CFR 56.107, 22 CFR 225.107, 24 CFR 60.107, 28 CFR 46.107, 32 CFR 219.107, 34 CFR 97.107, 38 CFR 16.107, 40 CFR 26.107, 45 CFR 46.107, 45 CFR 690.107, or 49 CFR 11.107

Although the documentation, authorization, or waiver for a covered entity to use or disclose PHI for research may be tracked by a software system, the subsequent list of 16 cross-references addresses how an institutional review board (IRB) is to be established. This establishment, as prescribed by these cross-references, is clearly outside the scope of software systems. Thus, we perform no further analysis on such cross-references.

In the case of a *referenced* text, if a legal statement cannot be operationalized as a software requirement, we set it aside and perform no further analysis on it. For example, the Privacy Rule, at §164.512(b)(1)(v)(C), cross-references 29 CFR 1904 through 1928. This referenced text regulates safety and health in the workplace. These regulations specify many rules related to various industries, some of which are not related to software systems. For example, 29 CFR 1910.25 regulates the type of portable wooden ladders that

[8] http://www.cms.gov/clia/
[9] 42 U.S.C. 300gg
[10] 5 U.S.C. 552a

can be used in the workplace, whereas 29 CFR 1912a establishes procedures for meetings of the National Advisory Committee on Occupational Safety and Health. Both of these references are unrelated to software systems governed by HIPAA, thus, we do not analyze such cross-references.

E. *Incorrect Cross-References*

Cross-references in legal texts may be erroneous. For example, §164.512(k)(3) states "A covered entity may disclose PHI to authorized federal officials for the provision of protective services to […] foreign heads of state or other persons authorized by 22 U.S.C. 2709(a)(3)." This text contains a cross-reference to 22 U.S.C. 2709(a)(3), part of which states: "special agents of the Department of State and the Foreign Service may protect and perform protective functions directly related to maintaining the security and safety of foreign missions (as defined in section 4302(a)(4) of this title)." This paragraph contains another cross-reference to the definition of "foreign missions" at 22 U.S.C. 4302(a)(4). However, the definition at this citation is "real property", not foreign missions. Foreign missions is defined in 22 U.S.C. 4302(a)(3). This is an obvious error. Although the error is documented via a footnote in the U.S. Code at 22 U.S.C. 2709(a)(3), the footnote was not immediately obvious to us and policy makers have yet to correct the legal text.

F. *General Cross-References*

May et al. [27] note that some cross-references do not mention a specific legal text by name. These cross-references are often stated as "other law", "state law", or "applicable law." For example, §164.502(g)(2) in the Privacy Rule allows covered entities to treat a parent as a representative of a minor "if, under applicable law, a parent […] has authority to act on behalf of an individual who is an unemancipated minor." No law is explicitly stated; instead, it is a general cross-reference to any applicable law. Requirements engineers are likely to require assistance from child law experts in resolving this/similar cross-references.

V. Discussion

Our analysis took approximately 32 man-hours and surfaced 5 critical conflicts that if not resolved would lead to non-compliance. This includes analyzing the 159 cross-references and developing the cross-reference taxonomy. As previously mentioned, to our knowledge, this is the first attempt to study the effects of cross-references on legal software requirements as we now discuss.

A. *Cross-References in the HIPAA Privacy Rule*

Table II displays the number of Pattern-C and Pattern-D cross-references by type that we identified in our study. Table II includes both the 108 external cross-references in the HIPAA Privacy Rule as well as the 69 cross-references identified by recursively applying our approach on the other legal texts that the Privacy Rule references. Of the 58 unrelated cross-references, 47 came from cross-references relating to the formation of an IRB, which is clearly beyond the scope of software systems (see Section IV.D).

TABLE II. Results from Applying the Taxonomy

Reference Type	Count
Refine	51
Exception	18
Definition	30
Unrelated	58
Incorrect	2
General	18
Total	**177**

B. *Identifying Conflicting Requirements*

Compliance requirements conflict when the requirements differ and may contradict each other. As previously discussed, cross-references introduce challenges to regulatory compliance [2, 5, 6, 7, 18, 27, 28], but researchers have yet to examine cross-references that introduce conflicting compliance requirements. For example in the HIPAA Privacy Rule, PHI must be retained by a covered entity for six years from the date when it was last in effect (§164.530(j)(2)), whereas in the Privacy Act of 1974, the information must be retained for five years or the life of the record, whichever is longer (§552a(c)(2)). Covered entities that must comply with both of these regulations, for example, a U.S. Department of Veteran's Affairs hospital, may be noncompliant if they focus on the five year minimum in the Privacy Act rather than the six year minimum under HIPAA. Analyzing such cross-references helps requirements engineers identify conflicting compliance requirements.

In our case study, we identified five sets of conflicting requirements (see Table III). Although five may not seem like a significant number, the conflicts are critical because they can lead to non-compliance. A method for identifying conflicts also can avoid the cost of building software systems that must later be re-engineered once a conflict is discovered, at greater expense. If requirements engineers limit their efforts to only examining the HIPAA Privacy Rule for legal compliance, the requirements they specify may be noncompliant with other laws, as in the case of the Privacy Act of 1974 and 29 CFR 1910.1020 example above.

Otto et al. note that definitions may conflict in regulations [28]. The HIPAA Privacy Rule references 30 definitions from other legal texts (see Table II). We identified one conflicting definition in our analysis—a conflict between the definitions of individual in the HIPAA Privacy Rule and the Privacy Act of 1974 as discussed in Section IV.C. We plan further studies to evaluate whether definitions do indeed conflict as much as previously thought [28] or if cross-references serve to reduce conflicts in laws (i.e. by creating a consistent definition across more than one legal text). Lamsweerde et. al outline a set of heuristics for identifying goal conflicts [34]. We employ these heuristics to identify conflicting legal requirements. These heuristics are summarized below [34]:

- Safety goals may conflict with satisfaction goals
- Goals that state information must remain confidential may conflict with goals that state the information should be shared. This heuristic helped us identify Conflict #4.
- Goals that optimize a value can conflict. This heuristic identifies Conflicts #1 and #2.

TABLE III. CONFLICTING REQUIREMENTS

Index	Conflicting Legal Texts	Summary of Conflict	Applicable Resolution Strategies
1	• HIPAA §164.530(j)(2) • Privacy Act of 1974 (cited at §164.524(a)(2)(iv)) • 29 CFR 1910.1020 (cited at §164.512(b)(1)(v)(C))	Length of data retention: • HIPAA: at least 5 years • Privacy Act: at least 6 years or the life of the record, whichever is longer • 29 CFR 1910.1020: at least 30 years if the employee worked for longer than a year	• Comply with most restrictive law • Keep data separate
2	• HIPAA §164.524(b)(2) • Privacy Act of 1974 (cited at §164.524(a)(2)(iv)) • 29 CFR 1910.1020 (cited at §164.512(b)(1)(v)(C))	Length of time an organization has to respond to a request for access to data: • HIPAA: in fewer than 30 days • Privacy Act: in fewer than 10 days • 29 CFR 1910.1020: in fewer than 15 working days	• Comply with most restrictive law • Keep data separate
3	• HIPAA §164.524(c)(4) • 29 CFR 1910.1020 (cited at §164.512(b)(1)(v)(C))	Under HIPAA, a covered entity may charge a reasonable, cost-based fee when providing copies of PHI to an individual , whereas in 29 CFR 1910.1020, employers must provide the first copy of an employees medical record free of charge	• Obligations supersede legal privileges • Keep data separate
4	• HIPAA §164.524(c)(4) • 29 CFR 1910.1020 (cited at §164.512(b)(1)(v)(C))	HIPAA and 29 CFR 1910.1020 contain different conditions that prevent the release of protected information to individuals. Even if an organization can withhold information under one law, they must release it under the other law.	Consult legal domain expert
5	• HIPAA §164.524(c)(4) • 29 CFR 1910.1020 (cited at §164.512(b)(1)(v)(C))	Under HIPAA, covered entities must de-identify health information before they release it, but under 29 CFR 1910.1020, they may release data to employees if personal identifiers cannot be removed.	• Do not exercise legal privileges • Keep data separate

- A goal that can have multiple instances can conflict by introducing competing goals among agents.
- Goals can conflict that have overlapping achieve and avoid constraints. We employ this heuristic to identify Conflicts #3 and #5.

C. *Resolving Conflicting Requirements*

In this section, we provide guidance to engineers for resolving conflicting compliance requirements. The strategies are descriptive in that they were developed based on our experiences in addressing the conflicts in Table III. Thus, the following strategies are based on our analysis to date and should be considered valid for the data set with which we worked [17]. Using the strategies described below, we were able to resolve four conflicts. The remaining conflict, #4 in, requires consultation with legal domain experts to resolve. We plan to identify additional strategies and further validate that our current strategies are applicable for resolving other kinds of conflicts that may exist in other legal domains such as for financial systems.

Multiple strategies may be used to resolve a given conflict. Engineers should select a conflict resolution strategy in conjunction with system stakeholders because the selected strategy may place additional requirements on the system or impact existing non-legal software requirements. We now discuss the conflict resolution strategies.

1) *Comply with the Most Restrictive Law*

Multiple regulations contain compliance requirements that govern the same kinds of software systems (e.g. EHR systems). When a compliance requirement expressed in one legal text is more restrictive than the corresponding compliance requirement expressed in another legal text, requirements engineers should choose to comply with the more restrictive of the two. To determine which legal text is more restrictive, each legal statement is classified as either:

- a ceiling rule, where the constraint is in the form "at least x years", or
- a floor rule, where the constraint is in the form "no more than y years"

Once each legal statement is classified, a simple matrix determines which legal statement is more restrictive (see Table IV). For example, consider Conflict #1 in Table III; each legal statement is a *ceiling rule*, therefore, the more restrictive option is to retain data for 30 years. Similarly, consider Conflict #2; each legal statement is a *floor rule*, thus, it can be resolved by responding to all requests within 10 days.

2) *Store Data Separately*

Different laws apply to different sets of records. For instance, consider Conflict #1 in Table III. HIPAA applies to records held by covered entities, whereas the Privacy Act applies to PII held by agencies of the federal government.

TABLE IV. RESOLVING TIME PERIOD CONFLICTS

		Legal Text 1	
		Ceiling rule	*Floor rule*
Legal Text 2	*Ceiling rule*	Comply with longer time period	Consult legal domain expert
	Floor rule	Consult legal domain expert	Comply with shorter time period

Likewise, 29 CFR 1910 applies to employee records. Recognizing the different scope of laws can permit compliance by holding records separately. For instance, employee records can be maintained in a separate database, and retained for 30 years independently of the database with HIPAA records, which must be retained for 6 years. Alternatively, the different kinds of data can be tagged using a markup; business rules can be developed for retaining the data elements for different time periods. If individual data elements are covered by conflicting legal requirements, this strategy should not be used. Requirements engineers should employ other strategies outlined in this section.

3) Obligations Supersede Legal Privileges

An obligation is an action that an actor is required by law to perform, whereas a privilege is an action an actor may perform but is not obligated to perform [20]. Legal texts denote obligations using natural language phrases such as "must" and denote privileges using phrases such as "may" [24]. Conflicts between obligations and privileges can be resolved by not exercising legal privileges. The conflict is resolved by performing the obligated action instead of the privileged action because an obligation trumps a privilege due its priority. Consider Conflict #3 in Table III, under HIPAA, covered entities have the privilege to charge a fee for copies of PHI, whereas under 29 CFR 1910.1020, employers are obligated to not charge for the first copy of an employee's medical record. This conflict can be resolved by specifying requirements to not charge for the first copy of PHI—complying with the obligation and not exercising the privilege. Likewise, engineers can resolve Conflict #5 by not releasing data that has not be de-identified.

4) Consult Legal Domain Experts

Some conflicts cannot be resolved with our current set of conflict resolution strategies. For example, consider Conflict #4 in Table III. Both the HIPAA Privacy Rule and 29 CFR 1910.1020 mandate that individuals and employees have access to their health information, respectively. Both regulations also provide mutually exclusive conditions under which a covered entity or employer can withhold information from an individual or employee. For example, HIPAA allows covered entities to withhold psychotherapy notes from an individual. An employer may deny employees' direct access to their medical records, if the medical record contains a diagnosis of a terminal or psychiatric illness (but they may be required to release the information to a third party such as the employee's primary physician). Thus, even if an organization has the privilege to withhold health information under one law, they are obligated under the other law to release it.

Using the obligations supersede legal privileges strategy, Conflict #4 could be resolved, in theory, by always releasing the information. However, releasing the information could be unethical or encourage healthcare professionals to violate professional codes of conduct if they believe the release will bring harm on someone. In this case, the obligations supersede legal privileges strategy does not adequately resolve the conflict, and requirements engineers should seek legal domain experts to assist in determining the priority between these conflicting compliance requirements. Engineers may have to consult multiple subject area experts, for instance, a tax law expert may be unable to address questions about Social Security law.

VI. Threats to Validity

When designing any case study, care should be taken to mitigate threats to validity. We make no causal references as a result of our study, so internal validity is not a concern [35]. External validity is the ability of a case study's findings to generalize to broader populations [35]. We use grounded theory analysis, so our cross-reference taxonomy is currently applicable to the 159 external cross-references we examined. We will refine and validate our taxonomy in future studies in different domains.

Construct validity addresses the degree to which a case study is in accordance with the theoretical concepts used [35]. Three ways to reinforce construct validity are: use multiple sources of reliable evidence; establish a chain of evidence; and have key informants review draft case study reports [35]. To establish a chain of evidence, we carefully documented the cross-reference classifications when performing our analysis; these classifications became the cross-reference taxonomy in Section IV. Finally, our draft case study report was reviewed by several ThePrivacyPlace members as well as by the law professor co-author who was a senior manager in drafting the HIPAA Privacy Rule.

Reliability is the ability to repeat a study and observe similar results [35]. To reinforce our study's reliability, we carefully document each cross-reference, the citing text, and its classification using our grounded theory approach.

VII. Summary

We have presented a taxonomy of legal cross-reference types. Engineers can use it to classify the effect of legal cross-references on compliance requirements. We developed this taxonomy based on a case study of the HIPAA Privacy Rule. We analyzed 177 total cross-references, which contained five sets of conflicting compliance requirements. To the best of our knowledge, we are the first to identify concrete examples of conflicting compliance requirements due to cross-references. Finally, we recommend strategies to resolve conflicts among compliance requirements.

We developed our cross-reference taxonomy through a grounded theory analysis of the HIPAA Privacy Rule cross-references, thus the taxonomy is currently valid for the HIPAA Privacy Rule only. We plan further studies using other legal texts to refine and further validate the taxonomy. In addition, we plan human subject experimentation to

measure the taxonomy's affect on requirements engineers' ability to classify cross-references and identify conflicts.

Requirements engineers need the ability to manage the evolution of the law across cross-references. For instance, consider a cross-reference that refines existing requirements by introducing constraints (see Section IV.A). When policy makers formulate changes to the referenced legal text, engineers need tools and techniques to update the requirements that were refined with the previous version of the text. In addition, HIPAA is being updated as a result of the HITECH Act—we plan to update our study using the new version of the regulation. Finally, we plan to integrate cross-reference analysis into production rule modeling [24, 26], which currently only codifies the internal cross-references.

ACKNOWLEDGMENT

This work was partially funded by NSF ITR grant #0325269 and NSF Science of Design Grant # 0725144. We thank the members of ThePrivacyPlace for their comments.

REFERENCES

[1] A.I. Antón, J.B. Earp, "A Requirements Taxonomy for Reducing Web Site Privacy Vulnerabilities", *Requirements Engineering Journal*, 9(3), 2004, pp. 169-185.

[2] T.J.M. Bench-Capon, G.O. Robinson, T.W. Routen, M.J. Sergot, "Logic Programming for Large Scale Applications in Law: A Formalisation of Supplementary Benefit Legislation", *1st Intl. Conf. on AI and Law*, 1987, pp. 190-198.

[3] B. Berenbach, D. Gruseman, J. Cleland-Huang "Application of Just In Time Tracing to Regulatory Codes", *8th Conf. on Systems Engineering Research*, 2010.

[4] B. Boehm, H. In, "Identifying Quality-Requirements Conflicts", *IEEE Software*, 13(2), 1996, pp. 25-35.

[5] T.D. Breaux. *Legal Requirements Acquisition for the Specification of Legally Compliant Information Systems*, Ph.D. Thesis, NCSU, 2009.

[6] T.D. Breaux, A.I. Antón, "Analyzing Regulatory Rules for Privacy and Security Requirements", *IEEE Trans. on Software Engineering*, 34(1), Jan.-Feb. 2008, pp. 5-20.

[7] T.D. Breaux, M.W. Vail, A.I. Antón, "Towards Regulatory Compliance: Extracting Rights and Obligations to Align Requirements with Regulations", *14th IEEE Intl. Requirements Engineering Conf.*, 2006, pp. 46-55.

[8] Cholvy, "Checking Regulation Consistency by using SOL-Resolution", *7th Intl. Conf. on AI & Law*, 1999, pp. 73-79.

[9] J. Cleland-Huang, A. Czauderna, M. Gibiec, J. Emenecker, "A Machine Learning Approach for Tracing Regulatory Codes to Product Specific Requirements", *32nd IEEE Intl. Conf. on Software Engineering*, 2010.

[10] M.L. Cohen, K.C. Olson, *Legal Research*, West, 2000.

[11] D.E. Damian, D. Zowghi, "Requirements Engineering Challenges in Multi-site Software Development Organizations", *Requirements Engineering Journal*, 2003, pp. 149-160.

[12] S. Easterbrook, B. Nuseibeh, "Managing Inconsistencies in an Evolving Specification", *Proc. of the 2nd IEEE Intl. Symposium on Requirements Engineering*, 1995, pp. 48-55.

[13] W. Emmerich, A. Finkelstein, C. Montangero, S. Antonelli, S. Armitage, R. Stevens, "Managing Standards Compliance", Trans. on Software Engineering, 25(6), Nov./Dec. 1999, pp. 836-851.

[14] 2010 Global Information Survey, Ernst & Young, 2010.

[15] S. Ghanavati, D. Amyot, L. Peyton, "Compliance Analysis Based on a Goal-Oriented Requirement Language Evaluation Methodology", Proc. of the 17th IEEE Intl. Conf. on Requirements Engineering, 2009, pp. 133-142.

[16] B.G. Glaser, Theoretical Sensitivity, Sociology Press, 1978.

[17] B.G. Glaser, A.L. Strauss, The Discovery of Grounded Theory, Aldine Transaction, 1967.

[18] M. Hamdaqa, A. Hamou-Lhadj, "Citation Analysis: An Approach for Facilitating the Understanding and the Analysis of Regulatory Compliance Documents," *6th Intl. Conf. on Information Technology: New Generations*, 2009, pp. 278-283.

[19] H.M. Hart, Jr., H. Wechsler, R.H. Fallon, Jr., J.F. Manning, D.J. Meltzer, D.L. Shapiro, *The Federal Courts and the Federal System*, 6th ed., 2009.

[20] W.N. Hohfeld, "Some Fundamental Legal Conceptions as Applied in Judicial Reasoning", *The Yale Law Journal*, 23(1), 1913, pp. 16-59.

[21] L. Karlsson, A.G. Dahlstedt, B. Regnell, J. Natt och Dag, A. Persson, "Requirements Engineering Challenges in Market-Driven Software Development-An Interview Study with Practioners", *Information and Software Technology,* 49, 2007, pp. 588-604.

[22] B. Krebs, "ChoicePoint Breach, Exposed 13,750 Consumer Records", *The Washington Post*, Oct. 19, 2009.

[23] A.K. Massey, P.N. Otto, A.I. Antón, "Prioritizing Legal Requirements", *2nd Intl. Workshop on RE and Law*, 2009.

[24] J.C. Maxwell, A.I. Antón, "Developing Production Rule Models to Aid in Acquiring Requirements from Legal Texts", 17th *Intl. IEEE Requirements Engineering Conf.*, 2009, pp. 101-110.

[25] J.C. Maxwell, A.I. Antón, "A Refined Production Rule Model for Aiding in Regulatory Compliance", NCSU Technical Report TR-2010-3, 2010, ftp://ftp.ncsu.edu/pub/unity/lockers/ftp/csc_anon/tech/2010/TR-2010-3.pdf.

[26] J.C. Maxwell, A.I. Antón, "The Production Rule Framework: Developing a Canonical Set of Software Requirements for Compliance with Law", *1st ACM Intl. Health Informatics Symposium*, 2010.

[27] M.J. May, C.A. Gunter, I. Lee, "Privacy APIs: Access Control Techniques to Analyze and Verify Legal Privacy Policies", *19th IEEE Computer Sec. Foundations Workshop*, 2006, pp. 85-97.

[28] P.N. Otto, A.I. Antón, "Addressing Legal Requirements in Requirements Engineering", *15th IEEE Intl. Requirements Engineering Conf.*, 2007, pp. 5-14.

[29] P.N. Otto, A.I. Antón, D.L. Baumer, "The ChoicePoint Dilemma: How Data Brokers Should Handle the Privacy of Personal Information", *IEEE Security & Privacy*, 5(5), 2007, pp. 15-23.

[30] W.N. Robinson, S. Fickas, "Supporting Multi-Perspective Requirements Engineering", *1st IEEE Intl. Requirements Engineering Conf.*, 1994, pp. 206-215.

[31] A. Siena, J. Mylopoulos, A. Perini, A. Susi, "Designing Law-Compliant Software Requirements", *28th Intl. Conf. on Conceptual Modeling*, 2009.

[32] A.K. Thurimella, B. Bruegge, "Evolution in Product Line Requirements Engineering: A Rationale Management Approach", *15th IEEE Requirements Engineering Conf.*, 2007, pp. 254-257.

[33] T.M. van Engers, M.R. Boekenoogen, "Improving Legal Quality: an Application Report", *9th Intl. Conf. on AI and Law*, 2003, pp. 284-292.

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

[35] R.K. Yin, *Case Study Research: Design and Methods*, in Applied Social Research Methods Series, Vol. 5, 2003, 3rd ed.

[36] J.D. Young, "Commitment Analysis to Operationalize Software Requirements from Privacy Notices", (in press) *Requirements Engineering Journal*, 2010.

[37] P. Zhang, L. Koppaka, "Semantics-Based Legal Citation Network", *11th Intl. Conf. on AI and Law,* 2007, pp. 123-130.