Reconciling Multi-jurisdictional Legal Requirements: A Case Study in Requirements Water Marking

David G. Gordon
Engineering and Public Policy
Carnegie Mellon University
Pittsburgh, United States of America
dggordon@cmu.edu

Travis D. Breaux
Institute for Software Research
Carnegie Mellon University
Pittsburgh, United States of America
tdbreaux@cs.cmu.edu

***Abstract*—Companies that own, license, or maintain personal information face a daunting number of privacy and security regulations. Companies are subject to new regulations from one or more governing bodies, when companies introduce new or existing products into a jurisdiction, when regulations change, or when data is transferred across political borders. To address this problem, we developed a framework called "requirements water marking" that business analysts can use to align and reconcile requirements from multiple jurisdictions (municipalities, provinces, nations) to produce a single high or low standard of care. We evaluate the framework in an empirical case study conducted over a subset of U.S. data breach notification laws that require companies to secure their data and notify consumers in the event of data loss or theft. In this study, applying our framework reduced the number of requirements a company must comply with by 76% across 8 jurisdictions. We show how the framework surfaces critical requirements trade-offs and potential regulatory conflicts that companies must address during the reconciliation process. We summarize our results, including surveys of information technology law experts to contextualize our empirical results in legal practice.**

Keywords-legal requirements, requirements comparison, requirements reconciliation, conflicts

I. INTRODUCTION

Emerging information systems increasingly leverage third-party data processing and storage. These third party services provide economies of scale that allow companies with minimal infrastructure to provide rich consumer experiences at relatively low cost. The emerging commodification of software as a service amplifies this phenomenon: Google Maps, Facebook, LinkedIn, and PayPal provide mapping, social network, and payment-processing services, to name a few. Composing software from services in this new ecosystem amplifies an old challenge: *how do business analysts identify those system requirements that govern their software in the presence of trans-border data flows?* This problem has received attention from government and industry with regards to privacy and security regulation [16, 18, 22]. Small companies and start-ups, in particular, frequently lack the resources to resolve this issue through legal guidance alone.

Consider an example scenario with data transfers across multiple jurisdictions in the United States. A New York State resident, while visiting relatives in Nevada State, accesses an online web account she has with a Wisconsin-based business. The business stores her data using a cloud service provider (CSP) that maintains the data in their Connecticut State facility. Each "step" in this data flow must address provincial laws that govern data access, retention, and breach notification. The laws are triggered by legal conditions, such as the geographical location of the business and data (Wisconsin, Nevada and Connecticut), as well as the location and legal residence of the data subject (Nevada, New York). These laws are written in semi-isolation: in some cases borrowing requirements from other jurisdictions, which was often the situation in U.S. data breach notification laws; in other cases competing with other jurisdictions by "racing" to the top or bottom[1] of best practice, such as the recent India privacy regulations that established stronger consent requirements than the European Union. While our examples in this paper are limited to U.S. regulations, the scope of this problem affects many industrialized countries worldwide.

We introduce an empirically validated framework that business analysts can use to reconcile regulatory requirements from multiple jurisdictions into a single standard of care. This reconciliation method, called requirements water marking[2], allows an analyst to establish a high or low water mark standard across two or more jurisdictions. The framework preserves traceability so that a business analyst can trace observed similarities and differences from requirements to specific sentences and phrases in the law. The collection of requirements produced by the framework can then be further evaluated by legal counsel and experts familiar with regional legal practices.

We developed and validated our framework by analyzing U.S. data breach notification laws. These laws have been enacted during the past eight years and have effectively created a U.S. nationwide information system that sends messages (notices) to consumers and regulatory agencies when a company discovers a breach of consumer data. While these laws support legacy systems for sending notices (e.g., telephone, postal mail, etc.), they also permit using electronic notices and many describe functional security requirements.

The remainder of this paper is organized as follows: in Section II, we discuss related work; in Section III, we introduce the framework, including the new water mark method; in Section IV, we introduce our case study design that we used to validate our framework; in Section V we discuss our summary findings, with a discussion of multi-jurisdictional conflicts discovered by the process presented in Section VI;

[1] In U.S. law, regulators who seek to establish the highest legal standard are observed to be in a "race to the top"

[2] A water mark refers to that line on the shore established by the fluctuations in water, or in this paper, by fluctuations in requirements. This is not to be confused with watermarks embedded in a document to demonstrate authenticity.

in Section VII we discuss threats to validity; in Section VIII, we report on interview with legal experts who reviewed our framework, with summary and future work in Section IX.

II. Related Work

The role of regulations in legal requirements has been a continuing topic of research [19]. We consider three related work topics: techniques for extracting requirements from legal texts, methods for comparing requirements to find similarities and differences, and research on the legal requirements semantics that have logical implications for reconciling differences across legal requirements sets. We note differences between our contribution and prior work.

Regulations and laws often conform to a stylized subset of natural language. Breaux introduced a frame-based method for systematically extracting requirements from legal texts [5]. The method includes validated phrase heuristics and a legal ontology that significantly improve requirements extraction by human analysts over traditional methods ($p < 0.001$) [5]. Based on this method, Breaux and Gordon developed a legal requirements specification language (LRSL) to assist analysts with the framework by formatting extracted requirements in a standard notation [4]. Herein, we reuse the LRSL to encode regulations as inputs to our multi-jurisdictional analysis framework.

In order to compare requirements across jurisdictions, analysts must compare textual requirements pairs to identify similarities and differences. Prior work to automatically identify equivalent requirements includes research in applied information retrieval (IR) [9, 24] and machine-learning [15]. Falessi et al. conducted an empirical evaluation of multiple IR-based NLP techniques to identify equivalent requirements pairs [9]. The evaluation compares different algebraic models, weighting and similarity metrics, and term extraction methods. The results found the "ideal" best technique is a vector-space model with the Cosine similarity metric, linear weighting and a Stanford part-of-speech noun and verb extractor. We evaluated this technique on our dataset and discuss the results in Section IX. Enhancements to IR-based techniques, such as project glossaries [24] and machine-learning [12, 15], or multi-word abstractions [11], may provide better automation to assist analysts with this step in our process. In particular, machine-learning methods that rely on training sets [15] are likely to show promise in multi-jurisdictional analysis over successive jurisdictions when comparing regulations from the same domain. We discuss this issue of scalability in Section V.

Dekhtya et al. studied human performance in tracing requirements to system tests [8]. They found that no single analyst was able to achieve the *gold standard*, which was the ideal solution, whereas the combined effort of all analysts did find all traces in the standard. We believe tracing requirements to test cases (or source code) is a conceptually different problem than comparing textual requirement pairs for similarity. To assist human analysts, we developed metrics that measure types of differences between requirements [3]. These metrics are used to measure terms and phrases that conceptually subsume the meanings of other terms and phrases, or dissimilar phrases that correspond to changes in modality (must, should, may). In addition to creating a "link" between two similar requirements, these metrics lead an analyst to rationalize and explain the similarity or difference that they observe.

Maxwell and Anton introduce a taxonomy of legal cross-references to identify conflicting requirements [17]. Cross-references are explicit phrases that appear in regulations and serve to link regulatory requirements within and across regulations. These links encode a specific semantic relationship, such as reusing a previously stated definition, conferring a priority to reconcile potential conflicts, or refining a requirement by describing required or recommended implementation strategies [5]. In our approach, we encode both explicit cross-references and implied links between requirements in our LRSL to identify relational dissimilarities. However, our metrics further identify phrase differences between requirements that are not encoded in cross-references. These comparisons are similar to work in model merging that examined inter- and intra-model properties before performing a merge [20].

III. Water Marking Framework

The water marking framework process overview appears in Figure 1 and consists of three steps performed manually by a human analyst. Arrows lead from inputs/outputs to each step, which are individually numbered: (1) the analyst extracts and encodes requirements from two regulatory documents S and R using a machine-readable LRSL that is parsed to yield itemized requirements; (2) the analyst conducts a gap analysis to compare requirements pairs across the two requirements sets to yield dissimilarity measures; and (3) the analyst applies the water marking constructs (union, disjoint, and minimum) to identify and reconcile consensus and conflict across these measures.

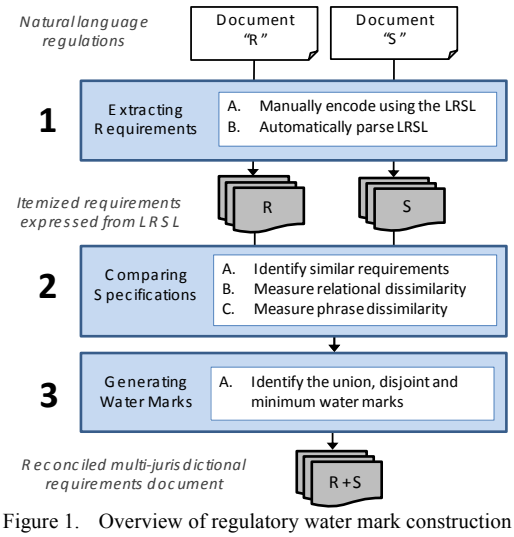


Figure 1. Overview of regulatory water mark construction

The framework combines and extends prior work to enable the water marking method. Step 1 is based on the frame-based requirements analysis method by Breaux [5], which is implemented in a legal requirements specification language (LRSL) to improve repeatability [4]. Step 2 extends metrics previously applied in an industry case study to compare

regulatory requirements with product requirements in the domain of accessibility [3]; in this paper we introduce a new metric (SE-PE) and validate the metrics in a new domain (data breach notification). Step 3 is a new method uniquely developed for the multi-jurisdictional requirements problem. We now briefly describe steps 1-2, before introducing step 3.

A. *Extracting Requirements*

In Step 1, the analyst translates a regulatory document into a legal requirements specification language (LRSL). The LRSL (see abridged example in Figure 2) serves to itemize legal requirements and maintain traceability to section and paragraph references in the original text. Definitions for terms-of-art, such as "data collector" are preceded by the equals sign (lines 1 and 2) and are later linked by the parser to uses in parsed requirements. Requirements begin with a left justified stakeholder role (see data collector, line 8), followed by one or more requirement clauses led by colons (lines 9, 11, 14, 17). To preserve context, requirements are linked to each other by relational keywords `REFINES`, `FOLLOWS`, or `EXCEPT` (line 13, 16, 18). For an extended LRSL discussion, see Breaux and Gordon [4].

```
1    data collector
2       = governmental agency
3           | institution of higher education
4           //...
5
6    SECTION 603A.210 //Security measures
7    PAR 1.
8    data collector
9       : maintains records which contain personal infor-
10        mation of a resident of this State
11      : shall disclose the breach of the security of the
12        system data to the resident of this State
13      FOLLOWS 1. #1
14      : must make the disclosure in the most expedient
15        time possible and without unreasonable delay...
16      REFINES 1. #2
17      : may delay the required notification
18      EXCEPT-TO 1. #3
```

Figure 2: Abridged LRSL Excerpt from Nevada §603A.210(1)

In the remaining paper, we present post-LRSL-processed requirements as text statements and graphs automatically generated by the LRSL parser. The corresponding graph for Figure 2 appears in Figure 3: nodes map to requirements, and arrows map to relations as follows: `REFINES` (solid line), `FOLLOWS` (finely dotted line), or `EXCEPT` (dashed line). Each requirement has a unique identifier: a shared label, e.g., the two-letter abbreviation NV for Nevada, and numerical index.

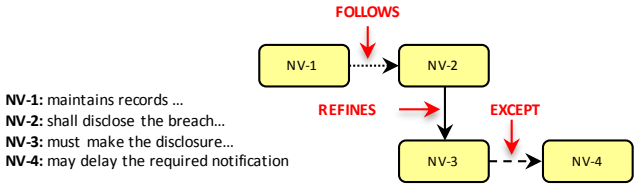


Figure 3: GraphML Representation of Nevada §603A.210(1)

Requirements described in legal texts may contain pre-conditions embedded in the statement. Step 2 requires the analyst to separate pre-conditions into implied permissions when they describe separate actions; a technique that we call *non-modal adaptation*. For example, consider the following excerpt from Connecticut §36a-701b(e)(4):

> The entity who demonstrates that the affected class of subject persons to be notified exceeds five hundred thousand persons may send notice using substitute notice.

The above excerpt maps to the LRSL in Figure 4, below, as follows: the underlined clause above is separated into a requirement clause (line 2) with the modal verb "may" and the annotation "implied-permission" (line 3) to indicate a non-modal adaptation produced this permission. Next, the instruction `PRECEDES` (line 4) indicates the prior requirement (line 2) is a pre-condition to the second requirement (line 5).

```
1    entity
2       : may demonstrate that the affected class of sub-
          ject persons to be notified exceeds five hun-
          dred thousand persons
3       ANNOTATE implied-permission
4       PRECEDES #2
5       : may send notice using substitute notice
```

Figure 4: Example of non-modal adaptation to map pre-conditions to implied permissions

B. *Comparing Specifications*

After encoding two regulations in the LRSL, the analyst performs a "gap analysis" using metrics to identify and rationalize similarities and differences between requirements pairs. For comparing two requirements A and B, the metrics in Table 1 are used; A refers to the first requirement, and B refers to the second requirement.

TABLE 1: REQUIREMENT COMPARISON METRICS

Metric	Metric Description
S-NE	(Near Equivalent): Requirements A and B are equivalent, with some portions of the requirements describing the same or a similar action
S-PE	(Pure Equivalency): Requirements A and B are equivalent and do not need further refinement through phrase metrics
P-G1	(Generalized Concept): The "phrase in B" describes a more general concept than the "phrase in A"
P-G2	(Missing Constraint): The "phrase in A" is missing from Requirement B
P-R1	(Refined Concept): The "phrase in B" describes a more refined concept than the "phrase in A"
P-R2	(New Constraint): The "phrase in B" is missing from Requirement A
P-M	(Modality Change): The "phrase in A" has a different modality than the "phrase in B"

The gap analysis is used to discover salient differences between two requirements sets. These differences occur *between* statements, called *relational dissimilarity*, and *within* statements, called *phrase dissimilarity*. Relational dissimilarity is measured when one requirement set contains a requirement not present in the other set (i.e., a requirement without an S-NE or S-PE metric) and phrase dissimilarity is measured when two near equivalent requirements (S-NE) are differentiated using the phrase-level metrics (e.g., P-G1, P-G2). If an organization operates information services in two jurisdictions governed separately by these requirements sets, resolving these differences is necessary to determine a single standard of care. For example, consider the following requirements from regulations CT and WI, respectively. Comparison of these requirements by the analyst yields the measurements shown in Figure 5.

CT-4: A person owns, licenses or maintains computerized data that contains personal information

WI-2: A person maintains or licenses personal information in this state

Measure	Phrase
S-NE (CT-4, WI-2) near equivalent	–
P-R1 (CT-4, WI-2) refined concept	"owns, licenses, or maintains" generalizes "maintains or licenses"
P-G1 (CT-4, WI-2) generalized concept	"personal information" generalizes "computerized data that contains personal information"
P-R2 (CT-4, WI-2) new constraint	"in this state" is missing from CT-4

Figure 5: Phrase-dissimilar Requirements from CT §36a-701b and WI §134.98

Because some portions of the requirements describe the same action, they are first asserted as being near equivalent (S-NE). Phrases in the requirements generalize one another; "owns, licenses, or maintains" is more general than "maintains or licenses," because it includes the extra action "owns" (P-R1) and "personal information" is a more general term than "computerized data containing personal information," because this data potentially contains other types of information (P-G1). Lastly, the P-R2 metric measures the new constraint "in this state" that does not appear in the CT-4.

C. Generating Water Marks

In prior work, we hypothesize that the differences made salient during a gap analysis could be generally resolved through three water mark techniques, called union, disjoint, and minimum [13]; in this paper, we implemented and evaluated this proposal. The union water mark technique yields a single practice from multiple jurisdictions, whereas the disjoint water mark technique maintains separate practices for each jurisdiction. The minimum water mark describes the lowest standard of care across multiple regulations. We now describe how to implement the water marks using the previously obtained measures.

1) Union Reconciliation. The union water mark consists of systematically merging requirements from multiple jurisdictions while addressing conflicts. The merger proceeds in two steps: (1) the analyst reviews the relational dissimilarities to identify requirements that are valid in both jurisdictions; and (2) the analyst merges phrase dissimilarities from two near-equivalent requirements to yield a single, combined requirement.

The analyst identifies relational dissimilar requirements by finding requirements in either requirement set that are not measured with S-NE or S-PE metrics. These requirements are reconciled by two techniques: *preservation*, which means practicing the requirement in both jurisdictions, and *omission*, or choosing to not practice a requirement in either jurisdiction. Preservation is typically applied to refinements linked by `REFINES` that describe how to implement a practice, or to post-conditions linked by `FOLLOWS` that describe follow-on permissions, obligations or prohibitions. In Figure 6, we preserve New York's (NY) requirement NY-25 to log notices in Connecticut's (CT) jurisdiction using a dashed-border node and maintaining the same refinement relation (a solid arrow). Omission is typically applied to exceptions linked by `EXCEPT` that appear in one jurisdiction and not another. In Figure 6, the omission of Mississippi's (MS) requirement MS-23 appears as a red cross through a node. The key in Figure 6 applies to subsequent figures in this paper.

The intuition for preservations is that relational dissimilar requirements linked with `REFINES` or `FOLLOWS` are subtasks, quality attributes, or additional tasks an organization performs to achieve compliance with one jurisdiction and that compliance with these requirements is permissible in another jurisdiction where they have no observed conflicts. This may incur an additional burden for those transactions covered by the second jurisdiction, but it may also streamline an organization's business practices. Contrarily, relational dissimilar requirements linked using `EXCEPT` describe alternatives or optional requirements from one jurisdiction that do not appear in a second jurisdiction. Thus, practicing such exceptions in the second jurisdiction may lead to violating an near-equivalent obligation in that jurisdiction.

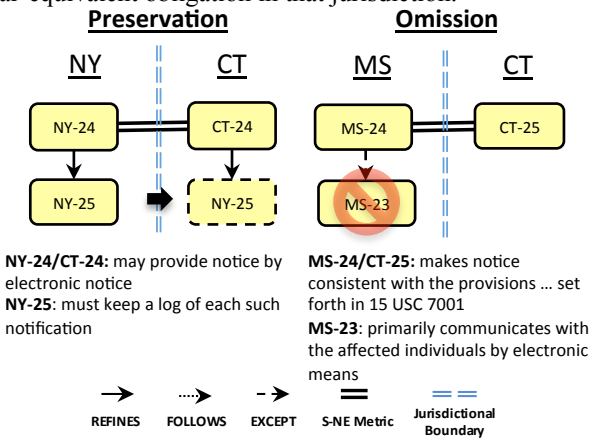


Figure 6: Union Reconciliation of Relational Dissimilarity Between MS-HB-583 and CT-36a-701b

Finally, the analyst merges near equivalent requirements by carefully combining dissimilar phrases into a single statement designed to encompass the details specified by both requirements. To facilitate this process, we developed heuristics (see Table 2) based on the phrase metric type (concept, constraint, modal) as well as the phrase topic in question: the *subject* of the requirement, such as "a person or business;" the *action* or a quality of the action, such as "notify the attorney general" or "notify expeditiously;" or the *object* of the action, including to whom or for whom the action is performed, such as "affected residents." The heuristics in Table 2 are intended to "take the union" of the meanings of two phrases, effectively yielding a requirement that covers two previously separated sets of circumstances.

TABLE 2: HEURISTICS FOR UNION RECONCILIATION OF PHRASE-DISSIMILAR REQUIREMENTS

	Conceptual Measures *(P-G1, P-R1)*	**Constraint Measures** *(P-G2, P-R2)*	**Modal Measures** *(P-M)*
Subject	Preserve more general subject phrase	Preserve constrained subject	Preserve obligations over permissions (e.g. "shall" or "must" over "may")
Action	Preserve more specific action	Preserve constrained action	
Object	Preserve more general object phrase	Preserve less constrained object	

Applying the heuristics yields a single requirement that maintains the original legal text with changes that can be traced back to the selected measures and heuristics. Further, the analyst must be aware of negations in the text which require a special technique not discussed in this paper.

2) Minimum and Disjoint Reconciliation. For relational dissimilarity, the minimum water mark technique consists of "omitting" requirements from one jurisdiction that do not appear in another jurisdiction. Omissions are excluded from consideration for the affected system implementation. Alternatively, the disjoint technique preserves these requirements. For example, the NY data breach law §899-aa(8)(a) specifies that an organization shall notify the state attorney general and other state entities regarding the "timing, content, and distribution of the notices and approximate number of affected persons" following notification of the affected individuals (see Figure 7); CT's data breach law §36a-701b has no such requirement. If an organization chooses the minimum standard, they will follow CT's lower standard of care and not notify the state attorney general in either jurisdiction, as shown in Figure 6. Otherwise, the organization may keep their practices disjoint, and use a separate procedure to summarize the breach for NY residents and the NY state attorney general.

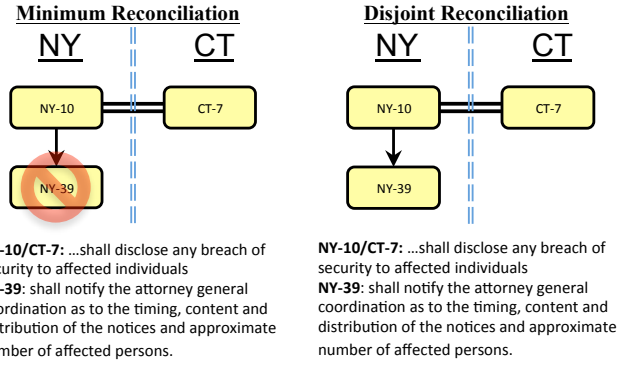


Figure 7: Minimum and Disjoint Reconciliation of Relational Dissimilarity between NY §899-aa and CT-36a-701b

The analyst applies the minimum technique to phrase-dissimilar requirements by omitting P-measured phrases that appear in one regulation but not another. In Figure 8, both CT and MS specify requirements covering entities that possess data on their consumers. However, nuances in the phrases affect the object that each requirement refers to:

CT-4: Person owns, licenses or maintains computerized data that contains personal information

MS-2: Person owns, licenses or maintains personal information…

Measure	Phrase
S-NE (CT-4, MS-2) (near equivalent)	–
P-G1 (CT-4, MS-2) (generalized concept)	"personal information" generalizes "computerized data that contains personal information"

Figure 8: Phrase-dissimilar Requirements from CT §36a-701b and MS HB-583

In Figure 8, adopting the minimum technique means preferring the more specific phrase from CT in the P-G1 measure "computerized data that contains personal information" for both jurisdictions over MS's more general phrase "personal information" that covers non-computerized information. The disjoint standard retains these requirements separately for data covered by each state.

D. Water Mark Chaining

The water mark method is a binary operation that accepts two sets of requirements and produces a single, reconciled requirement set for two jurisdictions. To analyze three or more jurisdictions, the analyst combines the output from two jurisdictions with the third requirements set using the same binary operation. These combinations produce "chains" that raise the question: is this process commutative? That is, does it matter which order we apply the operation over three or more jurisdictions to compute the outcome?

Consider an organization that has data on residents from three jurisdictions to which the organization is subject to their regulations: A, B, and C. Preferring to determine a single standard of care (if one exists), the organization's business analyst applies the water mark method. First, the analyst compares requirements sets A and B (denoted A/B) and generates the A-B water mark for the aggregate of two jurisdictions. The A-B water mark can then be reconciled with the requirements set C (denoted A-B/C), which reflects a comparison between the water mark A-B and regulation C and yields the A-B-C water mark. This left-associative notation is used throughout our paper to describe the order of operations. In Section IX, we discuss interview findings about how legal experts order jurisdictions in their analysis and our results from evaluating the commutative property.

IV. Case Study Design

We now discuss our case study design, including research questions, dataset selection criteria, units of analysis, and analysis procedure. To guide our research, we established the following research questions:

R1: What techniques exist to align requirements from multiple jurisdictions?

R2: How do these techniques scale?

Regarding question R1, we discovered that business and legal analysts presently lack a systematic method for comparing requirements across jurisdictions. To discover such a method, we employed grounded analysis [7], in which a theory is derived from a data set, and then we chose to evaluate the method using additional data sets and subject matter expert review. The selected data sets consist of U.S. data breach notification laws: these laws have been enacted across 46 U.S. states and territories from 2002-2011, each governing personal information about state residents. While the laws govern a common theme (data breach), they also vary considerably. We down-selected to eight regulations based on guidance from a legal expert with 7 years of privacy and security law expertise to highlight regulations that have been a priority for U.S. companies:

AR: Personal Information Protection Act, Arkansas Chapter 4.110, enacted 2005.

CT: Breach of Security Regarding Computerized Data Containing Personal Information: Connecticut Chapter 669, section 36a-701b. Enacted 2006.

MA: Security Breaches, Massachusetts Chapter 93H, enacted 207.

MD: Personal Information Protection Act, Maryland Subtitle 14-35. Enacted 2008.
MS: (No title given) Mississippi House Bill 583. Enacted 2011.
NV: Security of Personal Information, Nevada Chapter §603A. Enacted 2006.
NY: Notification of Unauthorized Acquisition of Personal Information, New York General Business Law § 899-aa. Enacted 2005.
WI: Notice of Unauthorized Access to Personal Information, Wisconsin §134.98. Enacted 2006.

All legal documents were mapped into the LRSL by the investigators (the authors), separately, and co-reviewed. The first author designed the reconciliation process with feedback from the second author to identify and address errors or concerns that arose throughout the process. The investigators kept a research notebook to record comments about unusual or notable artifacts in the translation; during comparison and reconciliation, a list of strategies was recorded to reflect how the investigator handled unusual cases, and upon acceptance of a new strategy, all previous resolutions were reviewed to ensure consistency across the dataset. A law expert was consulted on legal questions that arose during the process.

The units of analysis consist of the translated requirements and their relations as expressed in the LRSL and the measures of relational- and phrase-dissimilarity produced by the gap analysis. In the analysis procedure, we first compared definitions and then requirements between the regulations, applying the metrics outlined in Section III. After near- and pure-equivalencies were determined, we applied phrase-level metrics to further differentiate constraints between the requirements. After determining the differences, we constructed the union- and disjoint-water marks by applying the water mark generation techniques to the measures to identify trade-offs. Finally, we invited three legal experts (two law scholars and one attorney) to review the final process and a subset of the generated water marks.

V. Summary Findings

Applying the method to the eight data breach regulations produced a total of 338 requirements with Maryland yielding the most (60 requirements) and Arkansas and Wisconsin the fewest (36, each) for an average 42 requirements per regulatory document. Requirements extraction from the eight regulations required approximately 2.2 hours per regulation. Additional time was expended to develop and refine the extraction method. The gap analysis to produce the measures required a total of 30.8 hours for the eight regulations. This effort required pairwise comparisons between the union of previously measured regulations and the entire next regulation (as shown in Figure 8, the size of the union grows slower, as a function of the total number of requirements covered), than disjoint. Figure 9 summarizes the number of requirements contained in the union water mark (a single standard) and the disjoint water mark (separate standards). Above each water mark, we display the average time in minutes required to analyze each requirement in the union water mark. Although this number rises moderately as each new jurisdiction is added, this increase suggests the process is linear. Note that our process employed no additional efficiencies over successive jurisdictions.

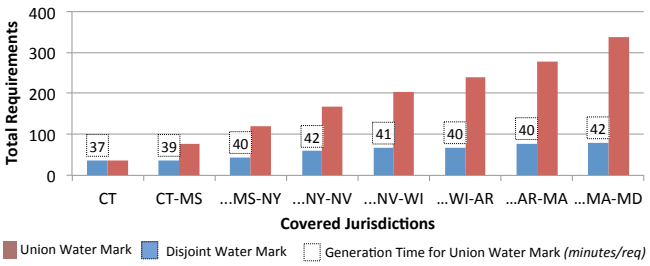


Figure 9: Requirement counts for union and disjoint high water marks

Figure 10 shows the relative breakdown of the comparison metrics (S-*, P-*) for each new jurisdiction when creating the union water mark. The first column, CT/MS, denotes the comparison between Connecticut (CT) and Mississippi (MS); the next column, CT-MS/NY, reflects a comparison between the generated water mark CT-MS and New York (NY), and so on. We now discuss interesting patterns observed during reconciliation.

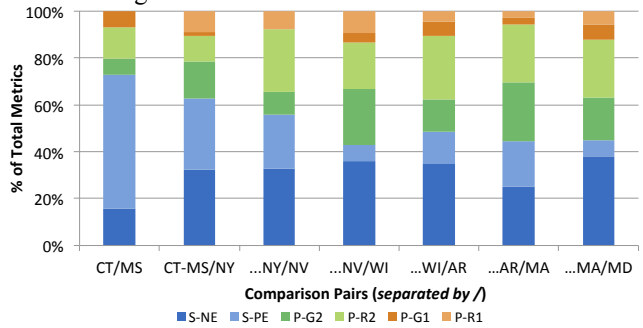


Figure 10: Breakdown of Comparison Metrics by Specification Pair

Increasing dominance of phrase metrics (P-) over statement metrics (S-*).* Initially, statement measures (S-*) contributed to over 70% of the total measures (see CT/MS); however, as additional regulations were added to the water mark, phrase measures began to dominate (P-*). As we seek to reduce comparison times between specifications for future work, we will begin with techniques that show promise in reducing phrase-level comparisons.

Increase in near equivalencies (S-NE) relative to Pure Equivalencies (S-PE). As the union water mark grew in size (see Figure 9), we see fewer pure equivalencies coupled with a rise in near equivalencies. The heuristics for reconciling phrase dissimilarities may produce this effect in the union water mark. The repeated merging of phrases produced requirements of increasing scope (e.g. "owns or licenses" changing to "owns, licenses, maintains, or uses") and thus reduced the likelihood of encountering two purely equivalent requirements. As all our comparisons occurred between generated specifications and a single jurisdiction (e.g. CT-MS-NY/NV) this may reflect the decreasing similarity of single jurisdictions with the water mark.

Increasing prevalence of constraint metrics (P-G2, P-R2) over concept metrics (P-G1, P-R1). Constraint metrics appear to be increasingly more prevalent than concept metrics (note the ratio of green to orange in each column) as we add specifications to the union water mark. This may indicate opportunities for future automation, as identifying concep-

tual generalizations is more difficult than identifying new or missing constraints.

VI. Patterns of Dissimilarity

During the water marking process, we observed multiple inter-regulatory conflicts that affect the system design or organizational processes, depending on the reconciliation technique employed: union or disjoint. In this section, we report these conflicts and observed impacts.

A. Variations Among Legal Definitions

Regulatory definitions serve as a "gateway" to deciding coverage. In the regulations we studied, the definitions for "personal information" produced several coverage conflicts (see Figure 11). The definitions have several overlaps, e.g., all include a first name, or first initial, and last name in combination with at least one "data element" as noted; however, individual states also note special inclusions and exclusions, such as covering medical information, or broadly covering any identifiable information. Furthermore, certain states differentiate *who* is or is not covered, such as making allowances for organizations subject to other laws, such as the Gramm-Leach-Bliley Act (GLBA). Explicit exclusions, such as Maryland's (MD) exclusion for information listed under HIPAA, are omitted in the union water mark, as these are in contention with other states' definitions, such as Arkansas, New York, and Wisconsin. Thus, we interpret this absence of coverage as discretionary, not mandatory.

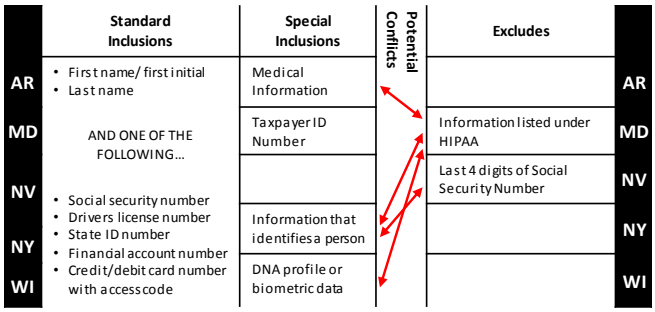


Figure 11: Inter-jurisdictional conflicts in personal information definitions

The CT-MS water mark was compared and reconciled with New York's §899-aa, which contains requirements NY-2 through NY-4 that prescribe how a data breach is determined (see Figure 12). These requirements clarify otherwise ambiguous requirements at the cost of flexibility within the organization. Because the relational dissimilar requirements are linked with the `REFINES` relation, they are retained and practiced in both jurisdictions. If kept disjoint, the two jurisdictions could implement different breach determination criteria, such as different security monitoring protocols to identify breaches.

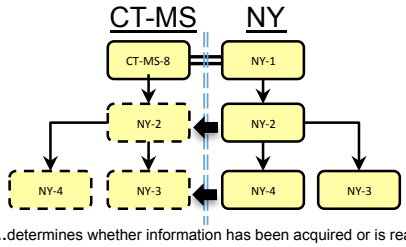
The disjoint water mark could involve different priority notification protocols for each data owner. For example, both CT-15 and MS-14 require notice to the data owner or licensee; however, Mississippi includes an additional constraint (indicated in red and measured by the P-R2 metric): provide notice "as soon as practicable following [the breach's] discovery". Using the union heuristics, the phrase was preserved in the reconciled requirement (CT-MS-15), placing a degree of urgency on the process.

CT-15: shall notify the owner or licensee of the information of any breach of the security of the data

MS-14: shall notify the owner or licensee of the information of any breach of the security of the data as soon as practicable following its discovery

CT-MS-15: shall notify the owner or licensee of the information of any breach of the security of the data as soon as practicable following its discovery

Furthermore, a notable case of relational-dissimilarity (see Figure 13) resulted in the omission of requirement CT-MS-11, which serves as an exception to the standard notification procedure shared across jurisdictions (CT-MS-7, NY-10). An exception only under CT-MS, this requirement cannot be applied to NY-10. Thus, CT and MS residents will be notified regardless of whether or not harm is likely as a result of a breach. However, the disjoint water mark would allow companies to only notify individuals from CT and MS under the more conservative distinction of *likely* harm. Our legal experts comment on this exception in Section VIII.



NY-1/CT-MS-8 :…determines whether information has been acquired or is reasonably believed to have been acquired by an unauthorized person…
NY-2 :may consider the factors outlined in this section, among others
NY-3 :may consider indications that the information is in the physical possession and control of an unauthorized person, such as a lost or stolen computer or other device containing information
NY-4 :may consider indications that the information has been downloaded or copied

Figure 12: Preservation of Refinement Series between CT-MS and NY §899-aa (GraphML)

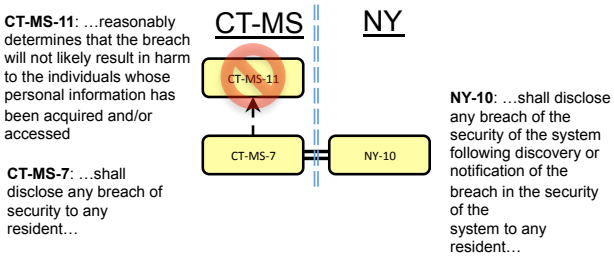


Figure 13: Removal of relational dissimilarity between union CT-MS and NY §899-aa (GraphML)

The requirements NV-47 and CT-MS-NY-48, below, present a complex phrase-dissimilar example pertaining to notification of consumer reporting agencies. We measure unique constraints in each requirement using the P-G2 and P-R2 metrics in blue and red, respectively. These constraints affect the action (how to notify), object (notice content) and the target (notice recipient). The more robust requirement CT-MS-NY-NV-47 preserves the highlighted phrases from each input requirement. Keeping such requirements disjoint may result in unnecessary duplication of effort in the determination of a consumer reporting agency or mix-ups about which notification content should be sent to whom.

CT-MS-NY-48: shall also notify consumer reporting agencies as to the timing, content and distribution of the notices and approximate number of affected persons

NV-47: shall notify, without unreasonable delay, any consumer reporting agency, as that term is defined in 15 U.S.C. ' 1681a(p), that compiles and maintains files on consumers on a nationwide basis, of the time the notification is distributed and the content of the notification

CT-MS-NY-NV-47: shall notify, without unreasonable delay, consumer reporting agencies, as that term is defined in 15 U.S.C. ' 1681a(p), that compiles and maintains files on consumers on a nationwide basis, as to the timing, content, an distribution of the notices and the approximate number of affected persons

B. *Variations in Practice*

During our reconciliation process we discovered unusual cases that merit additional care from the analyst. These cases include uncommon coverage mechanisms that indicate certain reconciliation techniques, the use of goal-based requirements that necessitate simultaneous reconciliation with multiple requirements, and the potential for reconciled definitions to have unintended implications as they are propagated throughout a requirements specification.

1) Variation in Coverage Mechanisms. The water mark generation process is used to reconcile requirements from different jurisdictions. Most of our regulations studied were limited to residents of the governed jurisdiction, however, Wisconsin (WI) §134.98 requires organizations that "have their principal place of business located in [WI]" (WI-1) to send notices to affected subjects, regardless of the subject's state of residence [6]. In this case, individuals are covered by both their state of residence and WI law, thus obstructing the disjoint water mark.

2) Goal-based Requirements. Goal-based requirements broadly describe what an organization must do, whereas means-based requirements describe how to achieve the goal. Reconciling similar goals and means yields numerous phrase dissimilarity measures when a single goal can be deemed equivalent to multiple means. In Figure 14, two states explicitly define criteria for the means of notification, including written (CT-22, MS-21), telephonic (CT-23, MS-22) and electronic notice (CT-24, MS-25). Alternatively, WI allows notification through a *reasonable* method (WI-27), which generalizes these means into a common goal.

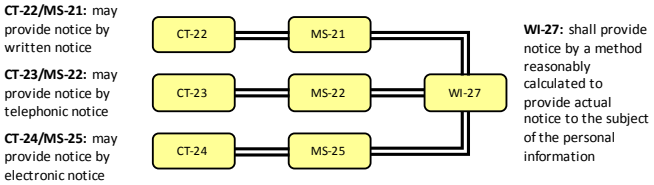


Figure 14: Multi-Statement Equivalency Between MS HB-583 and WI §134.98 (GraphML)

3) Deference to Standard. Regulations may defer to other regulations, such as the GLBA, as an alternative compliance standard. These external cross-references are problematic for requirements engineers, because they can yield errors and conflicts [17]. External cross-references can be inconsistently defined as well, as shown in the excerpts below. While we did not incorporate these external regulations into our analysis, a comprehensive analysis would examine cross-referenced regulations and may determine these external standards to be much higher.

MS-39: [an entity that] maintains such a security breach procedure pursuant to the rules, regulations, procedures or guidelines established by the primary or federal functional regulator, as defined in 15 USCS 6809 (2)

NV-45: [an entity that] is subject to and complies with the privacy and security provisions of the Gramm-Leach-Bliley Act, 15 U.S.C. " 6801 et seq.

VII. Threats to Validity

We now discuss threats to validity and our mitigations.

A. *Construct Validity*

Construct validity reflects whether the construct we propose to measure is indeed what we measured. In this paper, we rely on previously validated methods to acquire our data, including the frame-based method for extracting regulatory requirements from laws [5], and the nominal metrics for performing a gap analysis [3]. In this study, both authors reviewed the extracted requirements for consistency and both authors measured a stratified sample of requirements and found a 100% overlap for statement-level equivalences. We plan to conduct further evaluation on the phrase level metrics that employ the newly discovered heuristics for merging phrases, as reported in Table 2.

B. *Internal Validity*

Internal validity is the extent to which observed causal relationships exist within the data and, particularly, whether the investigator's inferences about the data are valid [23]. Each nominal measure is an inference that some statement or phrase can be assigned to a corresponding unary or binary relationship based on the metric's definition. Because the binary metrics are asymmetric, an alternative explanation for the findings is that the water marks are due to the order in which the comparisons occur, which is a threat to internal validity. Thus, we tested the water mark chaining for the commutative property and found the same water marks despite the order of comparisons. We intend to further test this assumption by examining other domains with less similarity in the domain phenomenology.

C. *External Validity*

External validity is the extent to which the framework generalizes. U.S. data breach notification laws are largely homogenous, as opposed to comparing laws from finance to healthcare, which describe different domains and risks to privacy and security. We selected data breach laws for this very reason, to assess very near-similarities in prototyping our water mark process. Thus, our evaluation reflects differences in laws that are extremely subtle. Future work should examine laws from multiple, different domains to assess external validity of our guidance and heuristics.

VIII. What the Legal Experts Say

In addition to repeatability, we are interested in how our results are reflected in applied settings. As noted by Siena et al. [21] and Bobkowska and Kowalska [1], legal and engineering viewpoints differ and these differences must be accounted for when prioritizing compliance decisions. To address this issue, we engaged with subject-matter experts to review our results through semi-structured interviews [2, 10]. Before the interviews, we provided each expert general descriptions of the reconciliation techniques and select con-

flicts from our dataset, and asked which techniques they would propose or they believe are currently practiced. We also surveyed the perceived legal validity of reconciliations produced using the union technique, given its complexity. The presented conflicts were chosen to demonstrate the different heuristics or strategies prescribed by the union technique, such as duplicating an action from one jurisdiction in a second jurisdiction (i.e., preserving a relational dissimilar requirement linked with `REFINES`). We organized their feedback around the following questions:

How do legal experts identify and resolve conflicts across jurisdictions? Our experts employ their past experience and training to resolve conflicts, often working directly with clients and their restricted abilities, budgets, organizational structure, etc. Companies may choose experts who are familiar with local jurisdictional sensitivities, including which requirements are routinely enforced or ignored, and experts may prioritize requirements differently based on their individual judgment. The prioritization process can include political, economic and technological issues, such as, is the State's Attorney General up for re-election, are the implementation costs for a requirement unreasonable, and has a technology changed to invalidate a regulatory requirement.

How do legal experts perceive the different reconciliation techniques of union, disjoint, and minimum? Our experts generally responded positively to the reconciliation techniques (high standard, separate standards, low standard) and grasped their intent, immediately. Respondents generally agree that the disjoint water mark may be cumbersome, but posed no additional legal concern, as the legal text in the requirements can remain unmodified. Although they agreed that the proposed union water marks for the requirements were "reasonable" and "legally fine", they offered a number of valuable caveats:

- Sending notice to individuals for every breach may appear as a higher standard (MA, NV, NY), than sending it only when there is a risk of harm (AR, CT, MS, WI); however, the latter approach avoids over-inundating residents with notices and losing their effectiveness. The aim of the notice is to encourage residents to act when there is a risk of identity theft. Thus, incorporating the rationale for a particular requirement can aide in resolving these conflicts, however, the elicitation and documentation costs can limit the preservation of rationale [14]. In general, the analyst should examine the underlying intent when considering trade-offs that involve different frequencies and not presume that better satisficing is always best, e.g., more notice means more consumer awareness, higher encryption key bits means more confidentiality, etc.
- The union can introduce other parties who have their own requirements into the business process, such as obligation CT-10 to consult with law enforcement in the event of a data breach. Implementing this practice may further limit company autonomy, because law enforcement can deliver advice that leads to new requirements that conflict with existing regulations.
- Preserving an action from one regulation not present in another may indirectly violate an unrelated requirement in the other. To reuse the above example, consulting with law enforcement may introduce an unacceptable delay in a certain jurisdictions where this practice is not prescribed. In this particular case, a preserved sub- or post-process (`REFINES` or `FOLLOWS`) may produce an undetected conflict with a quality attribute, e.g., delay the notice conflicts with notifying consumers, expeditiously.
- Requirements that are particularly difficult to reconcile may best be resolved by choosing a self-imposed standard that is higher than both, rather than risk choosing one or the other and yield a gap in compliance. For example, choosing to provide notice within a specific time frame (e.g., "48 hours") rather than allowing the system to default to the legally required time frame "as soon as practicable" or "immediately following discovery".

What do legal experts recommend to businesses? Our experts indicated that the union technique "is a familiar approach in law" as an organization will often pick the most onerous standard, particularly if the regulations are large. This remark is tempered with the belief that businesses only take on the more onerous standard provided that there is not a "significant cost difference." Regardless, the organization "will always back up its decision by having [a] business justification for [the decision]." Although respondents recognize the multiple standards created using the disjoint technique, they often prefer this approach over union, because it introduces less risk than reinterpreting the law. No respondent advocated for the minimum technique, citing its lack of compliance; however, one recognized that, due to resource constraints, organizations may prioritize meeting certain jurisdictional standards before others; e.g., "we have affected [individuals] in every state but the majority of them are from [this state and that state]; we want to avoid legal trouble in these locations in particular." When asked further about this, the respondent admirably indicated "[I] would much rather a client tries to do the best they can as opposed to saying 'I can't afford this' or 'I can't do anything.'" Respondents acknowledged that differences in experience, past clients, and area of focus could contribute to different opinions between legal experts.

IX. Discussion and Summary

In this paper, we present a new method that combines previously validated techniques for extracting legal requirements from regulations, measuring differences between two requirements sets, and inferring legal water marks for high and low standards of care across multiple corresponding jurisdictions. We applied the technique to eight U.S. data breach notification laws. We found that performing the union across this domain yielded a reduction from 338 total requirements down to 80 (a 76% reduction). Based on our interviews with legal experts, we believe most companies appear somewhere between the union and disjoint water marks, and may appear below the minimum standard when faced with resource constraints or when initially setting up their internal compliance regime for a new domain.

We discovered the process is commutative; this was done by generating specifications for a subset of our jurisdictions (CT, MS, and NY) in which the jurisdictions were recon-

ciled in different orders: CT-MS-NY and NY-MS-CT. The requirement counts were identical (48) and prescribed the same standard of care measured using the metrics in this paper. Differences between water marks were purely aesthetic, such as the order of reconciled phrases (e.g., "owns, licenses, or maintains" vs. "owns, maintains, or licenses") or the identifier assigned to the requirement (CT-MS-NY-14 vs. NY-MS-CT-13). Intermediate specifications (CT-MS, NY-MS) vary, but this is expected since they cover different jurisdictions.

During this analysis, we identified several opportunities for improving the method. For example, by grouping requirements into named categories (e.g., notification, access, encryption, disposal) based on their action verbs, we may be able to reduce the number of pairwise comparisons required with a small loss in precision and recall. In addition, our expert reviewers noted how rationale can be used to resolve trade-offs by appealing to tacit or undocumented regulatory and industry goals. For trade-offs where any decision would yield a non-compliant outcome with one or more decisions, this expert feedback may be used to justify that the decision is a best effort to an otherwise impossible legal landscape.

Our method is primarily manual with limited tool support to encode the extracted requirements, produce visualizations and record the comparison measures reported by the analyst. During our study, we applied the "ideal" best IR-based technique reported by Falessi et al. [9] to trace equivalent requirements pairs with the aim to improve performance in Step 2 of our method in Figure 1. This technique is based on vector-space models with a Cosine similarity measure, linear-incidence term weighting and Stanford part-of-speech noun and verb extractor. With respect to their dataset, this technique exhibited 0.935 precision and 0.936 recall with a 0.75 Lag, which measures the number of true positives within a proportion of the highest ranked results. Using our manually acquired results as the gold standard, the NLP technique performed very poorly, with a 0.077 precision and 0.300 recall. The reason for this discrepancy may be the small size of a legal requirement (typically 10-20 words), whereas, NLP-based techniques were originally developed from analyzing large corpus of thousands of words. To our knowledge, automated traceability methods have not yet advanced to implement our phrase-level measures. We see improvements in NLP-based analysis as a welcome improvement in our research.

ACKNOWLEDGMENT

This research was supported by the U.S. Department of Homeland Security (Grant Award #2006-CS-001-000001) and Hewlett-Packard Labs Innovation Research Program (Award #CW267287).

REFERENCES

[1] Bobkowska, A., Kowalska, M. "On efficient collaboration between lawyers and software engineers when transforming legal regulations to law-related requirements," *2nd Int'l Conf. Info. Tech*., pp.105-109, 2010.

[2] Bogner, A., Littig, B., Menz, W. *Interviewing Experts*. Palgrave Macmillan, 2009.

[3] Breaux, T.B., Anton, A.I., Boucher, K., Dorfman, M., "Legal requirements, compliance and practice: an industry case study in accessibility." *IEEE 16th Int'l Req'ts Engr. Conf.*, pp. 43-52, 2008

[4] Breaux, T.D., Gordon, D.G., "Regulatory requirements as open systems: structures, patterns and metrics for the design of formal requirements specifications." *Carnegie Mellon University Technical Report CMU-ISR-11-100*, 2011.

[5] Breaux, T.D., *Legal Requirements Acquisition for the Specification of Legally Compliant Information Systems.* Ph.D. Thesis, North Carolina State University, Apr. 2009

[6] Bryan Cave LLP. "Wisconsin Data-Security Law Imparts Obligation to Issue Consumer notification in Case of Security Breach." *Data Security Bulletin*, 2006.

[7] Corbin, J., Strauss, A. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, Sage Publications, 2007.

[8] Dekhtyar, A., Dekhtyar, O., Holden, J., Hayes, J.H., Cuddeback, D., Kong, W-K. "On human performance in assisted requirements tracing: statistical analysis." *19th IEEE Int'l Req'ts Engr. Conf*., pp. 111-120, 2011.

[9] Falessi, D., Cantone, G., Canfora, G. "Comprehensive characterization of NLP techniques for identifying equivalent requirements." *ACM-IEEE Int'l Symp. Empir. Soft. Engr. & Measm't*, pp. 18:1-10, 2010.

[10] Flick, U., *An Introduction to Qualitative Research*, 4th ed. Sage Publications Ltd., 2009.

[11] Gacitua, R., Sawyer, P., Gervasi, V. "On the effectiveness of abstraction identification in requirements engineering," 18th IEEE Int'l Conf. Req'ts. Engr., pp. 5-14, 2010.

[12] Gervasi, V., Zhowghi, D. "Mining requirements links." *Req'ts Engr'ing: Fnd. Soft. Qual., LNCS* 6606: 96-201, 2011.

[13] Gordon, D.G., Breaux, T.D. "Managing Multi-Jurisdictional Requirements in the Cloud: Towards a Computational Legal Landscape." *3rd ACM Cloud Computing Security Workshop (CCSW'11)* pp. 83-94, 2011.

[14] Greenspan, S., "Panel on recording requirements assumptions and rationale," *IEEE Int'l Symp. Req'ts Engr.*, pp. 282-285, 1993.

[15] Cleland-Huang, J., Czauderna, A., Gibiec, M., Emenecker, J. "A machine-learning approach for tracing regulatory codes to product specific requirements." *IEEE Int'l Soft. Engr. Conf*., pp. 155-164, 2010.

[16] Urquhart, J. "Regulation, Automation, and Cloud Computing," CNET, Aug-2011. [Online]. Available: http://news.cnet.com/8301-19413_3-20086081-240/regulation-automation-and-cloud-computing/. [Accessed: 05-Mar-2012].

[17] Maxwell, J.C., Anton, A.I., Swire, P. "A legal cross-references taxonomy for identifying conflicting software requirements." *19th IEEE Int'l Req'ts Engr. Conf.*, pp. 197-206, 2011.

[18] Kroes, N. "The clear role of public authorities in cloud computing," *Digital Agenda Comissioner* - Neelie Kroes. Apr-2011.

[19] Otto, P.N., Anton, A.I., "Addressing legal requirements in requirements engineering." *15th IEEE Int'l Req'ts Engr. Conf.*, pp. 5-14, 2007.

[20] Sabetzadeh, M., Nejati, S., Liaskos, S., Easterbrook, S., Chechik, M. "Consistency checking of conceptual models via model merging." *15th IEEE Int'l Req'ts. Engr. Conf.*, pp. 221-230, 2007.

[21] Siena, A., Mylopoulos, J., Perinir, A., Susi, A. "From laws to requirements," *1st Int'l Work. Req'ts Engr*. & Law, pp. 6-10, 2008.

[22] Weitzner, D., Privacy Law Scholars Conference Keynote Address, Deputy Chief Technology Officer in the White House Office of Science and Technology Policy, 2011.

[23] Yin, R.K. Case Study Research: Design and Methods, 4th edition, Sage Publications, 2009.

[24] Zou, X., Settimi, R., Cleland-Huang, J. "Improving automated requirements trace retrieval: a study of term-based enhancement methods." *Empir. Soft. Engr.,* 15:119-146, 2010.