Mapping Legal Requirements to IT Controls

Travis D. Breaux[1], David G. Gordon[2]
Institute for Software Research[1], Engineering & Public Policy[2]
Carnegie Mellon University
Pittsburgh, USA
{breaux@cs,dggordon@andrew}.cmu.edu

Nick Papanikolaou, Siani Pearson
Cloud and Security Lab
HP Labs
Bristol, UK
{nick.papanikolaou,siani.pearson}@hp.com

***Abstract*—Information technology (IT) controls are reusable system requirements that IT managers, administrators and developers use to demonstrate compliance with international standards, such as ISO 27000 standard. As controls are reusable, they tend to cover best practice independently from what specific government laws may require. However, because considerable effort has already been invested by IT companies in linking controls to their existing systems, aligning controls with regulations can yield important savings by avoiding non-compliance or unnecessary redesign. We report the results of a case study to align legal requirements from the U.S. and India that govern healthcare systems with three popular control catalogues: the NIST 800-53, ISO/IEC 27002:2009 and the Cloud Security Alliance CCM v1.3, as well as the CCHIT EHR Certification Criteria. The contributions include a repeatable protocol for mapping controls, heuristics to explain the types of mappings that may arise, and guidance for addressing incomplete mappings.**

Keywords-requirements engineering, privacy requirements, healthcare requirements, HIPAA, NIST 800-53, ISO 27002, CCM, CCHIT.

I. INTRODUCTION

Growth in information technology has led to measurable productivity growth and other economic benefits [5]. These benefits have since been attributed to the impact of IT on workforce decentralization and on creating an information rich workplace [15]. While this growth is partially attributable to increased automation, the introduction of IT into daily life also yields new societal concerns and in turn new regulation to ensure that IT practices are consistent with society norms. For example, Canada, the United States and Europe, among several other countries, have introduced several new laws governing IT privacy over the last decade. In addition to new laws, older laws that govern legacy business practices have new relevance to IT systems that supplement these practices through automation. The challenge for IT managers, administrators and developers – whether they are planning, configuring or evolving IT systems and their designs – is to determine what steps are needed to comply with these laws.

Industry best practice includes demonstrating that IT systems are aligned with industry standards, such as the ISO 27000 Series security standards. In order to demonstrate such alignment, companies will often carry out compliance checks in-house, as well as consult third parties to obtain accreditations and various certifications that are recognized within industry. Control catalogues are standards that contain itemized controls, which appear as prescriptive or descriptive statements of a standards-compliant IT-enabled organization. An analyst who aims to demonstrate compliance with a standard can attempt to map these controls onto their information practices or product requirements, while collecting evidence to justify, support or otherwise rationalize these mappings. In practice, however, we find that such control mappings are *ad hoc* and heavily dependent on tacit domain knowledge, which limits repeatability and increases the level of effort, respectively.

In addition to standards-compliance, companies may use previously mapped controls as a means to demonstrate compliance with laws. When companies reuse previously mapped controls, the savings can justify the investment to construct the original mapping, as companies only need to modify their information practices to comply with the difference between the law and the mapped controls. To facilitate this reuse, we conducted a small case study exploring the process of mapping regulatory requirements in the healthcare domain to three industry standards: NIST 80-53, draft revision 3; ISO/IEC 27002:2009, and the Cloud Security Alliance Common Control Matrix (CSA/CCM), version 1.3, as well as the EHR system Certification Criteria proposed by CCHIT in 2011. The contributions of this experience include a protocol for conducting control mappings to regulations, heuristics to explain the types of mappings that may arise, and guidance for evaluating incomplete mappings. The remainder of the paper is summarized as follows: in Section 2, we review related work; in Section 3, we present our research method and alignment process; in Section 4, we present our research findings; with a discussion and summary in Section 5.

II. RELATED WORK

Related work includes frameworks and methods to guide business analysts in the process of aligning regulatory requirements with IT systems. Islam et al. describe a framework to align regulations with security requirements that are elicited using Secure Tropos [12]. During the elicitation process, the analysts can use trust assumptions to restrict the analysis scope [11]. Sackmann et al. describe a 5-layer model to align laws with IT systems that includes mapping from regulations to control objectives, such as ITIL, CoBIT, and COSO, policies, monitors and IT system artifacts [14]. The model highlights reusing a company's

existing control objectives, and we re-interpret this model in the context of the types of mappings that we identified in our case study. Finally, methods exist to map CoBIT objectives to RBAC policies [7], to map legal requirements to state-based monitors [1], and to map legal requirements to product-requirements [3].

Huang et al. applied natural language processing and machine learning to trace legal requirements from the Health Insurance Portability and Accountability Act (HIPAA) Security Rule to product requirements of various Electronic Health Record (EHR) systems [6]. While the average precision scores vary widely across category of legal requirement (0.15-0.55), the results show a significant improvement by comparing different methods.

Gandhi and Lee describe a method to calculate the risk of non-compliance with security regulations using a lattice algebraic computational model [8].

III. Research Method & Alignment Process

In preparation for studying how industry practitioners align regulations with IT controls, we first sought to apply and generalize from our own ad hoc methods – hence, why this paper describes an experience report that is limited to the results of a few investigators. However, to compare our results and insights, we employ an exploratory case study design [16] to map regulations to IT controls. In this study, we seek to experience how an analyst can conduct an IT control mapping, to reflect on any reusable heuristics to guide an analyst, and what an analyst might do when an obstacle or slight difference arises between a regulation and a control description. Thus, we sought to answer four research questions (RQs):

RQ_1. What assumptions must an analyst make to realize a mapping from a legal requirement to an IT control?

RQ_2. What heuristics can analysts use to motivate or justify that a mapping is correct?

RQ_3. What types of gaps exist between legal requirements and IT controls in a given mapping?

RQ_4. What is the overall complexity of performing and maintaining an IT control mapping?

We developed an alignment process by attempting to perform an IT control mapping. Four subject-matter experts, who have varying levels of regulatory and security expertise, independently developed their own processes as an exercise. To coordinate the experts, we employed a scenario to narrow the scope and realization of IT control mappings: a patient seeking ambulatory care in a hospital; the scenario was adapted from the sample profiles for the Health Level[1] (HL) 7 Electronic Health Record System Functional Model, which is a standard model from the healthcare industry.

[1] HL7 is an international industry trade group working to improve healthcare automation by establishing IT standards.

We began the study using a stratified sample of regulatory requirements. Because regulations describe a range of activities, from high-level business practices to low-level functional requirements, we chose a stratification that reflects this range and to cover a variety of topics. The categories that we used are:

- *Retention*, which restricts how long data is retained;
- *Personnel Training*, which describes training for staff and IT system users;
- *Authentication*, which restricts how data may be accessed and by whom;
- *Validation*, which requires that data be accurate and up-to-date;
- *Consent*, which restricts when individual consent is needed before data collection and use; and
- *Quality assurance*, which concerns documenting security practices.

In addition, we selected legal requirements from five jurisdictions to complement our scenario based on a hypothetical manufacturer who targets EHR system deployment in California, Florida and New York and who plans to outsource a medical transcription service to India. This enhanced, cross-border scenario requires the manufacturer to also consider the HIPAA Security Rule, which is a U.S. national regulation that creates a "security floor" or minimum set of security requirements. We sampled 15 legal requirements from the following five laws that were encoded using the legal requirements specification language [4]:

FL: Florida Administrative Code, 59A-3.270
CA: California Code Regs. tit. 22 § 70751
NY: New York Comp. Codes R. & Regs. title 10, §405.10
HIPAA: U.S. Health Insurance Portability and Accountability Act (HIPAA), Security Rule, Par 164, Subpart C, §§164.302-318, 2003
ITR: India Technology Rules: Reasonable security practices and procedures and sensitive personal data or information, 2011

Finally, we chose four IT control catalogues from which to identify relevant mappings:

- ISO/IEC 27002:2009, titled "Information technology — Security techniques — Code of practice for information security management"
- National Institute of Standards and Technology (NIST) Special Publication (SP) 800-53, draft revision 4, titled "Security and Privacy Controls for Federal Information Systems and Organizations"
- Cloud Security Alliance (CSA) Cloud Control Matrix (CCM) version 1.3. (Note that this version is intended as a baseline for comparison, and is not intended to be comprehensive in its current state.)
- CCHIT Certified 2011 Ambulatory EHR Certification Criteria.

The ISO/IEC 27002:2009 is an established, international security standard containing 133 controls. The NIST SP

800-53 governs U.S. government agencies and many contractors and was first released in February 2005 with the fourth revision containing 228 controls that cover a wide range of security and privacy topics. In this draft of the fourth revision is a new Appendix J on privacy. Finally, the CSA CCM v1.3 contains 98 security controls that have been mapped to multiple other standards, including CoBIT v4.1, HIPAA, ISO/IEC 27001-2005, NIST SP800-53, PCI DSS v2.0 and many more. The CSA CCM is specifically designed for cloud computing and represents the most recent control catalogue in our study.

A. The Alignment Process

The alignment process consists of identifying and documenting control mappings. To identify a mapping, we performed *analysis pooling*, a technique developed by Gandhi and Lee [8], to locate relevant controls for each legal requirement. Analysis pooling consists of three steps: (1) search the IT control lists for related keywords from the legal requirement statement; (2) conduct *focused-hierarchical browsing,* which is a search within the category or neighborhood of any IT controls that were identified in step 1; and (3) conduct *multi-dimensional browsing*, which is a search outside the category for non-taxonomical relationships to controls. Step two is often about finding the right level of abstraction to map a requirement to control, whereas the third step is driven by deeper process and architectural implications that we discuss in Section IV. To document a mapping, the analyst records the following information: the requirement ID, the control ID, the rationale for the mapping, whether the control under- or over-specifies what is stated in the requirement or whether the two statements were equivalent, and what additional steps must be taken to address the gap between the requirement and control, if any.

IV. Research Findings

The analysts produced different mappings between the 15 regulatory requirements and the control catalogues; the total number of mappings is shown in Table I.

TABLE I. TOTAL NUMBER OF MAPPINGS PER ANALYST

Analyst	NIST 800-53	CSA CCM	ISO 27002	CCHIT EHR Cr.
1	29	34	19	51
2	31			15
3	35	44	18	

In Table I, all three analysts produced mappings from the regulatory requirements and the NIST 800-53 standard, however, due to limited time, analyst #2 was only able to map to NIST 800-53 controls and CCHIT requirements. While it may appear that all three analysts identified similar mappings under the NIST 800-53 control set, there was remarkably little overlap. When we compare analysts #1 and #2's mappings in the NIST 800-53 dataset, we found only 6 equivalent mappings; analysts #1 and #3 had 2 equivalent mappings, and analysts #2 and #3 had 6 equivalent mappings, with one mapping shared among all analysts. Similarly, comparing analyst #1 and #3's mappings in the CSA CCM and ISO 27002 data sets only yield 7 and 3 equivalent mappings, respectively. On closer inspection, it appears that some portion of the incompatibility is due to the wide range in "levels of abstract," such whether a required to ensure confidentiality maps to a control for developing confidentiality policies or a control to ensure cryptographic functions to protect data. We discuss a possible solution to this mismatch using capability-based mapping in Section V.B. We now discuss our detailed analysis of the mappings and rationales provided by each analysts to uncover assumptions, heuristics and types of gaps experienced during the mapping. Afterwards, we discuss lessons learned in Section V and our proposal for a solution to this problem.

A. Assumptions and Domain knowledge

We began by asking, *what assumptions must an analyst make to realize a mapping from a legal requirement to an IT control?* When mapping a regulatory requirement to a control statement, we discovered that the analysts made different assumptions about the requirement and target system when determining which controls were adequate. For example, one assumption was that certain classes of policy needed to exist before a policy-governed action could occur. Upon reflection, we discovered that an analyst could decompose a regulatory requirement into preliminary actions or procedures that are necessary to fulfill that requirement. Consider California requirement CA-5; legal requirements appear in the **Sans Serif**, and IT Controls in `monospace with dashed border`:

> **CA-5**: shall keep the records of unemancipated minors at least one year after such minor has reached the age of 18 years and, in any case, not less than seven years

Requirement **CA-5** assumes that patient records exist, which entails the existence of certain functions to enable data collection, access, retention and so on. To implement this **CA-5**, the system designer must be able to identify patient records of unemancipated minors, which are a class of patient who are 18 years of age and still under the supervision of a legal parent of guardian. In addition, the designer must have access to the patient's birth date (or age) and the first date for which the patient was seen in order to compute the retention period. Prior to performing the mapping with an IT control catalogue, the analyst may decompose this legal requirement into a set of implied requirements as follows that govern medical records:

> **CA-5.1**: shall maintain the patient's birth date
>
> **CA-5.2**: shall maintain the patient's emancipated status
>
> **CA-5.3**: shall maintain the date of record origination

This expanded set was used to perform the mapping and yielded additional control mappings. For example, one analyst mapped **CA-5** to the following two controls:

> **CCM DG-02:** Data, and objects containing data, shall be assigned a classification based on data type…
>
> **NIST DM-2:** The organization: a. Retains personally identifiable information (PII) for [Assignment: organization-defined time period] to fulfill the purpose(s) identified in the notice or as required by law;

The control CCM DG-02 requires classifying data types, which could then be referenced when applying various policies and, in this case, legal requirements. Control NIST DM-2 refers to personally identifiable information, which includes a person's birth date. This last control triggers alignment with any relevant privacy notices and laws that may be outside the scope of the original legal requirement **CA-5**. For this scope of this study, we restricted ourselves to surfacing assumptions based on keywords contained in the requirements; however, a subject matter expert may also infer relationships more broadly. For example, by recognizing that any retained data may need to be access restricted, and thus see links to access controls. We did not intentionally attempt to explore these broader implications.

B. Mapping Heuristics and Types of Gaps

We reflected on each of our mappings and sought to answer two critical questions: what heuristics do analysts use to motivate or justify that a mapping is correct, and what types of gaps exist between legal requirements and IT controls in a given mapping? When an analyst identified a control, they recorded their rationale for the control and described what they believed would need to be done to bridge the gap between the control and the requirement, if any. We coded this information and discovered five heuristics and corresponding gap types, which we now discuss.

1) Near Equivalence. Controls can appear similar to a requirement in many ways. In general, the control description may be *too narrow* or *too broad*, however, the actions taken can be near synonymous to the class of action required by the regulation. Consider the first situation, where a control is too narrow, by comparing requirement **FL-5** to control AT-3 to provide training.

> **FL-5:** [shall establish a process that] provides education and training in information management principles to decision-makers and other hospital personnel who generate, collect, and analyze information

> **NIST AT-3:** The organization provides role-based security training to information system users

Control AT-3 requires security training to information system users, as opposed to more general training on information management principles to decision-makers. To reuse this control, the analyst's organization would benefit from having developed (or requisitioned the development of) training material previously on a narrower IT topic. However, this training material falls short of what the legal requirement entails, as information management can include a much broader set of issues, especially for healthcare. Moreover, training may be based on generic, principles-oriented material or tailored to the functions and operations of a particular system.

The second type of near equivalence is a mapping to a control that is too broad in scope. Consider requirement **HIPAA-76** and control SI-2, below. The requirement for periodic security updates is one of many activities required by this control, specifically activity (b) in SI-2 that concerns security-related updates. The other activities within this control include additional benefits, such as monitoring flaws, generally, and testing updates before rolling them out across one's organization. In this case, the gap yields additional benefits that are not required by the regulation.

> **HIPAA-76:** [an entity must] implement periodic security updates.

> **NIST SI-2:** The organization: a. Identifies, reports, and corrects information system flaws; b. Installs security-relevant software and firmware updates; c. Tests software and firmware updates related to flaw remediation for effectiveness and potential side effects on organizational information systems before installation...

A third type of near equivalence is a solution to a related problem that demonstrates the ability to meet constraints of the same class of problem. Recall **CA-5**, above, which requires retaining medical records. An analyst mapped this regulation to NIST AU-11, below, which also describes retention and refers to a records retention policy. This control is focused on security incidents and would entail very different kinds of design choices: e.g., monitoring incidents on networks and computer systems, such as viruses, rootkits and phishing e-mails, as opposed to collecting medical information from healthcare professionals. However, the considerations needed to retain records in general are very similar.

> **NIST AU-11:** The organization retains audit records for [Assignment: organization defined time period consistent with records retention policy] to provide support for after-the-fact investigations of security incidents and to meet regulatory and organizational information retention requirements.

In this case, the gap is a matter of design abstraction: can the designer abstract the fundamentals of record retention into a general mechanism that could be reused across product lines or within the same product?

2) Policy Refinement. Analysts may see similarities between the regulatory requirement and a policy control. These controls require organizations to develop policies of a particular kind, such as audit, authorization or cryptographic key management policies. If the legal requirement falls within one of these existing policy areas, but has no obvious equivalent statement, then the analyst may create a mapping to the policy control. Consider legal requirement **NY-36**, which requires a quality assurance process based on record sampling to detect inaccuracies, and control ISO 10.10.2, which requires general procedures for monitoring information use and regular activity reviews.

> **NY-36:** [shall implement] a process implemented as part of the hospital's quality assurance activities that provides for the sampling of records for review to verify the accuracy and integrity of the system.

> **ISO 10.10.2:** Procedures for monitoring use of information processing facilities should be established and the results of the monitoring activities reviewed regularly.

To address this kind of gap, the analyst can extend the existing policy and procedures to include the specific steps required by the legal requirement. Because the policy establishes the basic responsibilities, resources and timelines necessary to conduct reviews, this gap may require less rework than starting a quality assurance program anew.

3) Technical Reuse. An analyst may find a technical control that can be used to implement the regulatory requirement. These controls may fall short of what the legal requirement entails, however, the controls also provide important insight into one mechanism for satisfying the requirement. Consider requirement *NY-34*, which requires a data verification process, and control ISO 12.2.1, which require data input validation in applications.

> **NY-34:** shall implement an ongoing verification process to ensure that electronic communications and entries are accurate

> **ISO 12.2.1:** Data input to applications should be validated to ensure that this data is correct and appropriate.

This mapping covers the requirement to validate computer "entries," but does not include the full range of communications. Other technologies, such as Secure-Socket Layer (SSL), may be needed to ensure that communications are not modified during transport – a topic not covered by this control.

C. Threats to Validity

We now discuss threats to validity.

Construct validity reflects whether the construct we propose to measure is indeed what we measured [16]. In this work, we relied on previously validated metrics to acquire the requirements used in our data set, including the frame-based method for extracting regulatory requirements from laws [2]. Mappings made by the analysts between requirements and controls were accompanied by a corresponding justification, which was reviewed by the other analysts in a group discussion.

Internal validity is the extent to which the inferences drawn from the data are valid [16]. As each investigator used their own technique for performing the mapping process, there is considerable variance among the mappings and accompanying annotations made, given differences in background knowledge, the ability to see connections, and the vigilance required to achieve a "complete" mapping. Additionally, the notion of a gap between a legal requirement and control was not explicitly defined prior to the mapping process. We seek to address these issues in future work by developing a more controllable and documentable process for each analyst to determine their mappings.

External validity is the extent to which the framework generalizes to other data sets [16]. Although the requirements draw on multiple bodies of law in different domains, including federal and state-level healthcare in the U.S. and Indian data privacy regulation, the number of potential regulations - as well as domains they govern - number in the hundreds if not thousands. Further, the sample of requirements (15) is small. To address these issues, in future work we will test our heuristics and recommended guidance using laws from other domains with larger requirements samples. We anticipate supporting these mappings with automated processes that could be used to reduce the number of comparisons made.

V. Discussion and Summary

During this case study, we discovered several lessons that guide how we approach control mappings. We now share these lessons with a proposal for how we could automate this process.

A. Lessons Learned

Lesson 1: Reusable controls can be less sensitive to specific requirements. Control sets are intended to be reusable across multiple projects. As discussed, there are many potential reasons for this: analysts may become increasingly familiar with the control set as it is reused across projects, reducing the effort required to conduct the mapping; projects can be compared to one another based on the controls used to satisfy their requirements; and costs and risks can be computed for each control, using feedback from completed projects, to assist in projected-related decision-making. Legal requirements, contrarily, are more specific in nature and pertain to a particular context or setting; this is achieved by the presence of constraints in the requirement restricting its applicability or meaning. Controls lack sensitivity to these constraints, which increases the likelihood of "imperfect" alignment between a requirement and a control - that is, the requirement will contain a constraint that the

control exceeds or does not meet, which may result in under- or over-utilized resources, respectively.

If an organization seeks to develop its own reusable control set, or perhaps to expand an existing set, it should consider the trade-off made between reusability of that control set and the degree of alignment between that set and the requirements contained in projects to which it will be applied. As the control set expands in size and controls become increasingly specific or nuanced (in order to achieve a greater degree of alignment with requirements) the reusability of that control set is reduced. Correspondingly, as a set is reduced in size and made increasingly generalizable, the greater the odds of over-alignment with requirements.

One approach to address these opposing forces is taken by NIST; the NIST controls feature variable fields enclosed in brackets that support reuse and the ability to tailor a control to the requirement's unique context. Although this makes the control more sensitive to the requirement, it only allows for a better (but not necessarily perfect) alignment, as with requirement **NY-36** and NIST control CA-7:

> **NY-36:** [shall implement] a process implemented as part of the hospital's quality assurance activities that provides for the sampling of records for review to verify the accuracy and integrity of the system.

> **CA-7:** The organization develops a continuous monitoring strategy and implements a continuous monitoring program that includes:
>
> a. Establishment of [Assignment: organization-defined metrics] to be monitored...
>
> d. Ongoing security status monitoring of organization-defined metrics in accordance with the organizational continuous monitoring strategy...

Lesson 2. Orthogonal control sets can be easier to map. Even though controls may be conceptually related, mapping can be difficult if they approach an issue from different viewpoints. Within our dataset, we found this to be most evident with regards to the subject of the control and the mechanism used to meet that control.

Although two laws may have the same high-level goal (e.g. maintaining the privacy of medical records), they may vary on the viewpoint taken to meet this goal. One law may impose requirements on medical records, and another may impose requirements on the personnel or departments that handle these records. This notion of varying viewpoints may also apply to control sets. Prior to performing the alignment process, the analyst should ensure that - if possible - the control set and legal requirements share similar viewpoints.

Relatedly, the control sets and legal requirements may use different mechanisms to achieve these goals; the more disparate the mechanisms, the more difficult the mapping will be. Mapping a technically oriented control set to a policy oriented set of legal requirements (or vice versa) requires additional effort on the part of the analyst due to the dissimilarity between the two sets.

Lesson 3. Controls are limited views of total functionality. While controls offer a standard means to assess a system, it is likely that the system itself affords other functions outside of a control catalogue that could be used to achieve legal compliance with requirements. If the analyst restricts herself solely to the control set, it is possible that she will exclude potential solutions that could achieve compliance with greater efficiency than those offered by the control set itself. These solutions may be unique to the system-to-be or the context in which it will exist, and can only be discovered by individuals with additional knowledge of these areas. When an existing mapping to a given control set is poor, the analyst should consider seeking these alternative means of compliance, as it justifies the effort necessary to discover their existence. Discovery of such means could further result in expansion or modification of the control set, though this raises the issue of reusability vs. applicability mentioned in Lesson 1.

Lesson 4. Mapping quality is based on the engineering gap. As noted in Section 3, each control mapping has some gap that a designer must address to reuse the control to satisfy a legal requirement. That is, even if every legal requirement may have a number of controls mapped to it, there may still exist some gap between the legal requirement and control requiring additional refinement or elaboration on the part of the designer. In some cases this gap may be minimal; for example, specifying retention durations for a control regarding backups. In many cases, however, this gap will be far larger, and could include developing or refining large-scale organizational policies, or re-engineering a function from one problem space, such as security incident auditing, to another problem space, such as medical record data quality.

B. Towards a Capability-based Approach

In this study, we discovered a capability-based approach that we believe could ease the mapping process for the analyst. A capability is the power or ability to accomplish a goal. Unlike goals, capabilities describe what an organization can do given their current resources. Regulations impose the need to develop, acquire or maintain a capability, whereas goals describe desires, wants and needs. When mapping legal requirements to IT controls, we recognize that each requirement and control describes multiple capabilities and that these capabilities interact at different levels of detail. Consider legal requirement **NY-34**:

> **NY-34:** shall preserve patient records including X-ray films or reproduction thereof safely for a minimum of seven years following discharge of the patient

To identify capabilities, we must identify primitive actions that express what must be done. This includes "preserve patient records" and "discharge patient." Quality

attributes and other constraints can be identified that apply to these capabilities, as well, such as "safely" and "a minimum of seven years;" however, we are primarily concerned with identifying capabilities. This legal requirement was mapped to the following ISO control:

> **ISO 15.1.3:** Important records should be protected from loss, destruction, and falsification, in accordance with statutory, regulatory, contractual, and business requirements

This control yields several capabilities, which an organization that has implemented this control could potentially reuse or adapt. This includes "avoid record loss; avoid record destruction; avoid record falsification; identify regulatory requirements; identify contractual requirements." These capabilities may be assigned to existing software components to provide this functionality, or it may be assigned to employee expertise. For example, a company's in-house legal counsel could have the ability to identify regulatory requirements, whereas a software developer may have worked on multiple projects that include avoiding record falsification using Public Key Cryptographic methods. Having this experience can reduce cost, as these employees can simply adapt this knowledge and experience to the context of a new regulation. Returning to **NY-34**, the most relevant capability in this mapping is "preserve patient records" and "avoid record loss." How these capabilities are implements may require additional considerations that we defer to the practice of system design.

The challenge in using this approach is that it does not relieve the analyst from requiring considerable domain knowledge. In Table II, we present several boundary cases to illustrate the type of domain knowledge required to use this approach. In the first column is the row index; in the second and third columns, we present the unique requirement and control index, followed by the capability phrase; in the fourth column, we present the delta as a goal-based relationship between the requirement and control: "R" means that the requirement capability is refined by the control capability; "G" means that the requirement capability can partially fulfill the IT control capability; and "E" means the two capabilities are more or less equivalent.

TABLE II. CAPABILITY MAPPING WITH DELTA CODING

#	Legal Requirements	ISO 27002 Controls	Δ
1	**CA-12**: Complete records	**12.2.2**: Validate information	G
2	**CA-12**: Sign records	**11.2.3**: Allocate passwords	R
3	**NY-33**: Record entry user	**10.10.1**: Log user activity	E
4	**NY-36**: Sample records	**10.10.2**: Review monitoring results	G
5	**HIPAA-67**: Authorize workforce	**6.1.3**: Allocate privileges	E
6	**ITR-12**: Obtain consent	**15.1.4**: Ensure privacy	G
7	**ITR-53**: Demonstrate control compliance	**13.2.3**: Collect legal evidence	E

In rows 1, 4 and 6, we determined that the IT control is a goal that is partially fulfilled by the legal requirement. In row 6, for example, the India ITR-12 requirement requires a company to obtain consent from the data provider before collecting the data. This is one capability that can be used to ensure privacy; other capabilities include encrypting the data and providing individual's access to view and correct their information. In row 4, the requirement requires a company to sample records as a quality assurance mechanism; once these records are sampled, it is assumed that some actor in the organization will review the samples to check for anomalies. This review capability, which is required by the ISO control, may be reused or adapted.

Refinements describe how to use an IT control capability to partially fulfill the legal requirement. In Table II, row 2, the allocating passwords implies the ability to authenticate users, which could then be used to have those users sign records to fulfill the legal requirement. Other mappings are more or less equivalent: the capability of collecting legal evidence may provide guidance to an organization that also has to collect evidence of control compliance; however, the audience to which these types of evidence will be presented are likely different (e.g., courts or attorneys as opposed to IT system auditors). Thus, equivalence is not an assurance that the IT control is a direct substitute for the legal requirement.

C. Control Mapping and Goals

Regulations and the legal requirements they contain often address implied goals relevant to the context for which they are enacted. In our analysis, we discovered that the presence of gaps between a legal requirement and a control could be detected by differences between the high-level goals they satisfy. For example, consider **CA-5** regarding retention of records for minors. This requirement was aligned by one analyst with MP-4 from NIST:

> **CA-5**: shall keep the records of unemancipated minors at least one year after such minor has reached the age of 18 years and, in any case, not less than seven years

> **MP-4:** The organization: a. Physically controls and securely stores [Assignment: organization-defined types of digital and/or non-digital media] within [Assignment: organization-defined controlled areas] using [Assignment: organization-defined security safeguards]...

A possible goal for this requirement could be to give young adults, who may be independent from their parents for the first time (see G1 in Figure 1), a period during which the individual will take on the responsibility of managing their own healthcare, which may include making regular visits to a medical professional. However, this goal differs from that of the implied goal for the NIST control, which is to generally ensure reasonable security and privacy (see G2 in Figure 1). This is not unexpected, as controls are designed to be applicable across industries, versus legal requirements that target specific social problems; as

addressed in Section V.A. Although each goal could be achieved by implementing similar functional requirements regarding data retention, their corresponding goals are dissimilar. Figure 1 shows how these implied goals for the legal requirement and control can be mapped to functional requirements. When a functional requirement is adopted in this fashion, the organization is effectively making the right decision for the wrong reason – and the implications of the goals may be different should that reason be considered in other contexts.

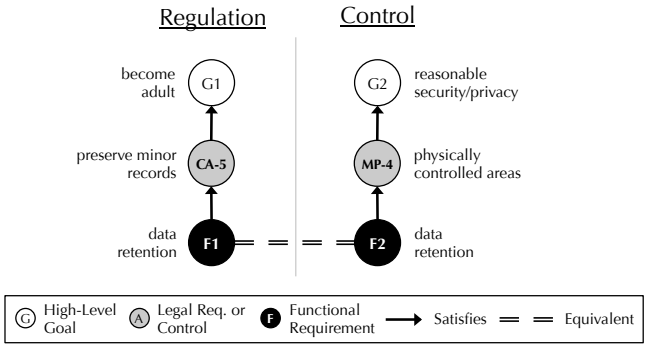

Figure 1: Dissimilar goals satisfiable by functional requirements regarding data retention

This can also be seen in requirement **CA-12**, which specifies that medical records be completed within two weeks of the patient's discharge:

> **CA-12:** [an entity] shall ensure that medical records be completed promptly and authenticated or signed by a licensed healthcare practitioner acting within the scope of his or her professional licensure within two weeks following the patient's discharge.

> **IA-2:** The information system uniquely identifies and authenticates organizational users (or processes acting on behalf of organizational users).

Although this requirement could be satisfied by functional requirements related to authentication and logging, one of its potential goals could be to avoid medical malpractice (see G3, in Figure 2) by ensuring that records are completed while the medical professional has recent knowledge of the patient. This goal does not relate to reasonable security and privacy, despite that both could be satisfied by similar underlying functional requirements (see F3-F6, in Figure 2). It is possible that the control may not have the additional logic necessary to do computations with logging data or provide ancillary mechanisms, such as notification, because they are aligned with a different goal.

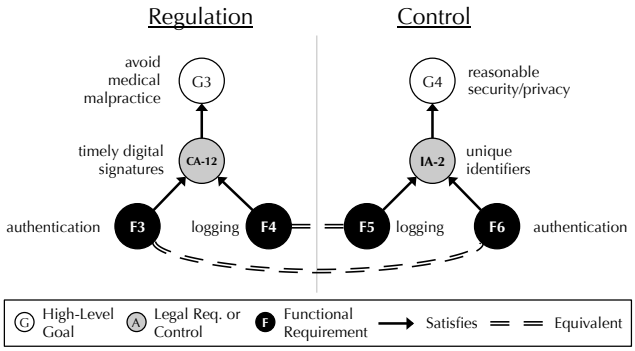

Figure 2: Dissimilar goals satisfiable by functional requirements regarding authentication and logging

D. Evaluation of IT Control Mappings

After the mappings were completed, we carried out a detailed analysis and comparison in order to identify overlaps and conflicts between them, and to justify why these occur. Consider the legal data retention requirements **CA-4** and **CA-5** (see also section IV.A).

> **CA-4:** shall preserve patient records including X-ray films or reproduction thereof safely for a minimum of seven years following discharge of the patient.

These requirements were mapped by the experts to subtly different IT controls in CCM, namely by one expert to DG-02 and DG-04, and by another expert to DG-03 and DG-05:

> **DG-02:** Data, and objects containing data, shall be assigned a classification based on data type, jurisdiction of origin, ...
>
> ...jurisdiction domiciled, context, legal constraints, contractual constraints, value, sensitivity, criticality to the organization and third party obligation for retention and prevention of unauthorized disclosure or misuse.
>
> **DG-03:** Policies and procedures shall be established for labeling, handling and security of data and objects which contain data. Mechanisms for label inheritance shall be implemented for objects that act as aggregate containers for data.
>
> **DG-04:** (v1.1) Policies and procedures for data retention and storage shall be established and backup or redundancy mechanisms implemented to ensure compliance with regulatory, statutory, contractual or business requirements. Testing the recovery of backups must be implemented at planned intervals.
>
> **DG-05:** Policies and procedures shall be established and mechanisms implemented for the secure disposal and complete removal of data from all storage media, ensuring data is not recoverable by any computer forensic means.

It is worth noting that the two experts mapped to two disjoint sets of related, but different, controls. This reflects

the assumptions they made, namely, that a proper classification of patient records is necessary so that archival can proceed effectively, and that secure disposal is necessary after the retention period has lapsed. However, while these requirements do relate to **CA-4**, it is likely that **CA-4** and **CA-5** could be satisfied by an organization implementing only DG-04.

Next we consider an example where the mappings performed were conflicting and/or more debatable. As shown previously (see left column), legal requirement **CA-12** is concerned with authentication. For this requirement, there were diverging views as to which controls were adequate. Among two of the three who studied it, there was agreement that CCM control IS-07 was suitable, although the third expert found this control only vaguely relevant. All three experts did agree that this requirement can only be met by having proper workflows and procedures in place in an organization, above and beyond just technical IT controls. One expert suggested that CCM control DG-01 might actually be preferable. These two controls are described as follows.

> **IS-07:** User access policies and procedures shall be documented, approved and implemented for granting and revoking normal and privileged access to applications, databases, and server and network infrastructure in accordance with business, security, compliance and service level agreement (SLA) requirements.
>
> **DG-01:** All data shall be designated with stewardship with assigned responsibilities defined, documented and communicated.

It is interesting to note that the first of these two controls focuses on access control specifically, thus covering only part of the legal requirement (not including signing), while the second control is generic and overspecifies the requirement. Interestingly, the requirement covers aspects that are not entirely related to security, and have more to do with process.

The CCHIT EHR Criteria, however, do mention IT controls that would be technically superior in meeting the CA-12 requirement, while explicitly noting that electronic signatures have not been standardized for use in this problem domain, e.g.:

> **AM 08.06** The system shall provide the ability to cosign a note and record the date and time of signature.

With reference to this and related controls, the CCHIT document explicitly notes *"*The words, "sign," "signature," "cosign," and "cosignature" are intended here to convey actions, rather than referring to digital signature standards. It is recognized that an electronic signature is useful here. However, a widely accepted standard for electronic signatures does not exist. Thus, the criteria calls for documenting the actions of authenticated users at a minimum. In the future, when appropriate digital signature standards are available, certification criteria may be introduced using such standards. ASTM has developed "2003 Updated ASTM Standard Guide for Electronic Authentication of Health Care Information" to address some of these issues."

There is a requirement in HIPAA concerning authorization and access to healthcare data, which also caused divergence during the analysis of our mappings:

> **HIPAA-67:** [an entity must] implement procedures for the authorization and/or supervision of workforce members who work with electronic protected health information or in locations where it might be accessed.

For this requirement, two completely different (but justifiable) mappings were produced. One mapping was to CCM controls IS-07 (which we have come across already in this section) and SA-13; the other mapping was to controls DG-01 (also mentioned above), IS-01, IS-03, IS-14. Of these controls it is interesting to notice the appearance of IS-14, as follows:

> **IS-14:** Managers are responsible for maintaining awareness of and complying with security policies, procedures and standards that are relevant to their area of responsibility.

Clearly, this control explicitly accounts for the 'supervision' aspect of the legal requirement, which is not covered directly by other controls.

As is evident from this discussion, an evaluation of the IT control mappings yields subtleties and different, often overlooked facets of requirements. Clearly any attempt to undertake such a control mapping must undergo extensive evaluation and verification.

E. Towards Tools to Guide Control Mapping in Organizations

In an organizational setting, we envisage the need for tools incorporating predefined mappings from regulations to implementable controls in such a way that they can be presented to experts for the purposes of risk assessment and decision support. For example, a company's legal team regularly needs to ensure compliance with the latest data protection regulations across all business processes; a tool which can automatically advise which controls need to be in place to achieve such compliance is essential.

We believe that some of the heuristics we have considered in this paper can be incorporated into a software tool that guides experts through the mapping of legal and regulatory healthcare requirements to IT controls. Certainly maintaining records and categories of all the applicable controls can be assisted significantly in such a way. Areas where conflict or confusion can occur could be flagged to an expert user and additional input provided, so that suitable IT controls can be selected to meet specified requirements in a semi-automated fashion. We are developing aspects of this

approach, including tools that allow decision makers to assess the risk level of undertaking a new project given the controls already implemented within the organization's IT infrastructure.

Tools such as these can either make use of a database of predefined mappings (such as those generated by the analysis presented in earlier sections), or can use "intelligent" algorithms to identify controls that are considered most relevant to a given regulation. The latter approach involves implementing natural language processing and machine learning techniques, which are outside the scope of this particular work, but have been developed to some extent in our earlier work [12]. We are planning to build databases of predefined mappings, while also providing helper tools for editing and maintaining these databases to reflect changes in laws and regulations.

Among the tools that are currently being developed, of note are (a) a visualization tool for the different layers of regulations and how they map to controls, (b) a verification tool to support the coverage model presented by Gordon and Breaux [10], which may be used to determine regulatory coverage for IT organizations.

VI. Summary and Conclusions

In this paper we have presented an approach to mapping legal and regulatory requirements for a particular problem domain (namely, electronic health records) to concrete security controls as defined in four different control catalogues. We discussed our methodology and current results and identified gaps, areas of overlap, and areas of conflict/confusion when mapping requirements to controls. Future work will include validating our results by consulting with additional practitioners from industry, and developing practical software tools to assist the mapping process.

Acknowledgment

This research was supported by Hewlett-Packard Labs Innovation Research Program (Award #CW267287).

References

[1] T.D. Breaux, "A method to aquire compliance monitors from regulations," *3rd IEEE Int'l W'shp Req'ts Engr. & Law*, pp. 1-10, 2010.

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

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

[4] T. D. Breaux and D. G. Gordon, "Regulatory requirements traceability and analysis using semi-formal specifications," *Int'l Wk'ing Conf. Req'ts Emgr.: Fnd. Soft. Q.*, 2013.

[5] E. Brynjolfsson, L.M. Hitt, "Beyond the productivity paradox: computers are the catalyst for bigger changes." Comm. ACM, 41(8): 49-55, 1998.

[6] J. Cleland-Huang, A. Czauderna, M. Gibiec, J. Emenecker. "A machine learning approach for tracing regulatory codes to specific product requirements," *Proc. 32nd ACM/IEEE International Conference on Software Engineering*, pp. 155-164. 2010.

[7] C. Feltus, E. Dubois, M. Petit. "Conceptualizing a responsibility-based approach for elaborating and verifying RBAC policies conforming with CobiT framework requriements." *3rd IEEE Int'l W'shp on Req'ts Engr. & Law*, pp. 34-43, 2010.

[8] R.A. Gandhi, S.W. Lee, "Discovering Multidimensional Correlations among Regulatory Requirements to Understand Risk," *ACM Trans. Soft. Engr. Method.* 20(4): Article 16, 2011.

[9] P. Giorgini, F. Massacci, J. Mylopoulos, N. Zannone, "Modeling security requirements through ownership, permission and delegation," *13th IEEE Int'l Req'ts Engr. Conf.*, pp. 167-176, 2005.

[10] D.G. Gordon, T.D. Breaux, "Assessing Regulatory Change through Legal Requirements Coverage Modeling", To Appear: IEEE Req'ts Engr. Conf., 2013.

[11] C.B. Haley, R.C. Laney, J.D. Moffett, B. Nuseibeh. "Using trust assumptions with security requirements." *Req'ts Engr. J.* 11: 138-151, 2006.

[12] S. Islam, H. Mouratidis, J. Jürjens, "A framework to support alignment of secure software engineering with legal regulations," *Softw Syst Model* (2011) 10:369–394

[13] N. Papanikolaou, Natural Language Processing of Rules and Regulations for Compliance in the Cloud." *Proceedings of DOA-SVI 2012*, 2012.

[14] S. Sackmann, M. Kahmer, M. Gillot, L. Lowis. "A classification model for automating compliance." *10th IEEE Conf. E-Comm. Tech. & 5th IEEE Conf. Ent. Comp., E-Comm. & E-Serv.*, pp. 79-86, 2008.

[15] P. Tambe, L.M. Hitt, E. Brynjolfsson, "The extroverted firm: how external information pracitces affect innovation and productivity." To Appear: *Management Science*, 2011.

[16] R.K. Yin. *Case study research*, 4th ed. In Applied Social Research Methods Series, v.5. Sage Publications, 2008.