ABSTRACT

SCHMIDT, JESSICA YOUNG. Specifying Requirements Using Commitment, Privilege, and Right (CPR) Analysis. (Under the direction of Annie I. Antón.)

Organizations have many documents, including policy documents (e.g., privacy policies and terms of use) and data use agreements (DUAs), with which their software must comply. Requirements engineers must incorporate these documents into the requirements phase of software in order to build in compliance from the start. In the United States, the Federal Trade Commission is empowered to monitor organizations’ compliance with the practices expressed in the organizations’ policy documents. Therefore, organizations must ensure that their software systems comply with these documents, which must serve as a primary source of compliance requirements early on in the software lifecycle. Regulations created pursuant to the U.S. Health Insurance Portability and Accountability Act specify that a DUA must exist for certain uses and disclosures of protected health information as a limited data set. For compliance reasons, it is important for requirements engineers to ask for and evaluate DUAs, as they are legally binding on the parties.

Herein, we present CPR (commitment, privilege, and right) analysis, our approach to acquiring compliance requirements from policy documents and DUAs. We present how we developed CPR analysis. We also describe the CPR analysis methodology, which includes a set of heuristics to guide analysts through each step of the methodology. We discuss our validation of the approach through its application to additional documents; comparison to a goal-based approach; its application by others; and its application in the development of a health system prototype to manage health data from a number of sources to support analytical decision-making. We found CPR analysis performed better in terms of ensuring compliance than goal-based analysis did. We also discovered that other requirements engineers could effectively apply CPR analysis to identify compliance requirements.
© Copyright 2012 by Jessica Young Schmidt

All Rights Reserved
Specifying Requirements Using Commitment, Privilege, and Right (CPR) Analysis

by
Jessica Young Schmidt

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Computer Science
Raleigh, North Carolina
2012

APPROVED BY:

__________________________  ___________________________
Julia B. Earp                  Munindar P. Singh

__________________________  ___________________________
Laurie Williams               Annie I. Antón
Chair of Advisory Committee
DEDICATION

To my amazing parents and husband, who have always supported me.
BIOGRAPHY

Jessica Young Schmidt grew up in Natural Bridge—a gorgeous little town in the mountains of Virginia. Since leaving the Shenandoah Valley for Raleigh, NC and Arlington, MA, she has realized just how lucky she was to grow up in such an amazing place. She was spoiled by the gorgeous views and tons of space to play. All of which led to her love of the outdoors—hiking, kayaking, mountain biking, running, snowboarding, fishing, playing soccer, and basically anything that lets her be outside with friends and family. Last year she married her husband Matt, a fellow NC State Computer Science Ph.D., on her family’s farm surrounded by family, friends, and, of course, the mountains.

Jessica graduated from Rockbridge County High School in 2003. She then attended Roanoke College where she double majored in Computer Science and Mathematics. She received her Bachelor of Science from Roanoke College in May 2007. After graduating from Roanoke College, she began graduate work at North Carolina State University. In December 2009, she received her Master of Science in Computer Science from NC State. In 2012, she received her Ph.D. in Computer Science from NC State. Jessica will be a full-time lecturer for the College of Computer and Information Science at Northeastern University starting in the fall of 2012.
ACKNOWLEDGEMENTS

The brick walls are there for a reason. The brick walls are not there to keep us out. The brick walls are there to give us a chance to show how badly we want something. Because the brick walls are there to stop the people who don’t want it badly enough. They’re there to stop the other people.

- Randy Pausch, The Last Lecture

Just keep swimming. Just keep swimming.

Just keep swimming, swimming, swimming. What do we do? We swim, swim.

- Dory, Finding Nemo

I believe these two quotes accurately explain my journey through grad school. It all started with the brick wall—my desire to get a Ph.D. and teach in hopes of inspiring students the way my professors inspired me. The journey wasn’t always easy but there was a purpose, and sometimes the only thing I could do was just keep swimming...

First, I want to thank my family and friends, who have always been there for me and will always be my top priority. Without them, life wouldn’t be nearly as fun, enjoyable, or worthwhile. I don’t regret any times I spent with them, even if at times I probably should have been working. They have made me the person I am today, and I love them dearly.

To Matt: You are absolutely amazing! I cannot even begin to say how thankful I am that you were persistent in getting me to come out with the group for drinks and in asking me out on a date without really asking me out on a date. You are the best thing that happened to me while at State. You managed to keep me sane, or at least somewhat sane, through grad school and everything else that has happened over the past four years. You were there for it all—success, excitement, screaming, crying—and always knew the right thing to say. And because you are as big of a nerd as I am, whenever I hit a roadblock with my research, you were there to talk through it with me. I love you!

To Mom and Dad: You have always been there to encourage me. With every success over the past 27 years, big or small, you have been there to celebrate with me and do the *happy dance.* I cannot thank you enough for your encouragement and the freedom to choose my own way. No matter what, each day I look forward to my call, or multiple calls, with you; they always brighten my day. You are so amazing! I want to be just like y’all when I grow up!
To Karla: Thanks for being such an amazing little sister! You excel at everything you do, which makes me incredibly proud and motivates me to be better.

To Cody: During tough times, you were exactly what the family needed.

To the Young, Wood, Gibson, Schmidt, and Steltenpohl families: I feel extremely lucky to have such large and loving families. Thank you for your support!

To Emily, Ashleigh, Kara, Amy, Haley, and Jess: Not only did you make my time at Roanoke amazing, but years later you are still my closest friends. You always remind me that having fun and being goofy are important because “everybody is somebody else’s weirdo.”

To Paul, Kelly, Brandon, April, Prachi, and my other Raleigh friends: Thanks for making Raleigh so enjoyable!

To Team LTS: Thanks for making my summer in Maryland so fun.

Next, I want to thank all of the people at NC State who supported me over the years. These individuals helped me with research, teaching, and grad school in general.

To Dr. Antón: Thank you for your advice, support, and encouragement over the past five years. You helped me to become a more confident researcher and presenter. I really appreciate all of the time you have spent helping me with my research, especially the time you spent performing goal analysis for my comparison study. Thank you as well for all of the semesters that you funded me as your RA.

To Dr. Earp: Thank you for all of the time you spent performing goal analysis for my comparison study and for always being available to answer any research questions I had, especially those involving statistics.

To Dr. Singh: Thank you for your advice and expertise in helping me to improve my research, particularly my definitions.

To Dr. Williams: Thank you for your advice and encouragement in applying CPR analysis within the development of the health system prototype. I also wanted to thank you for the semesters you funded me as an RA.

To The Privacy Place members and my labmates, especially Aaron, Jeremy, Ramya, Andrea, Lauren, Mei, Andy, Sarah, and Lucas: Thanks for your suggestions with my research and for always being there to distract me from my work (or, more likely, let me distract you from your work).
To Aaron: On my first day at State, you were tasked with showing me around the lab and introducing me to the research of the group. Five years later... and you are still an amazing resource to me—always answering my questions, as silly as they may be. Thanks for being a great labmate and friend!

To Paul: You were already in law school by the time I started at State, but you have been a ginormous help to me. Thank you so much for putting up with all of my emails that contained countless legal questions. Your help has proved to be essential for positioning my work and showing the relation to legal concepts. Thanks!

To Sarah: Thank you for all of your guidance with teaching CSC 116. Without you, the experience would not have gone as well as it did.

To Dr. Thuente and Dr. Reeves: Thank you for providing me with the opportunity to teach.

To Margery and Carol: Thank you for all of your help. Your help and advice have saved me countless hours of headaches over these past five years.

I also want to thank my professors at Roanoke who not only persuaded me to major in Computer Science but also inspired me to get my Ph.D. and pursue a career in teaching.

Finally, I want to once again sincerely thank everyone that helped me along the way with support, encouragement, and guidance because I would not have been able to make it to this point without each of you.
# TABLE OF CONTENTS

| List of Tables | x |
| List of Figures | xi |

## Chapter 1 Introduction
1. Motivation
  1.1 Policy Documents
  1.2 Data Use Agreements
2. Related Work
  2.1 Policies and Requirements
  2.2 Goal-Based Analysis
  2.3 Traceability
  2.4 Compliance
  2.5 Natural Language
  2.6 Commitments
  2.7 User Studies
3. Overview of Remaining Chapters

## Chapter 2 CPR Analysis Theory and Methodology
2.1 Theory of Commitments, Privileges, and Rights
  2.1.1 Classification Terminology
  2.1.2 Rationale Terminology
  2.1.3 Terminology Comparison
2.2 CPR Analysis Methodology
  2.2.1 Parse
  2.2.2 Classify
  2.2.3 Operationalize
3. Summary

## Chapter 3 CPR Analysis Heuristics
3.1 Parse
3.2 Classify
  3.2.1 Item Heuristic (IH)
  3.2.2 Helper Heuristics (HH)
  3.2.3 Classifying Heuristics (CH)
  3.2.4 Documenting Attribute Heuristics (DAH)
  3.2.5 Documenting Rationale Heuristics (DRH)
  3.2.6 Forming CPR Heuristic (FH)
3.3 Operationalize
3.4 Summary

vii
## Chapter 4 Multi-Case Studies

4.1 Formative Multi-Case Study
   - Research Questions
   - Documents Analyzed
   - Results

4.2 Summative Multi-Case Study
   - Research Questions
   - Documents Analyzed
   - Results

4.3 Threats to Validity

4.4 Summary

## Chapter 5 Examining CPR Analysis for Ensuring Compliance

5.1 Methodology
   - Goal-based Analysis
   - Differences Between Analysis Approaches

5.2 Preserving Context and Intended Meaning

5.3 Case Study Results
   - Language
   - Specificity
   - Modality

5.4 Threats to Validity

5.5 Discussion

## Chapter 6 Empirical User Study

6.1 Experimental Design
   - Design
   - Subjects
   - Materials
   - Procedure
   - Measurements

6.2 Pilot Study

6.3 Data Analysis
   - Subject Responses to Open-ended Questions for Analysis Tasks
   - Results of Empirical Evaluation

6.4 Threats to Validity

6.5 Summary

## Chapter 7 CPR Analysis within the Development of a Health System Prototype

7.1 Overview of HealthSystem

7.2 Requirements
   - Software Requirements
   - Operational Requirements
   - Data Custodian Requirements
7.3 DUA Template ......................................................... 125
7.4 Summary ............................................................. 125

Chapter 8 Conclusion .................................................. 127
  8.1 Chapter Summaries ................................................... 127
  8.2 Contributions ....................................................... 129
    8.2.1 Primary Contributions ....................................... 129
    8.2.2 Demonstrated Experiences ................................. 129
  8.3 Future Work ....................................................... 130

References .............................................................. 131

Appendices .............................................................. 141
  Appendix A CPR Analysis Heuristics .............................. 142
    A.1 Parsing Heuristics ............................................. 142
    A.2 Classification Heuristics .................................... 143
    A.3 Operationalization Heuristics .............................. 150
  Appendix B User Study Materials & Data ........................ 151
  Appendix C DUA Template ............................................ 156
LIST OF TABLES

Table 6.1 Subject Demographics ................................................. 93
Table 6.2 Subject Software Engineering Experience ......................... 94
Table 6.3 Artifacts Submitted by Subjects ..................................... 96
Table 6.4 Possible Coding Combinations for Related Pairs ................... 99
Table 6.5 Possible Codings for Subject-Produced Requirements Not in Related Pair. 101
Table 6.6 Correctness and Completeness for Conditions ....................... 108

Table 7.1 Requirements Counts for the HealthSystem Prototype .............. 120
Table A.1 Parsing Heuristics (PH) ................................................ 142
Table A.2 Item Heuristic (IH) ..................................................... 143
Table A.3 Helper Heuristic (HH) .................................................. 143
Table A.4 Classifying Heuristic (CH) ........................................... 143
Table A.5 Classifying Heuristic for Policy Documents (CH_{PD}) - Part 1 .... 144
Table A.6 Classifying Heuristic for Policy Documents (CH_{PD}) - Part 2 .... 145
Table A.7 Classifying Heuristics for DUAs (CH_{DUA}) ........................ 146
Table A.8 Documenting Attribute Heuristics (DAH) ............................ 147
Table A.9 Documenting Rationale Heuristics for Policy Documents (DRH_{PD}) .... 148
Table A.10 Documenting Rationale Heuristics for DUAs (DRH_{DUA}) ........... 149
Table A.11 Forming CPR Heuristic (FH) .......................................... 149
Table A.12 Requirement Type Heuristic (RTH) .................................. 150
Table A.13 Operationalizing Heuristics (OH) .................................... 150

Table B.1 User Study Data Discussed in Section 6.3 .......................... 155
LIST OF FIGURES

Figure 1.1  NIH Summary of the HIPAA §164.514(e)(4) [104]  ...................................... 4
Figure 1.2  Excluded identifiers in an LDS—45 C.F.R. §164.514(e)(2)  ....................... 5
Figure 2.1  CPR Analysis Methodology ................................................................. 20
Figure 2.2  Types of Items within a Document ..................................................... 21
Figure 4.1  Policy Documents Analyzed in Formative Study ................................. 56
Figure 4.2  DUAs Analyzed in Formative Study ..................................................... 57
Figure 4.3  Formative Study - Policy Documents Case - Types of Items ................. 59
Figure 4.4  Formative Study - Policy Documents Case - Classifications ............... 60
Figure 4.5  Formative Study - Policy Documents Case - Rationale ....................... 61
Figure 4.6  Formative Study - Policy Documents Case - Actors ........................... 61
Figure 4.7  Formative Study - DUA Case - Types of Items ................................. 62
Figure 4.8  Formative Study - DUA Case - Classifications ................................. 63
Figure 4.9  Formative Study - DUA Case - Rationale .......................................... 64
Figure 4.10 Formative Study - DUA Case - Actors .............................................. 64
Figure 4.11 Policy Documents Analyzed in Summative Study .............................. 66
Figure 4.12 DUAs Analyzed in Summative Study .................................................. 67
Figure 4.13 Summative Study - Policy Document Case - Types of Items ................ 68
Figure 4.14 Summative Study - Policy Document Case - Classifications ............... 69
Figure 4.15 Summative Study - Policy Document Case - Rationale ....................... 69
Figure 4.16 Summative Study - Policy Document Case - Actors ........................... 70
Figure 4.17 Summative Study - DUA Case - Types of Items ................................ 71
Figure 4.18 Summative Study - DUA Case - Classifications ................................ 71
Figure 4.19 Summative Study - DUA Case - Rationale .......................................... 72
Figure 4.20 Summative Study - DUA Case - Actors .............................................. 73
Figure 4.21 Multi-Case Studies - Types of Items .................................................. 74
Figure 4.22 Multi-Case Studies - Classifications .................................................. 75
Figure 4.23 Multi-Case Studies - Rationale ......................................................... 76
Figure 4.24 Multi-Case Studies - Actors ............................................................... 77
Figure 5.1  Analyzed Policy Documents ............................................................... 80
Figure 6.1  Total Number of Subject-Produced Requirements by Condition and Subject 104
Figure 6.2  Total Number of Expected Requirements to which Subject-Produced Requirements Related ................................................................. 105
Figure 6.3  Correctness by Condition and Subject ............................................... 106
Figure 6.4  Completeness by Condition and Subject ......................................... 107
Figure 6.5  Total Number of Related Pairs with No Coding ............................... 108
Figure 6.6  Total Number of Related Pairs with $Mult$ Coding ......................... 109
Figure 6.7  Total Number of Related Pairs with $PoV/Mul$ Coding .................... 110
Figure 6.8  Total Number of Subject-Produced Requirements with $NP$ Coding .... 111
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 6.9</td>
<td>Derived Requirements</td>
<td>111</td>
</tr>
<tr>
<td>Figure 6.10</td>
<td>Synthesized Requirements</td>
<td>112</td>
</tr>
<tr>
<td>Figure 7.1</td>
<td>Classifications of Classifiable Items in HealthSystem DUAs</td>
<td>118</td>
</tr>
<tr>
<td>Figure 7.2</td>
<td>Actors of Classifiable Items in HealthSystem DUAs</td>
<td>118</td>
</tr>
<tr>
<td>Figure 7.3</td>
<td>Classification/Actor Combinations of Classifiable Items in HealthSystem DUAs</td>
<td>119</td>
</tr>
<tr>
<td>Figure B.1</td>
<td>User Study Questionnaire</td>
<td>152</td>
</tr>
<tr>
<td>Figure B.2</td>
<td>Problem Analysis Questions</td>
<td>153</td>
</tr>
<tr>
<td>Figure B.3</td>
<td>Requirements Analysis Questions</td>
<td>153</td>
</tr>
<tr>
<td>Figure B.4</td>
<td>SocialNet Policy Document</td>
<td>154</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

You never have a second chance to make a good first impression.
- Anonymous

The research presented herein addresses the problem of developing compliant software systems. Specifically, given that organizations’ policy documents and data use agreements express commitments, privileges, and rights that can be legally binding, requirements engineers must specify these commitments, privileges, and rights as software requirements. In order to do so, requirements engineers must incorporate these documents into their requirements activities to ensure that the requirements fulfill the items within the documents. We incorporate policy documents and data use agreements (DUAs) into the requirements phase of the software lifecycle through the use of CPR analysis. During CPR (commitment, privilege, and right) analysis, requirements engineers extract commitments, privileges, and rights from documents and operationalize them as requirements. The objective of this chapter is to provide motivation for the inclusion of policy documents and DUAs in the requirements phase of the software lifecycle and a summary of related work.

1.1 Motivation

Software requirements are typically obtained from a variety of sources. In the realm of software compliance, requirements sources often include policies, law, and contracts [27, 30, 33, 93, 94, 95, 96, 109, 118, 132, 134]. These sources are particularly significant because there are legal consequences for noncompliance. Requirements engineers must leverage any documents they receive: they are the first line of defense for ensuring that systems comply with governing documents. In order to ensure compliance, we must employ organizations’ documents early on
in the software lifecycle [19]. Herein, we focus on two types of documents with which software systems must comply—policy documents and DUAs. We will now discuss each type of document and the motivation for complying with each.

1.1.1 Policy Documents

We use the term policy document to refer to any policy that an organization has. The policy documents we examined were all posted on the organizations’ websites. These policy documents describe the organizations’ practices, including how consumers’ or users’ personal information will be collected, disclosed, protected, shared, stored, and used. Organizations post different types of policy documents online, including privacy policies, privacy statements, Internet privacy statements, web privacy statements, notices of information practices, notices of privacy practices, terms of use, and terms and conditions.

In the United States, the Federal Trade Commission (FTC) is empowered to monitor organizations’ compliance with the practices expressed in their public policy documents. The FTC Act prohibits “unfair or deceptive acts or practice” [70]. This means the FTC can hold organizations liable for the statements they make in their policy documents; therefore, organizations need to comply with their policy documents [70]. In order to build policy-compliant system, policy documents need to be considered as a primary source of requirements.

Recent sanctions by the FTC against organizations that fail to comply with their policies highlight the importance of actively addressing compliance as a best practice that avoids such penalties [37, 110]. The FTC has investigated policy compliance failures and brought enforcement actions against at least seventeen organizations since 2002 [37]. These noncompliance cases are an important motivation for organizations to evaluate software compliance with their policy documents.

We will discuss two enforcement examples and two investigation examples. The FTC took enforcement actions against ChoicePoint, a data broker, when the personal records of 163,000 consumers were compromised because ChoicePoint did not comply with its privacy policy [67, 110]. Specifically, ChoicePoint did not comply with the following statements in its privacy policy about who would have access to information and its credentialing process [67]:

- ChoicePoint allows access to your consumer reports only by those authorized under the FCRA.
- Every ChoicePoint customer must successfully complete a rigorous credentialing process. ChoicePoint does not distribute information to the general public and monitors the use of its public record information to ensure appropriate use.

As software engineers, we want these statements to be built into the system to ensure compliance. In order to settle with the FTC, ChoicePoint paid $10 million in civil penalties and $5 million in
consumer redress [67]. The settlement also required the implementation of procedures to avoid releasing consumer reports to illegitimate businesses, the creation of an information security program, and audits of the organization for 20 years.

The FTC also took enforcement actions against Gateway Learning Corporation (GLC), which sells educational tools, when the organization did not honor commitments it made to its users within its policy document, including [65, 66, 122]:

- We do not sell, rent or loan any personally identifiable information regarding our consumers with any third party unless we receive a customer’s explicit consent.

- We do not provide any personally identifiable information about children under 13 years of age to any third party for any purpose whatsoever.

Contradicting its policy document, GLC rented consumer information, which resulted in consumers receiving direct mailings and telemarketing calls [65, 66]. The FTC fined GLC the amount of money that it retained from renting user information; and the FTC mandated audits, strict record keeping of policy documents, and employee training [65, 66, 122]. To quote from the FTC decision, GLC is also prohibited from “disclosing to any third party any personal information collected on its Web site prior to the date it posted its revised privacy policy permitting third-party sharing (June 20, 2003) without the express affirmative (‘opt-in’) consent of the consumers to whom such personal information relates” [65].

The FTC investigated Facebook and Google because they did not live up to their promises. Specifically, Facebook told users that they could keep their personal information private, but Facebook shared users’ information that was marked as private [68]. Google came under FTC investigation when it launched Google Buzz, which violated some of Google’s privacy promises [69]. Facebook and Google were not charged with fees to settle their investigations, but, like ChoicePoint and GLC, the FTC required Facebook and Google to implement comprehensive privacy programs and to obtain audits for 20 years among other requirements relating to consumer privacy [68, 69].

Policy documents play an important role in identifying policy compliance requirements. A policy compliance requirement is a requirement that must be met in order for an organization to comply with its policy documents. These requirements can be traced to the policy documents from which they were derived or extracted. Policy compliance requirements can be system-related or operational (related to business practices). Herein, we use compliance requirement to refer to policy compliance requirements.

### 1.1.2 Data Use Agreements

A data use agreement (DUA) is a legally binding agreement between a covered entity and a limited data set recipient that specifies how the limited data set (LDS) will be used and to
whom it will be disclosed [104]. Regulations adopted pursuant to the United States Health Insurance Portability and Accountability Act (HIPAA)\(^1\) specify that a DUA must exist when certain protected health information is used or disclosed. In the context of the HIPAA, DUAs are critical, legally binding agreements that must be examined by requirements engineers in order to develop compliant software.

Under 45 C.F.R. §164.514(e) (part of the HIPAA Privacy Rule), data use agreements must be made between covered entities and limited data set recipients if a limited data set is being used or disclosed. The LDS can be used or disclosed only for the following purposes: research, public health, or healthcare operation. DUAs must contain certain information, described concisely by the National Institutes of Health (NIH) in Figure 1.1.

<table>
<thead>
<tr>
<th>The Privacy Rule requires a data use agreement to contain the following provisions:</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Specific permitted uses and disclosures of the limited data set by the recipient consistent with the purpose for which it was disclosed (a data use agreement cannot authorize the recipient to use or further disclose the information in a way that, if done by the covered entity, would violate the Privacy Rule).</td>
</tr>
<tr>
<td>- Identify who is permitted to use or receive the limited data set.</td>
</tr>
<tr>
<td>- Stipulations that the recipient will</td>
</tr>
<tr>
<td>- Not use or disclose the information other than permitted by the agreement or otherwise required by law.</td>
</tr>
<tr>
<td>- Use appropriate safeguards to prevent the use or disclosure of the information, except as provided for in the agreement, and require the recipient to report to the covered entity any uses or disclosures in violation of the agreement of which the recipient becomes aware.</td>
</tr>
<tr>
<td>- Hold any agent of the recipient (including subcontractors) to the standards, restrictions, and conditions stated in the data use agreement with respect to the information.</td>
</tr>
<tr>
<td>- Not identify the information or contact the individuals.</td>
</tr>
</tbody>
</table>

**Figure 1.1:** NIH Summary of the HIPAA §164.514(e)(4) [104]

An LDS is not the full data set of protected health information (PHI), nor does it qualify as de-identified health information\(^2\) under the HIPAA. An LDS instead “refers to PHI that excludes 16 categories of direct identifiers and may be used or disclosed, for purposes of research, public health, or health care operations, without obtaining either an individual’s Authorization or a waiver or an alteration of Authorization for its use and disclosure, with a data use agreement”


\(^2\)The standards for de-identification of PHI are contained in 45 C.F.R. §164.514(b).
Specifically, an LDS excludes the identifiers listed in Figure 1.2 from the PHI.

- Names
- Postal address information, other than town or city, State, and zip code
- Telephone numbers
- Fax numbers
- Electronic mail addresses
- Social security numbers
- Medical record numbers
- Health plan beneficiary numbers
- Account numbers
- Certificate/license numbers
- Vehicle identifiers and serial numbers, including license plate numbers
- Device identifiers and serial numbers
- Web Universal Resource Locators (URLs)
- Internet Protocol (IP) address numbers
- Biometric identifiers, including finger and voice prints
- Full face photographic images and any comparable images

**Figure 1.2:** Excluded identifiers in an LDS—45 C.F.R. §164.514(e)(2)

Simply reading the HIPAA Privacy Rule, requirements engineers would not know whether they need to care about DUAs. Given the existence of a DUA, however, a requirements engineer knows that the system must support the exchange of an LDS, along with the other terms specified in the legally binding DUA.

As a result of the 2009 Health Information Technology for Economic and Clinical Health (HITECH) Act\(^3\) that amended the HIPAA, KPMG will be auditing 150 covered entities to assess the compliance \([59, 128]\). During the audits, KPMG will check that the covered entities comply with HIPAA and any of their documents, including policy documents and DUAs.

DUAs play an important role in identifying contractual compliance requirements. A contractual compliance requirement is a requirement that is agreed upon by two (or more)\(^4\) parties and is enforceable by law.


\(^4\)Although it is possible for DUAs to be multilateral agreements, our work examines and discusses only bilateral agreements.
1.2 Related Work

We now examine areas of related work for CPR analysis, including: policies and requirements; goal-based analysis; traceability; compliance; natural language; commitments; and user studies.

1.2.1 Policies and Requirements

Researchers have previously addressed the similarities and differences between policies and software requirements [21]. Policies and software requirements both convey intended outcomes, rather than fact. Policies, however, are broader in scope than software requirements. Whereas policies can govern multiple systems, software requirements are typically specified for one system. Because of the similarities between requirements and policies, policies are amenable—indeed, even well suited—to analysis using traditional requirements engineering techniques [23]. We derive software requirements from policy documents and DUAs that govern systems or websites.

1.2.2 Goal-Based Analysis

There are several goal-based approaches within requirements engineering [18, 20, 22, 24, 31, 35, 36, 73, 74, 135]. Dardenne et al. developed a goal-directed approach for requirements acquisition—Knowledge Acquisition in autOmated Specification, KAOS [56]. The approach includes three modeling levels: the domain-independent meta level, the domain level, and the instance level. Yu’s $i^*$ framework models goals and other intentional elements (tasks, resources, and softgoals) with the relationship between actors and intentional elements [135]. Giorgini et al.’s Secure Tropos framework supports security requirements modeling and analysis in terms of three relationships—ownership, trust, and delegation—between actors and services (goals, tasks, or resources) [74].

The Goal-Based Requirements Analysis Method (GBRAM) aids requirements engineers in identifying goals from requirements sources [18]. Applying the GBRAM results in goals that are operationalized as requirements using schema templates. Antón and Earp extended the GBRAM to support goal-mining (the extraction of pre-requirements goals from post-requirements text artifacts) to derive privacy- and security-related goals from requirements source documents [20]. Goal-mining has been successfully applied to analyze e-commerce, financial, and healthcare privacy policies [20, 22, 24]. Goal-mining is the goal-based approach against which we compared CPR analysis with respect to the ability for each approach to ensure compliance as discussed in Chapter 5.

Antón and Earp introduced a taxonomy of website privacy requirements [20]. The taxonomy is comprised of privacy protections and privacy vulnerabilities. Organizations express how they protect consumer privacy through protection goals [20]. In contrast, the privacy vulnerabilities
describe instances when consumer privacy may be threatened [20]. Antón and Earp developed the taxonomy primarily to aid requirements engineers in ensuring better coverage of privacy requirements by identifying vulnerabilities that need mitigation via new requirements to address these vulnerabilities. In contrast, CPR analysis focuses on aiding organizations in ensuring that they do what their policy documents and DUAs say they do, thus helping organizations to avoid “unfair or deceptive” practices [70], thereby ensuring compliance.

Breaux et al. proposed Semantic Parameterization, a process for representing domain descriptions in first-order predicate logic [35]. They also introduced Knowledge Transformation Language (KTL), a context-free grammar, to analyze the most frequently expressed goals in over 100 online policy documents to derive semantic models [31, 36]. Commitments, rights, and privileges are richer than goals in that they preserve more context from the documents, as we discuss in Chapter 5. Goals examine targets of achievement whereas commitments, privileges, and rights address actions that are required or entitled.

1.2.3 Traceability

Traceability is critical in requirements engineering to link the requirements for the system and the other important entities of the software [54]. To establish software compliance with all governing legal texts, traceability from relevant regulations to requirements specifications is essential [33, 109]. The traceability helps to demonstrate due diligence in a court of law [94]. Regulatory compliance has been examined using automated traceability links [26, 55]. Our work supports both forward and backward traceability across policy documents, DUAs, commitments, privileges, rights, and requirements through documentation.

1.2.4 Compliance

Requirements engineering researchers have investigated software compliance [34, 73, 94, 95, 115]. Robinson identified a need for software requirements to comply with policies and developed a framework,ReqMon, to monitor software requirements at runtime [115]. His approach focused on runtime compliance with system requirements, whereas we focus on the initial derivation of compliance requirements. Ghanavati et al. introduced an approach to track legal compliance; they connected models of business processes with legislation [73]. They used a goal-based approach, whereas we employ a CPR-based approach. They focused on checking compliance through their use of links, while we build in compliance during the requirements activities.

Researchers have examined existing software requirements for compliance with legal texts [94, 95]. Massey et al. proposed an approach to evaluate and improve existing system requirements that must comply with legal texts [94]. They applied their approach to the requirements for an open-source electronic health record system that must comply with the HIPAA and
demonstrated improved compliance via improved security and privacy requirements for the system [94]. Maxwell and Antón developed a production rule model of the HIPAA Privacy Rule that enables analysts to query whether their requirements are compliant with the regulations [95]. It facilitates communication between requirements engineers and legal domain experts [95].

Breaux et al. performed a comparative evaluation between legal and product requirements based on Section 508 of the Rehabilitation Act Amendments of 1998, called the Accessibility Standards [34]. They compared the two requirements sets with respect to qualitative statement and phrase metrics. Statement metrics determine whether: requirements are equivalent; one requirement describes why the other should be implemented; or one requirement describes how the other should be implemented. Phrase metrics determine whether: concepts are generalized; concepts are refined; or the modality changes. Their study demonstrated limitations in existing requirement acquisitions methods. Their approach identified compliance gaps between previously specified product requirements and the Accessibility Standards. They also identified additional knowledge sources to help refine legal requirements into product requirements that comply with law.

Rather than checking the legal compliance of existing requirements or comparing two sets of requirements, within our comparison study (Chapter 5), we compare the results of two requirements engineering approaches—both of which extract requirements artifacts (CPRs and goals) from policy documents—for their ability to ensure compliance with the associated policy documents.

1.2.5 Natural Language

Berry et al. examined ambiguities in natural language software requirements specifications and legal contracts that arise from the use of ambiguous words, such as all, each, every, and, and or [28, 29]. They discussed how to address these types of ambiguities in natural language and how to avoid them. For example, or can be ambiguous whether it refers to the inclusive OR or exclusive OR; to avoid this ambiguity, they suggest explicitly stating either inclusive OR or exclusive OR. As requirements engineers extract commitments, privileges, and rights from natural language documents, they should be aware of ambiguities that may exist and know how to handle the ambiguities.

Goldin and Berry’s AbstFinder tool can be employed to identify abstractions in natural language documents [79]. The abstractions express a document’s main ideas by ignoring the details during requirements elicitation for negotiation with customer. AbstFinder allows phrases to contain any number of words that can be spaced throughout the sentence in any order without concern for semantics. We are concerned with the specificity of the documents that we analyze.

using CPR analysis. Our CPR analysis heuristics employ natural language patterns or phrases. These phrases differ from the phrases used within AbstFinder because the word ordering and context are important for compliance requirements.

Abbott transformed natural language descriptions into computer program by identifying patterns and parts of speech [1]. He linked descriptive noun phrases to a program’s data types and objects. Similarly, our CPR heuristics (Chapter 3) help to identify attributes by asking questions that often relate to parts of speech.

1.2.6 Commitments

Singh employs social commitments between agents\(^6\) for multi-party agreements [120]. He defines a commitment as follows [120]:

\[
A \text{ commitment is a four-place relation involving a proposition } (p) \text{ and three agents } (x, y, \text{ and } G). \text{ Let } c = C(x, y, G, p) \text{ denote a commitment from } x \text{ toward } y \text{ in the context of } G \text{ and for the proposition } p. \text{ Then, } x \text{ is the debtor, } y \text{ the creditor, } G \text{ the context group, and } p \text{ the discharge condition of commitment } c.
\]

In contrast, our multi-party agreements exist between the organization, expressing its practices in policies, and the user, interacting with the organization’s policy-governed system, or between the data custodian and recipient for data use agreements. We specify the commitments, privileges, and rights in terms of actions instead of conditions because as requirements engineers, we want to know the actions that parties can or will do within the system. We discuss Singh’s definitions more in Section 2.1.3, where we discuss the justification for our terminology definitions.

Haddadi examined commitments as they relate to agents in order to form a “potential cooperation,” where agents work together to achieve a goal [84]. Haddadi discussed formation conditions or the conditions under which the commitment is formed. In our work, formation conditions relate to the rationale of the classifiable item.

1.2.7 User Studies

Antón conducted an empirical study in which she compared the GBRAM to alternative analysis methods—Object Modeling Technique (OMT) and non-method-assisted (control) [18]. Her hypothesis was that analysts performing GBRAM are better able to identify requirements than analysts using the alternative analysis methods. Antón examined the total number of requirements and total number of critical requirements (requirements that if not met cause the system to fail) produced by the subjects. GBRAM subjects produced significantly more critical

\(^6\)“Agents are active, persistent (software) components that perceive, reason, act, and communicate” [89].
requirements than the OMT and control subjects. The user study discussed in Chapter 6 is structured in a similar way.

Breaux completed a user study to examine the effect of his upper ontology and phrase heuristics in legal requirements acquisition by measuring completeness and consistency [30]. The requirements produced by the participants were compared against a set of expected legal requirements to determine the number of expected legal requirements produced by the participants using a set of qualitative metrics—statement and phrase. Breaux found that participants using the ontology performed better in terms of completeness and consistency than participants performing traditional practices. In our user study, we compare the subject-produced requirements against a set of expected compliance requirements (Chapter 6).

1.3 Overview of Remaining Chapters

The remainder of this dissertation is organized as follows: Chapter 2 presents our theory of commitments, privileges, and rights along with our CPR analysis methodology; Chapter 3 presents our CPR analysis heuristics; Chapter 4 presents our multi-case studies, which we conducted to develop and validate CPR analysis theory and methodology; Chapter 5 discusses our comparison study in which we compared CPR analysis to goal-based analysis based on ensuring compliance; Chapter 6 describes our empirical user study that we completed to examine the ability of requirements engineers to extract compliance requirements from policy documents; Chapter 7 presents our work with the development of a health system prototype to manage health data from a number of sources to support analytical decision-making; and Chapter 8 summarizes our work, contributions, and future work.
Chapter 2

CPR Analysis Theory and Methodology

I meant what I said
And I said what I meant...
An elephant’s faithful
One hundred per cent!
- Dr. Seuss, Horton Hatches the Egg

The objective of our work is to develop an approach to analyze policy documents and data use agreements (DUAs) for requirements. Initially we attempted to extract rights and obligations from policy documents based on Breaux et al.’s legal-based work with legal texts, where a right is “an action that a stakeholder is conditionally permitted to perform” and an obligation is “an action that a stakeholder is conditionally required to perform” [36]. During this initial analysis, we discovered that Breaux et al.’s classifications—rights and obligations—were insufficient to characterize the types of items contained within policy documents which are important for requirements. Many of the items within the policy documents are actually pledges, or commitments from the organization to the user, rather than rights and obligations as found within legal texts. For example, many statements contained the phrase “we will,” which is used to express a commitment to a position or action [38].

When we discovered that Breaux et al.’s legal-based approach was insufficient for policy documents, we restarted our analysis using a grounded theory approach to better understand and comprehensively characterize the kinds of items contained in the policy documents. In grounded theory, a data set is systematically analyzed to obtain theory [75, 76]. Therefore, we did not use a distinct preconceived theory or hypothesis as the basis of our work. Instead, we
developed new theory based on scientific analysis. Requirements engineering researchers have successfully relied on grounded theory approaches in their research of legal- and policy-based software requirements [20, 33, 35, 36, 64].

By applying a grounded theory approach to analyze policy documents and DUAs, we developed our theory of commitments, privileges, and rights (the classifications discovered through analyzing these documents) and our methodology (how to extract CPRs from policy documents [132, 134] and DUAs [119] and operationalize them as requirements). In our work, commitments reflect an actor’s pledges, whereas privileges and rights reflect actions that an actor is entitled to perform. In Section 2.1.1, we provide full definitions and discuss how commitments, privileges, and rights differ from one another.

When developing our theory and methodology, we completed two types of multi-case studies—formative and summative, which we discuss in Chapter 4. The formative multi-case study was conducted to form ideas and gain context. The summative multi-case study was conducted to validate the theory and methodology. The cases within the studies relied on two types of documents—policy documents and DUAs. During our formative study, we examined each available document three times. First, we examined the documents to understand the type of information they contained—pledges and entitlements. This analysis yielded classifications—commitments, privileges, and rights. During the second pass through the documents, we examined the natural language phrases that signaled each classification. This list of natural language phrases formed the basis for our classifying heuristics discussed in Chapter 3 [119, 132]. During the final pass through the documents, we applied our heuristics to confirm that each item could be classified using the heuristics.

For the remainder of this chapter we discuss our theory of CPRs and the CPR analysis methodology.

2.1 Theory of Commitments, Privileges, and Rights

Our theory contains three classifications: commitment, privilege, and right, which are defined in Section 2.1.1. Commitments are pledges; privileges and rights are entitlements. Each CPR has a rationale—internal or external, which are defined in Section 2.1.2. In Section 2.1.3, we provide a comparison of CPR terminology and definitions with terminology and definitions from other researchers in the following areas: legal concepts, legal-based requirements engineering, and multi-agent agreements.

2.1.1 Classification Terminology

Our analysis of policy documents and DUAs for information of importance for requirements and compliances yielded *pledges*—commitments—and *entitlements*—privileges and rights:
A commitment of Party A with respect to Party B is an action, a, that Party A pledges to Party B.

Consider the following item from a health system prototype DUA:

Moreover, all access to the system and its data will be stored in a secure log file for auditing.

The item contains a commitment:

store all access to the system and its data in a secure log file for auditing

A privilege of Party A with respect to Party B is an action, a, that Party A is entitled to perform, where a is an independent action. An independent action is one which has no complementary action from another party (Party B) that must be completed by Party B in order for the action to take place. In other words, Party A has the privilege of performing a only if performing a does not imply that a commitment from Party B to Party A must exist.

Consider the following item from Pfizer Privacy [111]:

Generally, you can browse our Web sites without disclosing any personally identifiable information.

The item contains a privilege:

browse the organization’s websites without disclosing any personally identifiable information

The item is a privilege rather than a right because the action browse is an independent action that does not have a complementary action that must be completed by another party.

A right of Party A with respect to Party B is an action, a, that Party A is entitled to perform, where a is a dependent action. A dependent action, a, is one which has a complementary action, b, from another party that must take place in order for a to take place. In other words, Party A has a right only if performing a implies that a commitment from Party B to Party A to perform b exists.

Consider the following item from Aetna Privacy Notices [12]:

---

1 The item is classified as a commitment using CH_{DU/A} on page 39.
2 Others researchers have also used complementary actions [25, 102, 103].
3 The item is classified as a privilege based on CH_{PD} 1 and CH 1 on pages 30 and 29.

13
The Health Insurance Portability and Accountability Act of 1996 (HIPAA) Privacy Rule allows members the right to receive a notice that describes how individual health information may be used and/or disclosed and how to acquire access to this information.

The item contains a right from legal text and will have external rationale:

receive notice in order to describe how individual health information may be used and/or disclosed and how to acquire access to this information

The complementary action and commitment of the organization is the following because in order for the user to receive notice, the organization must send notice:

send notice to user in order to describe how individual health information may be used and/or disclosed and how to acquire access to this information

In software systems, it is important to distinguish between privileges and rights. If Party A has a right, then the system not only needs to incorporate the action, \( a \), of Party A’s right but also the action, \( b \), of Party B’s implied commitment that exists because \( a \) is a dependent action. The inclusion of \( b \) within the system is important because without it, Party A will not be able to perform \( a \). For example, receive notice is a right rather than a privilege because if the user is entitled to receive a notice, it is dependent on the organization sending notice. Note that the action of the implied commitment, send, is the complementary action of receive, the action of the right. Given a privilege, the system only needs to incorporate the action, \( a \), of the privilege without concern for actions of another party because \( a \) is independent and does not require the action of another party. An example privilege of the organization is change the terms of policy. This is a privilege because it expresses an action, changing, that the organization is entitled to perform, but that does not imply an action on the user, or other actor, because the action is independent.

CPR analysis provides requirements engineers with an understanding of the requirements contained within policy documents and DUAs by focusing on pledges and entitlements, which are important because the pledged and entitled actions should be required and allowed, respectively, by the organization and within the system.

### 2.1.2 Rationale Terminology

Each classifiable item has a rationale, or basis for the classifiable item being included within the document; specifically:

\[ \text{The item is classified as a right based on CHPD 6 and CH 1 on pages 33 and 29.} \]
• A classifiable item has an *internal* rationale, if it delineates organizational practices or procedures as the basis for the item.

• A classifiable item has an *external* rationale, if it delineates legal texts as the basis for the item.

### 2.1.3 Terminology Comparison

We now compare our terminology and definitions to those of legal concepts, legal-based requirements engineering, and multi-agent agreements. Hohfeld, who was a law professor early in the twentieth century, proposed eight fundamental legal concepts that describe legal relations—*duty, right, privilege, no-right, power, liability, immunity*, and *disability* [87]; however, herein, we focus on the three concepts that are relevant to our analysis of policy documents and DUAs—duty, right, and privilege. These are relevant to our commitment, right, and privilege, respectively. Breaux et al. presented a legal-based approach to analyzing legal texts for software requirements [30, 36]. As previously discussed, originally we attempted to apply Breaux et al.’s legal-based approach [36] to policy documents but discovered that their classifications were insufficient to characterize the kinds of items within policy documents, which are inherently different from the kinds of items found in legal documents. Singh discusses agreements within multi-agent systems, employing social commitments, privileges, and rights between agents [120].

**Commitment**

First, we examine the terminology and definitions of legal concepts, legal-based requirements engineering, and multi-agent agreements that are relevant to our commitment classification.

Hohfeld’s duty and Breaux’s obligation legally require the actor to perform the action. Hohfeld defines a duty as follows [87]:

> A *duty* or a legal obligation is that which one ought or ought not to do.

Breaux defines an obligation as follows [30]:

> *Obligation* – an act that an actor is required to perform.

As we previously discussed, CPR analysis was developed because policy documents contain actions that are pledged rather than legally required. In our work, commitments are pledges that either have an internal or external rationale. The rationale is used to track whether the commitment is based on organizational practices or legal texts. For this reason, most directly related to Hohfeld’s duty and Breaux’s obligation are our commitments with external rationale as they are legally pledged or required. Our commitments with internal rationale are not required until pledged or included within the policy document or DUA.
Singh employs social commitments between agents within multi-agent systems [120]. Singh defines a commitment as follows [120]:

A commitment is a four-place relation involving a proposition \( p \) and three agents \((x, y, \text{ and } G)\). Let \( c = C(x, y, G, p) \) denote a commitment from \( x \) toward \( y \) in the context of \( G \) and for the proposition \( p \). Then, \( x \) is the debtor, \( y \) the creditor, \( G \) the context group, and \( p \) the discharge condition of commitment \( c \).

Singh’s commitment contains two parties—debtor and creditor. “[T]he debtor is the agent who is committed, and the creditor is the agent who receives the commitment” [120]. In CPR analysis, we do not use ‘debtor’ and ‘creditor’ because while the terms can have meanings that are outside of the financial domain, we use Party A and Party B in order to more generally note the parties. Party A and Party B within our work relate to Singh’s debtor and creditor, respectively. Hohfeld and Breaux do not name the parties associated with duties and obligations.

Like our definitions, Hohfeld’s and Breaux’s definitions are stated in terms of actions that take place. Singh’s definitions are stated in terms of conditions. We specify commitments, privileges, and rights in terms of actions instead of conditions, because as requirements engineers, we want to know the actions that parties can or will do within the system [30, 36, 96, 97, 121]. Considering the condition that holds exactly when a given action has been completed, one should be able to switch between the use of conditions and actions.

Singh discusses operations on commitments—create, discharge, cancel, release, delegate, and assign [120]. Given our concern with compliance, it is important to note that if the policy document or DUA changes, the commitments also change. Similarly, within our work most of the actions the actors will complete multiple times, not just once; therefore we do not have discharge conditions. For these reasons, we do not perform operations on our commitments.

**Right**

We examine the terminology and definitions of legal concepts, legal-based requirements engineering, and multi-agent agreements that compare to our right classification. Like Breaux’s permission, our right is an action that the actor is entitled or permitted to perform. Breaux defines permission as follows [30]:

*Permission* – an act that an actor is permitted to perform.

Our right is a dependent action, \( a \), that a party is entitled to perform that has a complementary action from another party that must be completed; this complementary action relates to a commitment from another party. Hohfeld’s right\(^5\) and Singh’s claim place a claim against

\(^5\)It should be noted that there could be other ways to model these relationships, e.g., our right as Hohfeld’s power (Munindar P. Singh, personal communication, June 13, 2012).
another party; in other words, they place a duty on another party. Therefore, they are similar to our right in that they are actions that place actions on another party. Hohfeld defines right as follows [87]:

A right is one’s affirmative claim against another

Singh defines claim as follows [120]:

A claim or right is what an agent can demand from another. It is like a commitment with respect to the relevant context, which is not made explicit by Hohfeld. Thus, we have Claim(x, y, p) \overset{\Delta}{=} C(y, x, G, p).

Hohfeld’s right and duty are correlatives, which means the legal relations are equivalent. For example, “if X has a right against Y that he shall stay off the former’s land, the correlative (and equivalent) is that Y is under a duty toward X to stay off the place” [87]. A right, by our definition, does have an implied commitment for another party based on the dependent action. However, commitments do not always provide others with rights; instead, they can simply say what Party A will do without entitling Party B to do anything. For example, the organization’s commitment to comply with law does not provide the user with a right. Breaux and Antón discussed the need for implied commitments with rights and the unnecessary implied rights as [32]:

Rights without complementary obligations are meaningless since governed parties are not required to respond to the invocation of such rights. In terms of designing and engineering software systems, these rights may be effectively ignored. On the other hand, obligations without an explicit and complementary right do have value and must be properly incorporated into system specifications.

Based on Breaux and Antón’s discussion of implied rights and because commitments do not always provide others with rights, we do not document implied rights within our work.

Privilege

We examine the terminology and definitions of legal concepts, legal-based requirements engineering, and multi-agent agreements with respect to our privilege classification.

Our privilege is an independent action, a, that a party is entitled to perform; an independent action is one which has no complementary action from another party (Party B) that must be completed by Party B in order for the action to take place. In other words, Party B need not perform an action, b, in order for Party A to perform a. Hohfeld’s privilege and Singh’s privilege both address the freedom from claims of another, meaning there is no commitment from Party A to Party B. Hohfeld defines privilege as follows [87]:

17
A *privilege* is one’s freedom from the right or claim of another

Singh defines privilege as follows [120]:

A *privilege* is a freedom an agent has from claims of another. In other words, it is an absence of a duty to refrain from the given act. In this sense, a privilege is the dual of a claim with the roles of the agents reversed, i.e., \( \text{Priv}(x, y, p) \triangleq \neg \text{Claim}(y, x, \neg p) \).

Although our privileges are classified using a different definition than Hohfeld’s and Singh’s, our privileges would satisfy their definitions. We can show this by contradiction. To do this, we assume that there exists an action \( a \) that *Party A* has the privilege to perform and *Party A* has the commitment not to perform \( a \). It is straightforward to see how this would be a contradiction of both Hohfeld’s and Singh’s definition of privilege. For this situation to be possible according to our definitions, one of two cases must exist: either *Party A* has committed to *Party B* not to perform \( a \), or *Party B* has the right to perform the complementary action of not performing \( a \). The first case would be a conflict in the policy document or DUA because the document would state that *Party A* would have the privilege of performing \( a \) and the commitment not to perform \( a \). In our studies we have not found an example of such a conflict, and thus can assume that such conflicts do not exist. The second case also cannot exist because it requires a complementary action for \( a \), which by our definition of privilege cannot exist. Thus, according to our definition of privilege, if *Party A* has the privilege to perform \( a \), *Party A* will not have the commitment not to perform \( a \), which agrees with Hohfeld’s and Singh’s definition of privilege.

Breaux defines exclusion as follows [30]:

*Exclusion* – an act that an actor has no express permission to perform or that an actor is not expressly required or prohibited from performing.

Breaux also states that an exclusion is “Any action that an actor is not permitted, required, or prohibited from performing” [30]. In contrast, our privilege is an independent action that the actor is entitled, or permitted, to perform.

**Summary of Comparison**

We have shown how our terminology and definitions differ from those of others. While there are some similarities between our terminology and others’, we were not able to simply adopt existing terminology because they did not hold true for our data set—policy documents and DUAs. Given that our theory and methodology are the result of grounded theory analysis, we must maintain validity within the analyzed documents. Our classifications must hold true for our data set. The specificity and uniqueness of our definitions of CPR is likely influenced by the difference in objective and source of the different existing approaches.
Our commitments could be considered equivalent to Singh’s. Our commitments with external rationale match Hohfeld’s duty and Breau’s obligation. Our privileges and rights would also be classified as privileges and rights, respectively, with Hohfeld’s and Singh’s definitions; however, all of Hohfeld’s and Singh’s privileges and rights would not be classified as privileges and rights, respectively, with our definitions.

2.2 CPR Analysis Methodology

CPR analysis employs a document as the input and produces a set of requirements as the output. The methodology consists of three steps (see Figure 2.1): (1) parse; (2) classify; and (3) operationalize.

2.2.1 Parse

The requirements engineer parses the source document into individual statements so that each statement may be independently analyzed. The documents are generally formatted as paragraphs and/or lists.

When considering paragraphs, the requirements engineer parses each sentence as a separate statement. To parse the documents based on sentences, we apply Manning and Schütze’s algorithm to identify the sentence boundaries [92]. We tentatively place sentence boundaries after all periods and question marks; boundaries are disqualified if the period follows an abbreviation or if the period or question mark is part of a URL or email address [92]. Each section heading is also parsed as a separate statement.

A continuation is a statement that begins in the fragment before a list and ends within a list item [30]. Herein, the fragment before a list is prepended to the beginning of each list item, forming statements with the corresponding sentences.

2.2.2 Classify

The classify step contains multiple substeps within it. First, the requirements engineer distinguishes between classifiable and unclassifiable items. Second, the requirements engineer classifies classifiable items as commitments, privileges, and rights. Third, the requirements engineer documents attributes for each classifiable item. Fourth, the requirements engineer documents the rationale of each classifiable item. Finally, the requirements engineer forms the commitments, privileges, and rights.

The requirements engineer examines each statement to determine whether it contains any classifiable items—i.e., commitments, privileges, or rights—which are important for requirements and compliance. Some statements may not contain classifiable items, or are not important for
Figure 2.1: CPR Analysis Methodology
requirements or compliance—for example, if the statements are section headings, definitions, clarifications, or warnings; we refer to these as unclassifiable items and do not operationalize them as requirements. Other statements may contain multiple CPRs—if they contain multiple actions, conditions, and/or purposes.

The types of items within a document can be seen in Figure 2.2. Each item within the document is either classifiable or unclassifiable. A classifiable item is either a pledge or an entitlement. A pledge is a commitment, whereas an entitlement is either a privilege or a right. Because our theory and methodology were developed using a grounded theory approach, we know that all items that are important for requirements and compliance are classifiable—commitments, privileges, and rights; all other items are unclassifiable as they are not important for requirements and compliance.

By classifying the items, the requirements engineer can gain a greater understanding of the types of items that are expressed within policy documents and DUAs. The requirements engineer classifies each classifiable item as either a commitment, a privilege, or a right through the use of natural language patterns and heuristics, Classifying Heuristics (CH) given in Section 3.2.3. The requirements engineer also documents the attributes and rationale of each classifiable item using natural language patterns and heuristics, Documenting Attribute Heuristics (DAH) and Documenting Rationale Heuristics (DRH) given in Sections 3.2.4 and 3.2.5.

![Figure 2.2: Types of Items within a Document](image)

Each classifiable item is further parsed to identify attributes, which the requirements engineer documents in order to form CPR and operationalize the CPR as a requirement. These attributes are defined as follows:
The **actor** is the responsible stakeholder who performs the given action.

The **action** is the action (verb) that the actor performs.

The **object** is the item upon which the actor’s action is acting.

The **object’s source** is the originator of the object.

The **target** is the intended recipient of the actor’s action.

The **purpose** is the reason for performing the action.

The **condition** is the restriction or pre-condition on performing the action.

The **examples**, or scenarios, illustrate how commitments, privileges, and rights are executed.

Values for these attributes are found by asking questions that relate to the definition of each attribute [114]. It is possible that some attributes may not have values for a given item. When an item does not include certain attributes or there are values that may be inferred, the item should be discussed with the organization to find out what is meant by—and ideally to clarify—the policy document or DUAs.

To form the CPRs, we use the CPR template (FH 1 on page 49) that incorporates the identified attributes, which we documented. If an attribute does not have a value, the related portion is omitted within the template. All of the attributes are used within the templates with the exception of examples. Software engineers gain a stronger understanding of the requirements by using these examples as instances of when and/or how the CPR expressed by the requirement is used within the system.

### 2.2.3 Operationalize

The requirements engineer operationalizes the CPRs into requirements using templates. *Operationalize* refers to translating classifiable items into requirements. The unclassifiable items are not operationalized into requirements. Others have used templates when operationalizing items into requirements [18]. This final step takes classifiable items as input and produces a set of requirements.

CPR analysis produces two types of requirements—system and operational. **System requirements** specify the capabilities of the system or software. **Operational requirements** specify “business rules or operational procedures” [20] outside of the system. We employ templates for each type of requirement based on whether the item is a pledge or an entitlement.
2.3 Summary

This chapter presented the CPR analysis theory and methodology. We discussed our theory that contains three classifications: commitment, privilege, and right. We compared these classifications to the concepts of other researchers in the following areas: legal concepts, legal-based requirements engineering, and multi-agent agreements. We also described the steps of CPR analysis methodology. In the next chapter, we discuss the heuristics that are employed to perform each of these steps.
Chapter 3

CPR Analysis Heuristics

Penny: You’ll never guess what just happened!
Leonard: Oh, I give up!
Sheldon: I don’t guess! As a scientist, I reach conclusions based on observations and experimentation.
- The Big Bang Theory

In this chapter, we discuss the heuristics employed to perform CPR analysis. The heuristics are rules that guide requirements engineers conducting CPR analysis. We organize this chapter into sections based upon the three steps of CPR analysis—(1) parse; (2) classify; and (3) operationalize. Each section discusses the heuristics employed for that step. Where separate heuristics exist for policy documents and data use agreements (DUAs), we note these heuristics with subscript $PD$ and $DUA$, respectively.

3.1 Parse

In order to parse the document into statements that can be analyzed independently, we employ three Parsing Heuristics (PH). These heuristics are used for both policy documents and DUAs.

The first heuristic (PH 1) is applied to paragraphs. It splits a paragraph into sentences based on Manning and Schütze’s algorithm [92]. We consider each sentence to be a statement.
Given a paragraph of the document, follow the following steps to parse it into sentences:

1. Tentatively place sentence boundaries after all periods and question marks.
2. Disqualify boundaries if the period follows an abbreviation.
3. Disqualify boundaries if the period or question mark is part of a URL or email address.

The result will be a set of statements based on sentences.

**PH 1: Parse Paragraph**

Consider the following paragraph from Dossia’s Privacy Statement [60]:

Dossia collects the personal information you voluntarily enter into the Dossia website, including the health information you enter, or authorize others to enter, into your Personally-Controlled Health Record (PCHR). Dossia protects the privacy and security of this information as described in this Privacy Statement and uses this information to provide you with your PCHR and associated services. Except for any narrow exceptions explained in this Privacy Statement, Dossia will not disclose information in PCHRs to third parties without your explicit permission.

Using PH 1 the paragraph is parsed into three statements based on the three sentences.

1. Dossia collects the personal information you voluntarily enter into the Dossia website, including the health information you enter, or authorize others to enter, into your Personally-Controlled Health Record (PCHR).
2. Dossia protects the privacy and security of this information as described in this Privacy Statement and uses this information to provide you with your PCHR and associated services.
3. Except for any narrow exceptions explained in this Privacy Statement, Dossia will not disclose information in PCHRs to third parties without your explicit permission.

Policy documents and DUAs tend to also contain lists. There are two types of lists within documents: lists that contain continuations and lists that do not contain continuations. A continuation is a statement that begins in the fragment before the list and ends within a list item [30]. PH 2 addresses how lists are parsed into statements for analysis.

- For lists that do not contain continuations, parse each list item using **PH 1**.
- For lists that contain continuations, complete the following steps:
  1. Prepend the fragment before the list to each list item.
  2. Use **PH 1** to parse the resulting list items.

**PH 2: Parse List**

25
Consider the following list from Aflac’s Privacy Policy [14]:

Accordingly, we will disclose Personal Information to employees, agents, or third parties only as described herein and:

1. To fulfill a transaction that you have requested on the Web Sites.
2. To service your policy.
3. To investigate or handle claims.
4. As permitted or required by law by regulatory and law enforcement authorities.
5. To enforce or apply our Terms and Conditions of Use and other agreements.
6. To protect the rights, property, or safety of Aflac, aflac.com, aflacny.com, the Web Sites’ visitors, or others.

Because the list is a continuation, using PH 2, we prepend the fragment before the list (Accordingly, we will disclose Personal Information to employees, agents, or third parties only as described herein and) to each list item. Then we parse each list item using PH 1, resulting in the following six statements:

1. Accordingly, we will disclose Personal Information to employees, agents, or third parties only as described herein and To fulfill a transaction that you have requested on the Web Sites.
2. Accordingly, we will disclose Personal Information to employees, agents, or third parties only as described herein and To service your policy.
3. Accordingly, we will disclose Personal Information to employees, agents, or third parties only as described herein and To investigate or handle claims.
4. Accordingly, we will disclose Personal Information to employees, agents, or third parties only as described herein and As permitted or required by law by regulatory and law enforcement authorities.
5. Accordingly, we will disclose Personal Information to employees, agents, or third parties only as described herein and To enforce or apply our Terms and Conditions of Use and other agreements.
6. Accordingly, we will disclose Personal Information to employees, agents, or third parties only as described herein and To protect the rights, property, or safety of Aflac, aflac.com, aflacny.com, the Web Sites’ visitors, or others.

The final parsing heuristic is for section headings. With PH 3, a section heading is considered its own statement.

Parse each section heading as a separate statement.

**PH 3:** Section Heading
A. Information Collected and Used by Dossia

3.2 Classify

The classify step of CPR analysis contains multiple substeps. First, the requirements engineer distinguishes between classifiable and unclassifiable items by using the Item Heuristic (IH). Second, the requirements engineer classifies each classifiable item as a commitment, privilege, or right by using the Classifying Heuristics (CH). Third, the requirements engineer documents the attributes for each classifiable item by using the Documenting Attribute Heuristics (DAH). Fourth, the requirements engineer documents the rationale for each classifiable item by using the Documenting Rationale Heuristics (DRH). Finally, the requirements engineer forms commitments, privileges, and rights by using the Forming Heuristic (FH). The requirements engineer also uses Helper Heuristics (HH) for tasks within the classify step.

3.2.1 Item Heuristic (IH)

Recall that items within a document are either classifiable or unclassifiable (Figure 2.2 on page 21). The classifiable items are classified as commitments, privileges, and rights, then operationalized as requirements. The unclassifiable items are not commitments, privileges, or rights and are not operationalized as requirements. After an item is documented as unclassifiable, no other heuristics are applied to it. Based on the grounded theory analysis that was used to develop our theory and methodology, we found that each unclassifiable item in policy documents and DUAs was either a clarification, a definition, a section heading, or a warning. We employ IH 1 to determine whether an item is classifiable or unclassifiable based on the types of items in documents that are unclassifiable.

If the item is one of the following:

- **Clarification**: item clarifies or explains another item.
- **Definition**: item defines a key word or phrase.
- **Section heading**: item denotes the start of a section.
- **Warning**: item provides a warning, advisory, or suggestion.

Note that the item is **unclassifiable**. Otherwise, note that the item is **classifiable**.

**IH 1**: Distinguish between Classifiable and Unclassifiable Items

27
The following item from Aetna’s Notice of Privacy Practices - Strategic Resource Company (SRC) [10] is unclassifiable based on IH 1 because it is a definition for *health information*:

By “health information,” we mean information that identifies you and relates to your medical history (i.e., the health care you receive or the amounts paid for that care).

### 3.2.2 Helper Heuristics (HH)

We employ two Helper Heuristics to assist us during the classify step. HH 1 describes what the requirements engineer does when a classifiable item from a policy document or DUA contains multiple actions. If an item contains multiple actions, it is split into multiple items that each express a single action. This heuristic can be applied with any of the Classifying Heuristics (CH) when the item contains multiple actions.

If an item contains multiple actions, split the item into multiple items—each of which contains a single action.

**HH 1: Multiple Actions**

For example, if the item is in the following format: [actor] [action 1] and [action 2] [object], then we separate it into two items: [actor] [action 1] [object] and [actor] [action 2] [object]. As such, each item will only have a single action. Consider the following item from Dossia’s Privacy Statement [60]:

Dossia collects and uses identifiable information about you for enrollment, ongoing account and system administration, communications with you about your account, and internal operations.

The item contains two actions—collect and use. Using HH 1, we split it into two items that each contain a single, testable action [72, 121]:

1. Dossia collects identifiable information about you for enrollment, ongoing account and system administration, communications with you about your account, and internal operations.
2. Dossia uses identifiable information about you for enrollment, ongoing account and system administration, communications with you about your account, and internal operations.

HH 2 relates to the use of pronouns in documents. For CPRs and requirements, we use nouns rather than pronouns to describe the parties and objects in order to be explicit about the party or object being referenced.
When a pronoun is present in an item, change the pronoun to the antecedent that the noun references.

**HH 2: Pronouns**

Policy documents often reference the *organization* with *we, us,* or *our.* Policy documents reference the *user* with *you* or *your.* We change the pronouns referencing the organization and the user to *organization* and *user,* respectively. In DUAs, pronouns were not as prominent as pronouns in policy documents.

### 3.2.3 Classifying Heuristics (CH)

We now discuss classifying heuristics that we employ to determine the classification of an item. These heuristics classify items based on the presence of natural languages phrases. While policy documents and DUAs contain some of the same phrases, the classification for the phrases are not identical. For this reason, each document type has its own set of heuristics.

There is one classifying heuristic that is the exception to the classifying heuristics being document specific. CH 1 applies to both policy documents and DUAs. As there are two types of entitlements—privilege and right, CH 1 is a helper classifying heuristic, which is used to distinguish between entitlements that are privileges and entitlements that are rights. This distinction is based on the differences in definitions discussed in Section 2.1.1. CH 1 is not used alone but instead is used by other classifying heuristics (CH\textsubscript{PD} 1, 2, 3, 6, and 9 & CH\textsubscript{DUA} 1, 3, and 6) to determine whether an entitlement is a privilege or a right.

- If the entitlement contains an independent action (no complementary action exists), classify it as a **privilege**.
- If the entitlement contains a dependent action (a complementary action from another party that must take place exists), classify it as a **right**.

**CH 1: Distinguish between Entitlements—Privileges and Rights**

### Classifying Heuristics for Policy Documents (CH\textsubscript{PD})

We now examine classifying heuristics for policy documents (CH\textsubscript{PD}). The presence of modal verbs, which express modality—necessity or possibility, can be used to determine an item’s classification. CH\textsubscript{PD} 1 classifies an item based on modal verbs—*can, could, do, may, might, must, shall, should, will,* or *would*—in the pattern [actor] [modal verb] [action]. In some cases,
CH 1 is needed to determine the type of entitlement within the item.

If the item is in the format [actor] [modal verb] [action], classify the item based on the modal verb and possibly the actor.

- If the item contains the modal verb must, classify it as a commitment because the item expresses necessity through a pledge.
- If the item contains the modal verb can, could, may, or might, then it expresses a possibility through an action that the actor is entitled to perform. Classify the entitlement as a privilege or a right using CH 1.
- If the item contains the modal verb do, shall, should, will, or would, then it is actor-dependent.
  - If the organization is the actor, classify the item as a commitment because the item expresses what the organization pledges to the user.
  - If the user is the actor, the item is an entitlement. Classify the entitlement as a privilege or a right using CH 1.

\[
\text{CH}_{PD} 1: \text{Modal Verb}
\]

Consider the following item from Cigna’s Notice of Privacy Practices - CIGNATURE Rx [47]:

If we change this Notice, we will send you the new notice if you are enrolled in a CIGNA HealthCare benefit plan at that time.

The actor organization and the modal verb will signal a necessity—sending notice. Given that the modality is a necessity, we classify the item as a commitment based on \( \text{CH}_{PD} 1 \). In this example, the action is send.

While modal verbs are an identifier of an item’s modality, they are not present in every item. For items without modal verbs, we use other heuristics to determine the classification of the item. \( \text{CH}_{PD} 2 \) classifies the item based on the actor when the item is in the format of [actor] [action].
If the item is in format [actor] [action], classify it based on the actor.

- If the organization is the actor, classify the item as a commitment because it expresses an action the organization pledges to the user.
- If the user is the actor, then the item expresses an action the user is entitled to perform and is an entitlement. Classify the entitlement as a privilege or a right using CH 1.

**CH\textsubscript{PD} 2:** Actor-Action Phrase

Consider the following item from Dossia’s Privacy Statement [60]:

Dossia collects the personal information you voluntarily enter into the Dossia website, including the health information you enter, or authorize others to enter, into your Personally-Controlled Health Record (PCHR).

Based on CH\textsubscript{PD} 2, the item is a commitment because the organization is the actor and collect is the action. The commitment expresses what the organization pledges to do—collect personal information.

Key phrases signal privileges and rights. CH\textsubscript{PD} 3 includes phrases that signal entitlements through the use of able to, choose to, authority to, and ability to.

If the item is in one of the following formats:

- [actor] is/are able to [action]
- [actor] may be able to [action]
- [actor] should be able to [action]
- [actor] will be able to [action]
- [actor] is/are entitled to [action]
- [actor] can choose to [action]
- [actor] may choose to [action]
- [actor] will have the authority to [action]
- [actor] has/have the ability to [action]

The item is an entitlement. Classify the entitlement as a privilege or a right using CH 1.

**CH\textsubscript{PD} 3:** Key Phrases for Privileges and Rights

Consider the following item from Dossia’s Privacy Statement [60]:

31
If you choose to receive more information about the enhancements, products, or services, you will be able to request it.

Based on CH_PD 3, the item is an entitlement because it is in the format \( \text{[actor]} \text{ will be able to [action].} \) The \emph{user} is the actor, and \emph{request} is the action. The entitlement is a privilege based on CH 1 because the action \emph{request} is independent—no complementary action.

Policy documents can explicitly state actions that the actor is required to do. CH_PD 4 includes phrases that signal commitments through the use of the word \emph{require}.

If the item is in one of the following formats:
- \( \text{[actor]} \text{ is/are required by law to [action]} \)
- \( \text{[actor]} \text{ is/are required to [action]} \)
- \( \text{[legal text]} \text{ requires [actor] to [action]} \)
- \( \text{[legal texts]} \text{ require [actor] to [action]} \)
- \( \text{[system]} \text{ requires [actor] to [action]} \)
- \( \text{[system]} \text{ will require [actor] to [action]} \)

Classify the item as a \textit{commitment}.

\textbf{CH_PD 4: Explicitly Required}

Consider the following item from CIGNA International Expatriate Benefits [41]:

\emph{We are required by law to maintain the confidentiality of health information that identifies you.}

According to CH_PD 4 and the phrase \( \text{[actor]} \text{ are required by law to [action]} \), the item is a commitment that is based on law. The \emph{organization} is the actor, and \emph{maintain} is the action.

Policy documents can also explicitly state the actions that an actor commits to performing. CH_PD 5 includes a phrase that signals commitments through the use of \emph{committed}.

If the item is in the following format:

\( \text{[actor]} \text{ is/are committed to [present participle of action]} \)

Classify the item as a \textit{commitment}.

\textbf{CH_PD 5: Explicitly Committed}

Consider the following item from GlaxoSmithKline (GSK) internet privacy statement [77] that
signals a commitment based on CH$_{PD}$ 5 and the phrase [actor] is committed to [present participle of action]:

GSK is committed to protecting the security of your personal information.

The organization is the actor, and protect is the action.

Policy documents also contain explicit privileges and rights. CH$_{PD}$ 6 includes phrases that signal entitlements through the use of the word right. It is important to note that although the items may explicitly state “right,” these items are not always truly rights based on our definitions. The first six phrases of CH$_{PD}$ 6 express entitlements from legal text and will have external rationale, which will be documented using DRH$_{PD}$ 1.

If the item is in one of the following formats:

- [legal text] affords [actor] the right to [action]
- [legal texts] afford [actor] the right to [action]
- [legal text] allows [actor] the right to [action]
- [legal texts] allow [actor] the right to [action]
- [legal text] gives [actor] the right
- [legal texts] give [actor] the right
- [actor] has/have the right to [action]
- [actor] reserve/s the right to [action]

The item is an entitlement. Classify the entitlement as a privilege or a right using CH 1.

CH$_{PD}$ 6: Explicit Privileges and Rights

Consider the following item from Aetna’s Privacy Notices [12]:

The Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule affords members the right to receive a notice that describes how health information may be used and disclosed and how to get access to this information.

The legal text is Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, the actor is the user, and the action is receive. Through the use of CH$_{PD}$ 6 along with CH 1, the item is a right rather than a privilege because receive is a dependent action as the organization must send notice in order for the user to receive the notice.

Items within policy documents also express actions and objects to which the actor agrees or consents. We classify these items as commitments. CH$_{PD}$ 7 and CH$_{PD}$ 8 address the agreement or consent to actions or objects.
If the item is in one of the following formats:

- [actor] agree to [action]
- [actor] consent to [action]
- [actor] has/have agreed to [action]

Classify the item as a commitment.

**CH\textsubscript{PD} 7: Agree/Consent to Action**

Consider the following item from Microsoft Online Privacy Statement [99]:

To demonstrate our commitment to your privacy, we have agreed to disclose our information practices.

Based on CH\textsubscript{PD} 7, we classify the item as a commitment, where the organization is the actor and disclose is the action.

If the item is in one of the following formats:

- [actor] agree to [object]
- [actor] consent to [object]

Classify the item as a commitment. The commitment has agree or consent as the action.

**CH\textsubscript{PD} 8: Agree/Consent to Object**

Consider the following item from DestinationRx’s Privacy Policy [57]:

By using our Web site, you consent to the collection and use of the information as outlined above by DestinationRx.

Based on CH\textsubscript{PD} 8, it is a commitment, where the user is the actor, consent is the action, and collection and use of the information is the object.

Policy documents can contain items that express privileges and rights of the user that are directed at the user through the use of imperative sentence structure. CH\textsubscript{PD} 9 addresses the use of imperative sentences in policy documents.
Consider the following item from Cigna’s Notice of Privacy Practices - CIGNA Medical Group of Arizona [45]:

If you wish to make any of the requests listed above under “Individual Rights,” please make the request at any of our health care centers.

Based on CH_PD 9, the item is an entitlement. Using CH 1, we classify the entitlement as a privilege because request is an independent action. The condition is user wishes to make any of the requests listed under "Individual Rights" in policy document, and the action is make.

Policy documents can express commitments of the organization through items that have the object as the subject of the sentence, as shown in CH_PD 10.
Consider the following item from Pfizer’s Privacy Policy [112]:

The backup media is stored in a physically secure storage facility.

Using CHPD 10, we classify the item as a commitment of the organization based on the format [object] is [past participle of action]. The action is store, and the object is backup media.

Classifying Heuristics for DUAs (CHDUA)

We now discuss classifying heuristics for DUAs (CHDUA). As with policy documents, modal verbs can be used to determine an item’s classification within DUAs (CHDUA 1).

If the item is in the format [actor] [modal verb] [action], classify the item based on the modal verb.

- If the item contains the modal verb must, shall, or will, classify it as a commitment because the item expresses necessity through a pledge.
- If the item contains the modal verb may, then it expresses a possibility through an action that the actor is entitled to perform. Classify the entitlement as a privilege or a right using CH 1.

CHDUA 1: Modal Verbs

Consider the following item from North Carolina Center for Health Statistics DUA [105]:

Recipient shall report to Covered Component any use or disclosure of the IHI that is not provided for in this DUA of which the Recipient becomes aware.

Based on CHDUA 1 and the modal verb shall, we classify the item as a commitment. The recipient is the actor, and report is the action.

CHDUA 2 classifies the item based the format [actor] [action], which expresses a commitment within DUAs.

If the item is in format of [actor] [action], classify the item as a commitment.

CHDUA 2: Actor-Action Verb Phrase

Consider the following item from a health system prototype DUA:

36
The data custodian maintains the Data System under the authority of the Health and Human Services, Division of Health Services Regulation, Office of Emergency Medical Services.

Based on CH\textit{DUA} 2, we classify the item as a commitment. The \textit{data custodian} is the actor, and \textit{maintain} is the action.

Some items in DUAs contain explicit privileges and rights. CH\textit{DUA} 3 gives natural language patterns that signal such privileges and rights. Similar to CH\textit{PD} 6, it is important to note that although a item may specify “right,” based on our definitions the proper classification may be as a privilege.

\begin{table}[h]
\centering
\begin{tabular}{ |p{0.8\textwidth}| }
\hline
If the item is in one of the following formats:
\hline
- [actor] has the authority to [action]  
- [actor] shall have the right to [action]  
\hline
The item is an entitlement. Classify the entitlement as a \textit{privilege} or a \textit{right} using CH 1. 
\hline
\end{tabular}
\end{table}

\textbf{CH\textit{DUA} 3: Explicit Privileges and Rights}

Consider the following item from North Carolina Center for Health Statistics DUA [105]:

\begin{quote}
Data Recipient shall have the right to use all IIHI provided to it by the Data Custodian for the Research, Public Health or Health Care Operations purposes as listed below:
\end{quote}

Based on CH\textit{DUA} 3, the item is an entitlement, where the \textit{data recipient} is the actor and \textit{use} is the action. Using CH 1, we classify the entitlement as a privilege because \textit{use} is an independent action.

CH\textit{DUA} 4 classifies items within which the actor explicitly expresses agreement to an action.

\begin{table}[h]
\centering
\begin{tabular}{ |p{0.8\textwidth}| }
\hline
If the item is in the following format:
\hline
[actor] agrees to [action]  
\hline
Classify the item as a \textit{commitment}. 
\hline
\end{tabular}
\end{table}

\textbf{CH\textit{DUA} 4: Agree to Action}

Consider the following item from University of Cincinnati’s HIPAA Privacy Data Use Agreement [129]:
Recipient hereby agrees to fully comply with the requirements under HIPAA as applicable with respect to Limited Data Set information, including, without limitation, 45 C.F.R. §164.514, throughout the term of this Agreement.

Based on CH\textsubscript{DUA} 4, we classify the item as a commitment, where the \textit{recipient} is the actor and \textit{comply} is the action.

CH\textsubscript{DUA} 5 classifies items within which the actor explicitly expresses agreement to an object. These commitments have \textit{agree} as the action.

<table>
<thead>
<tr>
<th>If the item is in one of the following format:</th>
</tr>
</thead>
<tbody>
<tr>
<td>- [actor] agrees to [object]</td>
</tr>
<tr>
<td>- [actors] agree to [object]</td>
</tr>
</tbody>
</table>

Classify the item as a \textbf{commitment}. The commitment has \textit{agree} as the action.

\textbf{CH\textsubscript{DUA} 5: Agree to Object}

Consider the following item from North Carolina Center for Health Statistics DUA [105]:

\textbf{The Parties agree to the provisions of this DUA in order to address the requirements of HIPAA and to protect the interest of both Parties.}

Based on the use of CH\textsubscript{DUA} 5, we classify the item as a commitment with the action \textit{agree} and the object \textit{provisions of DUA}.

Items containing the natural language patterns in CH\textsubscript{DUA} 6 have the object, rather than the actor, as the subject of the item. We found that the actor of these items is implied as the recipient. We classify these items as commitments, privileges, or rights depending on the pattern contained within them.
- If the item is in one of the following formats:
  - [object] shall [action]
  - [object] is [past participle of action]
  - [object] must be [past participle of action]
  - [object] will be [past participle of action]
  - [object] shall be [past participle of action]
  - [object] should be [past participle of action]

Classify the item as a commitment of the recipient.

- If the item is in the following format:
  
  [object] may be [past participle of action]

The item is an entitlement of the recipient. Classify the entitlement as a privilege or a right using CH 1.

\textbf{CH}_DUA \textit{6}: Object as Subject

Consider the following item from a health system prototype DUA:

Data for presentation will be aggregated as needed to mask identification of individuals and hospitals.

Using \textit{CH}_DUA \textit{6}, we classify the item as a commitment. The action is aggregate, and the object is data for presentation.

\textit{CH}_DUA \textit{7} classifies items that contain embedded commitments: (1) a commitment from the actor with the action agree, and (2) commitments from each of the parties with the given action.

If the item is in one of the following format:

- [actor] agrees that [parties] shall [action]
- [actor] agrees that [parties] will [action]

The item contains multiple commitments:

- a commitment from the actor with the action agree, and
- commitments from each of the parties with the given action.

\textbf{CH}_DUA \textit{7}: Embedded Commitments
Consider the following item from a health system prototype DUA:

Recipient agrees that it, and any employees, agents and subcontractors to whom it discloses the IIHI, will not use the IIHI other than as permitted by this DUA.

Using CH_PD 7, the item contains multiple commitments. The recipient is the actor, the parties are recipient, employees, agents, and subcontractors, and the action is use.

### 3.2.4 Documenting Attribute Heuristics (DAH)

The attributes defined in Section 2.2 are found by asking questions that relate to the attribute definitions [114]. The same documenting attribute heuristics are used for policy documents and DUAs.

We use the following two items, Item 1 and Item 2, from Dossia’s Privacy Statement [60] as examples for documenting attributes.

---

**Dossia collects the personal information you voluntarily enter into the Dossia website, including the health information you enter, or authorize others to enter, into your Personally-Controlled Health Record (PCHR).**

**Item 1:** Item from Dossia’s Privacy Statement [60]

---

**Unless you explicitly and specifically consent, Dossia will not disclose your health information or contact information to third parties for them to use for marketing purposes.**

**Item 2:** Item from Dossia’s Privacy Statement [60]

---

The actor is the responsible stakeholder who performs the given action. DAH 1 provides the questions employed to identify the actor of an item.

Find the actor, or responsible stakeholder, by asking:

- *Who is the subject of this item?*
- *Who is performing this action?*

The answer to these questions will be the actor.

**DAH 1:** Actor
When we ask the questions in DAH 1 of Item 1 and Item 2 from Dossia, we find that the subject of both items is the organization.

The action is the action (verb) that the actor performs. DAH 2 provides the question employed to identify the action of an item.

<table>
<thead>
<tr>
<th>Find the action by asking:</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>What action is the actor performing?</em></td>
</tr>
<tr>
<td>The answer to this question will be the <em>action</em>.</td>
</tr>
</tbody>
</table>

**DAH 2: Action**

When asking the question in DAH 2 for Item 1, we find that the action is *collect*. For Item 2, we need the next heuristic to determine the action.

Items often contain negative actions (actions containing *not*), which express actions that cannot and/or will not be done. We attempt to have positive actions whenever possible. When a negative action is present, there are two separate cases depending on the presence of conditions, which we address with DAH 3.

<table>
<thead>
<tr>
<th>For negative actions (actions containing <em>not</em>), there are two separate cases—those with conditions and those without conditions.</th>
</tr>
</thead>
<tbody>
<tr>
<td>- If the item contains a negative action and a condition, use the contrapositive. Document the negated action (resulting in a positive action after removing the double negative) and the negated condition with <em>only</em> added to it.</td>
</tr>
<tr>
<td>- If the item contains a negative action and no condition, document the negative action.</td>
</tr>
</tbody>
</table>

**DAH 3: Negative Actions**

For Item 2, we must use DAH 3. The item contains a negative action and a condition; therefore, we use the contrapositive. The action becomes *disclose* and the condition is *only if user explicitly and specifically consent*.

The *object* is the item upon which the actor’s action is acting. DAH 4 provides the question employed to identify the object of an item.
Find the object by asking:

*What is the object of the actor’s action?*

The answer to this question will be the object.

**DAH 4: Object**

Using DAH 4, Item 1 and Item 2 have the following objects, respectively:

1. *user’s personal information*
2. *user’s health information or contact information*

The *object’s source* is the originator of the object. DAH 5 provides the questions employed to identify the object’s source of an item.

Find the object’s source by asking:

- *Who is the originator of the object?*
- *What is the originator of the object?*

The answer to these questions will be the *object’s source*.

**DAH 5: Object’s source**

By answering the second question for Item 1, we find that the object’s source is *organization website*. Item 2 does not have an object’s source.

The *target* is the intended recipient of the actor’s action. DAH 6 provides the question employed to identify the target of an item.

Find the target by asking:

*Who is the intended recipient of the actor’s actions?*

The answer to this question will be the *target*.

**DAH 6: Target**

Item 1 does not contain a target. Using DAH 6, we find the target to be *third parties* for Item 2.
The purpose is the reason for performing the action. DAH 7 provides the question employed to identify the purpose of an item.

Find the purpose by asking:

*What is the reason for performing this action?*

The answer to this question will be the purpose.

**DAH 7: Purpose**

Item 1 does not have a purpose. Using DAH 7, we found that the purpose of Item 2 is *marketing purposes*.

If an item contains multiple purposes, we split the item into multiple items that each express a single purpose using DAH 8 [72, 121].

**DAH 8: Multiple Purposes**

Item 1 and Item 2 do not contain multiple purposes.

The condition is the restriction or pre-condition on performing the action. DAH 9 provides the question and terms employed to identify the condition of an item.

Find the condition by asking:

*What is the restriction on performing the action?*

The answer to this question will be the condition.

Additionally, the condition is often found within an item by identifying the following terms:

- if
- unless
- when

**DAH 9: Condition**
Using DAH 9, we found the condition of Item 1 to be the following: voluntarily entered into user’s Personally-Controlled Health Record (PCHR). For Item 2 with DAH 3, we noted the condition as only if user explicitly and specifically consent.

Like purposes, if an item contains multiple conditions, we split the item into multiple items that each express a single condition using DAH 10 [72, 121]. This splitting occurs during the classification step because when the item is split the resulting items will have the same classification and many of the same attribute values.

| If an item contains multiple conditions joined by the logical OR, then the item is split into multiple items—each of which contains a single condition. |
| DAH 10: Multiple Conditions |

Neither Item 1 or Item 2 had multiple conditions.

The examples, or scenarios, illustrate how commitments, privileges, and rights are executed. DAH 11 provides the terms employed to identify the examples of an item.

| Examples are indicated by one of the following formats: |
| - including [...] |
| - For example, [...] |
| - such as [...] |
| - e.g., [...] |

Where [...] is the example.

| DAH 11: Examples |

Item 1 contained examples which were signaled by including. The examples are health information user enters, or authorizes others to enter, into your Personally-Controlled Health Record (PCHR). Item 2 did not have examples.

### 3.2.5 Documenting Rationale Heuristics (DRH)

Requirements engineers note the rationale, or the basis for the classifiable item being included within the document. In policy documents and DUAs, the rationale can be determined to be external based on domain knowledge or the presence of phrases. Otherwise the items are considered to have internal rationale. Recall, classifiable items that have external rationale are
Documenting Rationale Heuristics for Policy Documents (DRHP)

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.

Consider the following item from CIGNA Group Insurance Privacy Notice [40]:

We do not disclose any protected information about our customers or former customers to anyone except as permitted by law.

Using DRHP 1, the item has external rationale because it contains the phrase permitted by law.

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.

Documenting Rationale Heuristics for Policy Documents (DRHP)

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.

The rationale for other classifiable items is determined by knowledge of specific legal texts. We found that policy documents sometimes express items based on the Children's Online Privacy Protection Act (COPPA). DRHP 2 uses COPPA to document the rationale as external.
If the classifiable item refers to children under the age of 13, then document the rationale of the classifiable item as **external** based on the Child Online Privacy Protection Act (COPPA).

**DRH$_{PD}$ 2: COPPA for Rationale**

Consider the following item from Microsoft Online Privacy Statement [99]:

> When a Microsoft site does collect age information, and users identify themselves as under 13, the site will seek to obtain consent from parents for the collection, use and sharing of their children’s personal information.

Using DRH$_{PD}$ 2, we document that the item has an external rationale.

Other classifiable items require legal domain knowledge in order to document the rationale as external. A minimal, high-level awareness of various relevant legal texts is necessary to document the rationale of these classifiable items. DRH$_{PD}$ 3 documents rationale based on domain knowledge.

Document the rationale of the classifiable item as **external** when the analyst knows that the classifiable item is based on legal texts, even though the classifiable item does not explicitly state legal texts as the rationale of the classifiable item.

**DRH$_{PD}$ 3: Domain Knowledge for Rationale**

Consider the following item from drugstore.com’s Pharmacy Notice of Privacy Practices [61]:

> You have the right to receive an accounting of disclosures of your PHI that we have made after April 14, 2003 for purposes other than (1) for drugstore.com’s treatment, payment, or health care operations, (2) to you or based upon your authorization and (3) for certain government functions.

Based on analyst’s knowledge of HIPAA and through the use of DRH$_{PD}$ 3, we document the item’s rationale as external.

While we use the first three documenting rationale heuristics for policy documents to document external rationale, we employ the final heuristic (DRH$_{PD}$ 4) to document internal rationale. DRH$_{PD}$ 4 denotes internal rationale when an external rationale is not noted with another DRH$_{PD}$.
If the rationale of the classifiable item is not documented as external using DRH\textsubscript{PD} 1, DRH\textsubscript{PD} 2, or DRH\textsubscript{PD} 3, document the rationale as \textbf{internal} because organizational practice or procedure is the basis of the action.

\begin{center}
\textbf{DRH}_{PD} 4: Internal Rationale
\end{center}

Consider the following item from Dossia’s Privacy Statement [60]:

Dossia will not call you for marketing purposes.

The item is not documented as having external rationale using DRH\textsubscript{PD} 1, DRH\textsubscript{PD} 2, or DRH\textsubscript{PD} 3; therefore, through DRH\textsubscript{PD} 4, we classify the item as having internal rationale.

\section*{Documenting Rationale Heuristics for DUAs (DRH\textsubscript{DUA})}

Similar to policy documents, the rationale for classifiable items within DUAs can be documented as external using key phrases. Two types of phrases signal external rationale: (1) phrases that contain the words \textit{law} and/or \textit{regulation} (DRH\textsubscript{DUA} 1) and (2) phrases that contain HIPAA (DRH\textsubscript{DUA} 2).

If the classifiable item includes one of the following phrases:

- in accordance with law
- in accordance with regulation
- permitted by law
- permitted by regulation
- required by law
- required by regulation
- specified by law
- specified by regulation
- with applicable law
- with applicable regulation

Document the rationale as \textbf{external} because a legal text is the basis of the action.

\begin{center}
\textbf{DRH}_{DUA} 1: Key Phrases for External Rationale
\end{center}

Consider the following item from a health system prototype DUA:

As specified by law, data provided under this DUA shall not be used for commercial purposes.
Using DRH\textsubscript{DUA} 1, we document the item’s rationale as external based on the phrase \textit{specified by law}.

Some DUA items, explicitly state the basis for the item as HIPAA. DRH\textsubscript{DUA} 2 uses phrases which include the HIPAA to document an item’s rationale as external.

If the classifiable item includes one of the following phrases:

- requirements of HIPAA
- under HIPAA
- with HIPAA

Document the rationale as \textbf{external} because legal text, specifically the HIPAA, is the basis of the action.

\textbf{DRH\textsubscript{DUA} 2: HIPAA Explicit Phrases}

Consider the following item from University of Cincinnati’s HIPAA Privacy Data Use Agreement [129]:

Recipient hereby agrees to fully comply with the requirements under HIPAA as applicable with respect to Limited Data Set information, including, without limitation, 45 C.F.R. 164.514, throughout the term of this Agreement.

Using DRH\textsubscript{DUA} 2, we document the rationale as external because it contains the phrase \textit{under HIPAA}.

Other classifiable items within DUAs have external rationale based on the portion of the HIPAA that discusses DUAs (45 C.F.R. §164.514(e) as discussed in Section 1.1.2).

Document the rationale of the classifiable item as \textbf{external} when the analyst knows that the classifiable item is based on 45 C.F.R. §164.514(e), even though the classifiable item does not explicitly state 45 C.F.R. §164.514(e) as the rationale of the classifiable item.

\textbf{DRH\textsubscript{DUA} 3: DUAs within HIPAA}

Consider the following item from a health system prototype DUA:

Recipient shall immediately report to Data Custodian any use or disclosure of the IIHI that is not provided for in this DUA of which the Recipient becomes aware.

Using DRH\textsubscript{DUA} 3 and the following knowledge:
§164.514(c)(4)(ii)(C)(3): “Report to the covered entity any use or disclosure of the information not provided for by its data use agreement of which it becomes aware”

we document the item as having external rationale.

While we use the first three documenting rationale heuristics for DUAs to document external rationale, we use the final DRH$_{DUA}$ to document internal rationale. DRH$_{DUA}$ 4 denotes internal rationale when an external rationale is not noted with another DRH$_{DUA}$.

If the rationale of the classifiable item is not documented as external using DRH$_{DUA}$ 1, DRH$_{DUA}$ 2, or DRH$_{DUA}$ 3, document the rationale as internal because organizational practice or procedures is the basis of the action.

**DRH$_{DUA}$ 4: Internal Rationale**

Consider the following item from a health system prototype DUA:

> For oral or written presentations or publications, the source of the data must be attributed to the covered entity.

The item is not documented as having external rationale using DRH$_{DUA}$ 1, DRH$_{DUA}$ 2, or DRH$_{DUA}$ 3; therefore, through DRH$_{DUA}$ 4, we document the item as having internal rationale.

### 3.2.6 Forming CPR Heuristic (FH)

The final substep of the classify step is forming the CPRs. We form CPRs through the use of FH 1, which employs the attributes (bracketed items) identified in the earlier substep (Section 3.2.4). Note that if an attribute does not exist in an item, then the attribute is not included in the template.

Form the commitment, privilege, or right using the following template:

```
[action] [object] from [object’s source] to/with [target] for/in order to [purpose] given/if [conditions]
```

**FH 1: Forming CPR**

Examining Item 1 and Item 2 from Section 3.2.4, we apply FH 1 to form the following CPRs.

1. collect user’s personal information from organization website if voluntarily entered into user’s Personally-Controlled Health Record (PCHR)
2. disclose user’s health information or contact information to third parties
   for marketing purposes only if user explicitly and specifically consent

It can be seen that when the items did not include certain attributes, these attributes were not present in the resulting CPRs.

3.3 Operationalize

To operationalize commitments, privileges, and rights as requirements, we employ templates that take the classification into consideration. We employ a Requirement Type Heuristic, RTH 1, to determine whether we operationalize a CPR into a system requirement or operational requirement. Recall, system requirements specify the capabilities of the system or software, and operational requirements specify “business rules or operational procedures” [20] outside of the system.

<table>
<thead>
<tr>
<th>If the CPR is a capability of the system or software, then operationalize it as a system requirement. Otherwise, operationalize it as a operational requirement.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RTH 1: Requirement Type</strong></td>
</tr>
</tbody>
</table>

Operationalizing Heuristics (OH) describe how to translate CPRs into requirements based on the classification and whether we operationalize them as system requirements or operational requirements. Based on the action being pledged, the system requirement template for commitments (OH 1) contains “require,” and the operational requirement template for commitments (OH 2) contains “shall.” Based on the action being entitled, the system requirement template for privileges and rights (OH 3) contains “allow,” and the operational requirement template for privileges and rights (OH 4) contains “shall be able to.”

<table>
<thead>
<tr>
<th>Given a commitment that should be operationalized as a system requirement based on RTH 1, operationalize it as a system requirement using the following template:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OH 1: System Requirement for Commitments</strong></td>
</tr>
</tbody>
</table>

Consider the following item from a health system prototype DUA:

Moreover, all access to the system and its data will be stored in a secure log file for auditing.
Using RTH 1, we find that the item will be operationalized as an system requirement because it explains a capability of the system. The item is classified as a commitment based on CH_{DU_A} 6. By applying the DAHs and FH 1, we have the following commitment:

store all access to the system and its data in a secure log file for auditing

The commitment is operationalized as the following requirement using OH 1:

The system shall require the recipient to store all access to the system and its data in a secure log file for auditing.

Given a commitment that should be operationalized as a operational requirement based on RTH 1, operationalize it as a operational requirement using the following template:

The [actor] shall [commitment].

**OH 2: Operational Requirement for Commitments**

Consider the following item from Dossia’s Privacy Statement [60]:

Unless you explicitly and specifically consent, Dossia will not call you for marketing purposes.

Using RTH 1, we find that the item will be operationalized as an operational requirement because it is not a capability of the system. Instead it relates to operational procedures. The item is classified as a commitment based on CH_{PD} 1. By applying the DAHs and FH 1, we have the following commitment:

call user for marketing purposes, only if user explicitly and specifically consent

The commitment is operationalized as the following requirement using OH 2:

The organization shall call user for marketing purposes, only if user explicitly and specifically consent.

Given a privilege or a right that should be operationalized as a system requirement based on RTH 1, operationalize it as a system requirement using the following template:

The system shall allow the [actor] to [privilege/right].

**OH 3: System Requirement for Privileges and Rights**
Consider the following item from Pfizer Privacy [111]:

Generally, you can browse our Web sites without disclosing any personally identifiable information.

Using RTH 1, we find that the item will be operationalized as an system requirement because it explains a capability of the system. The item is classified as a privilege based on CH$_{PD}$ 1 and CH 1. By applying the DAHs and FH 1, we have the following privilege:

\[
\text{browse the organization's websites without disclosing any personally identifiable information}
\]

The privilege is operationalized as the following requirement using OH 3:

\[
The \text{system shall allow the user to browse the organization's websites without disclosing any personally identifiable information.}
\]

Given a privilege or a right that should be operationalized as a operational requirement based on RTH 1, operationalize it as a operational requirement using the following template:

\[
The \text{[actor] shall be able to [privilege/right].}
\]

**OH 4:** Operational Requirement for Privileges and Rights

Consider the following item from a health system prototype DUA:

Such research may be conducted, subject to relevant Institutional Review Board approval.

Using RTH 1, we find that the item will be operationalized as an operational requirement because it is not a capability of the system. Instead it relates to operational procedures. The item is classified as a privilege based on CH$_{DUA}$ 6 and CH 1. By applying the DAHs and FH 1, we have the following privilege:

\[
\text{conduct research if approved by Institutional Review Board}
\]

The privilege is operationalized as the following requirement using OH 4:

\[
The \text{recipient shall be able to conduct research if approved by Institutional Review Board.}
\]
3.4 Summary

This chapter presented heuristics for performing CPR analysis on policy documents and DUAs. We discussed three Parsing Heuristics (PH) that aid analysts in parsing a document into statements. We presented multiple types of heuristics that are employed during the substeps of the classify step:

- Item Heuristic (IH) to distinguish between classifiable and unclassifiable items;
- Helper Heuristics (HH) to assist with various tasks within the classify step;
- Classifying Heuristics (CH) to classify each classifiable item as a commitment, privilege, or right based on natural language phrases;
- Documenting Attribute Heuristics (DAH) to document the attributes of the classifiable items;
- Documenting Rationale Heuristics (DRH) to document the rationale of the classifiable items; and
- Forming Heuristic (FH) to form the commitments, privileges, and rights using the documented attributes.

The final step of CPR analysis, operationalize, contained two types of heuristics: Requirement Type Heuristic (RTH) and Operationalizing Heuristics (OH). A summary of the heuristics is provided in Appendix A, beginning on page 142.
Chapter 4

Multi-Case Studies

We keep moving forward, opening new doors, and doing new things, because we’re curious and curiosity keeps leading us down new paths.
- Walt Disney

This chapter discusses the development and validation of CPR analysis theory and methodology through two multi-case studies—a formative multi-case study and a summative multi-case study. We conducted the formative multi-case study to shape our theory and methodology (discussed in Chapter 2) and develop our heuristics (discussed in Chapter 3) by examining the context of the documents. We conducted the summative multi-case study to validate the theory and methodology developed during the formative multi-case study by examining whether our CPR analysis could be applied to documents outside the data set of the formative study.

The cases examined with the studies relate to two types of documents—policy documents and data use agreements (DUAs). Recall, we use the term policy document to refer to any policy that an organization has. The policy documents we examined were all posted on the organizations’ websites. These policy documents describe the organizations’ practices, including how consumers’ or users’ personal information will be collected, disclosed, protected, shared, stored, and used. Also, recall, a data use agreement is a legally binding agreement between the covered entity and the limited data set recipient that specifies how the limited data set will be used and to whom it will be disclosed [104].

The remainder of this chapter is organized as follows: Section 4.1 discusses the formative multi-case study; Section 4.2 discusses the summative multi-case study; Section 4.3 describes threats to validity for the study; and Section 4.4 summarizes the multi-case studies.
4.1 Formative Multi-Case Study

As previously stated, we conducted the formative multi-case study to develop the theory and methodology for extracting requirements from policy documents and DUAs. In this section, we discuss the research questions, documents analyzed, and results of the formative multi-case study.

4.1.1 Research Questions

The formative multi-case study was designed to answer the following questions:

*RQ*$_{F1}$: What information within policy documents and DUAs is important for requirements and compliance?

*RQ*$_{F2}$: How can this information, which is important for requirements and compliance, be extracted from the documents and operationalized as requirements?

The first question (*RQ*$_{F1}$) investigates the types of information within the documents that are important for requirements and compliance. The answer to this question became our classifications—commitments, privileges, and rights (CPRs)—or theory. The second question (*RQ*$_{F2}$) addresses how to extract CPRs from the documents and operationalize them as requirements, which relates to CPR analysis—our methodology and heuristics.

4.1.2 Documents Analyzed

We examined twenty-one documents for the formative multi-case study. For the policy document case, we analyzed seventeen policy documents [132, 133]. For the DUA case, we analyzed four DUAs [119].

**Policy Documents Analyzed**

We analyzed seventeen policy documents from four healthcare organizations representing four healthcare sectors—Aetna$^1$ (a health insurance company), drugstore.com$^2$ (an online drugstore), GlaxoSmithKline (GSK)$^3$ (a pharmaceutical company), and Dossia$^4$ (a Personal Health Records, PHR, system). We analyzed later versions of the same policy documents that were previously analyzed using a goal-driven approach [24]. For Antón et al.’s goal analysis, the policy documents represented three healthcare sectors: health insurance companies, online drugstores, and

---

$^1$http://www.aetna.com  
$^2$http://www.drugstore.com  
$^3$http://www.gsk.com  
$^4$http://www.dossia.org
pharmaceuticals [24]; we also analyzed policy documents from a fourth healthcare sector that did not exist at the time of the previous study—Personal Health Records (PHR) systems [16, 131]. We list the analyzed documents in Figure 4.1. Two analysts familiar with analyzing policy documents for requirements artifacts and knowledge of healthcare legal domain performed this analysis.

- Aetna (Health Insurance)
  - Notice of Information Practices by Plan Type
    * Large Case Pension [3]
    * Life and Disability [4]
  - Notice of Privacy Practices by Plan Type
    * Employee Assistance Plan [5]
    * Insured Health Benefits Plan [7]
    * Long-Term Care Plan [8]
    * Rx Home Delivery [9]
    * Strategic Resource Company (SRC) [10]
    * Student Health Plan [11]
  - Privacy Notices [12]
  - Web Privacy Statement [13]
- drugstore.com (Drugstore)
  - Pharmacy Notice of Privacy Practices [61]
  - Privacy Policy [62]
  - Terms of Use [63]
- GlaxoSmithKline (GSK) (Pharmaceutical)
  - Internet Privacy Statement [77]
  - Terms and Conditions [78]
- Dossia (PHR)
  - Privacy Statement [60]

Figure 4.1: Policy Documents Analyzed in Formative Study

The seventeen analyzed documents ranged in length from one to fourteen pages. Aetna had eleven documents accounting for thirty pages, whereas drugstore.com had a total of twenty-seven pages within its three documents. GlaxoSmithKline’s two documents were each two pages in length, and Dossia’s one document accounted for nine pages. These policy documents included a variety of document types, which are shown in Figure 4.1. Most documents contained a
combination of natural language paragraphs and lists, but some documents were completely in natural language paragraph format.

Seven of Aetna’s documents were “Notice of Privacy Practices” for different healthcare benefits plans. When we analyzed these seven documents, we found that they contained the same commitments, privileges, and rights, only differing on the plan name. Based on this, we only discuss one of the documents, “Notice of Privacy Practices - Strategic Resource Company (SRC),” when we discuss the results of the study in Section 4.1.3. Aetna also had two “Notice of Information Practices” documents; however, these documents were structured differently and contained different information. Thus, we analyzed both of these documents.

The Pharmacy Notice of Privacy Practices for drugstore.com was unique within the analyzed policy documents because a portion of it included specific commitments, privileges, and rights of the organization with external rationale based on each U.S. state’s laws. We have not included the portion of document relating to specific state laws within our analysis because state laws were not discussed in any other document by any other organization.

**DUAs Analyzed**

We analyzed four HIPAA-governed DUAs. We examined a HIPAA-governed DUA from a health system prototype on which we were working and discuss in Chapter 7. To ensure generalizability, we also examined three sample HIPAA-governed DUAs that were publicly available online. The DUAs are different from the policy documents in that DUAs are signed and therefore establish a mutually agreed to contractual relationship. The DUAs analyzed are shown in Figure 4.2.\(^5\)

We analyze fewer DUAs than policy documents in our formative and summative studies because we only have access to a limited number of DUAs with a health system prototype to manage health data from a number of sources and support analytical decision making, and DUAs, other than samples like we analyze in our formative study, are typically not publicly available.

---

\(^5\)HealthSystem (HS) is a healthcare organization with which we were able to aid in their efforts to develop a software prototype to aid in predicting potential outbreaks of diseases. We discuss HealthSystem in Chapter 7.
The documents' lengths ranged from two to five pages, accounting for a total of thirteen pages. HS-I DUA consisted of five pages. AcademyHealth Model DUA accounted for two pages. The North Carolina Division of Public Health DUA was three pages in length. The length of the University of Cincinnati HIPAA Privacy DUA was three pages.

4.1.3 Results

In this section, we discuss the results of the formative study. Recall that the study was designed to answer two questions—$RQ_{F1}$ and $RQ_{F2}$. $RQ_{F1}$ asked what information within the documents is important for requirements and compliance. Our finding that commitments, privileges, and rights were important for requirements and compliance resulted in our classifications and theory. $RQ_{F2}$ asked how this information (commitments, privileges, and rights) is extracted from the documents and operationalized as requirements. Our CPR analysis methodology, including heuristics, were developed as a result of this formative study.

This section will also discuss the knowledge gained by applying CPR analysis. By examining the distribution of classifiable and unclassifiable items in the documents, the requirements engineer gains an understanding of what portions of the documents result in compliance requirements with which the system needs to comply. This distribution motivates the inclusion of these documents in the requirements phase of software in order to build compliant systems. By examining the distribution of commitments, privileges, and rights within the classifiable items, the requirements engineer also gains an understanding of the modality of the actions. In order to comply with the documents, it is important that the system maintain the modality of the actions. The commitments will translate to actions that the actor will perform within the system. The privileges and rights will translate to actions that the actor may perform within the system. Because the documents express items that are not strictly legally-based, we examine the distribution of internal and external rationale. By examining the distribution of the actors in the classifiable items, the requirements engineers gains an understanding of the parties that will be interacting with system.

Policy Document Results

We investigate the policy documents of four organizations based on several areas that we discussed above—types of items, classifications, rationales, and actors. We examine these areas for each organization individually and for all four organizations combined. The types of items refer to whether the item is classifiable or unclassifiable, as discussed in Section 2.2.2. We refer to commitment, privilege, and right as the classifications. As previously mentioned, a classifiable item either has an internal or external rationale. In policy documents, there are two possible actors: the organization and the user. In DUAs, the actors are the data custodian, the recipient,
and other various actors, which we group as “other.”

The percentage of items that were classifiable and unclassifiable for each organization and all organizations combined are shown in Figure 4.3. Examining Aetna’s policy documents, we found that 55% of the 283 items were classifiable, or important for requirements and compliance. drugstore.com had more total items (416) than the other organizations; 57% of which were classifiable. Of GSK’s 98 items, only 40% were classifiable. Thirty-eight percent of the 196 items in Dossia’s policy document were classifiable. For all of the organizations combined, 51% of the 939 items were classifiable. Therefore, just over half of the items within the policy document case of the formative study were important for requirements and compliance. These classifiable items were operationalized as requirements. This further supports the need to include policy documents in the requirements phase of the software lifecycle. These compliance requirements are important to ensure compliance with governing policies.

The distribution of classifications among classifiable items for each organization and all organizations combined are shown in Figure 4.4. Of Aetna’s 156 classifiable items, 46% were commitments, 52% were privileges and 2% were rights. Of drugstore.com’s 239 classifiable items, 48% were commitments, 50% were privileges, and 2% were rights. Most of GSK’s 39 classifiable items (64%) were commitments, with the remainder being privileges. Dossia’s 75 classifiable items only contained commitments (56%) and privileges (44%). For all of the organizations combined, the 509 classifiable items had the following classifications: 50% were commitments, 49% were privileges, and 1% were rights. Therefore, for all of the organizations combined, half of the classifiable items were pledges and half of the classifiable items were entitlements. The health insurance and online drugstore organizations had more entitlements than pledges, whereas the pharmaceutical and PHR organizations had more pledges than entitlements. Recall, one of
the motivations for developing the CPR methodology was our inability to apply the legal-based approach to the policy documents because legal-based approaches do not address the pledges contained in policy documents. The classification results for the policy document case of the formative study show that the pledges, or commitments, comprise a large part of requirements expressed by policy documents. Thus, an approach that does not actively uncover these pledges would leave a significant portion of the compliance requirements unaddressed.

![Formative Study - Policy Documents Case - Classifications](image)

**Figure 4.4:** Formative Study - Policy Documents Case - Classifications

The rationales of classifiable items for each organization and all organizations combined are shown in Figure 4.5. The majority (85%) of Aetna’s 156 classifiable items had internal rationale. Most of drugstore.com’s 239 classifiable items had internal rationale (92%). Most of GSK’s 39 classifiable items (95%) had internal rationale. The rationale for 92% of Dossia’s 75 classifiable items was internal. For all the organizations combined, 90% of the 509 classifiable items had internal rationale. The fact that the policy documents, across the board, were comprised of mostly classifiable items with internal rationale suggests that the policy documents express mostly organizational practices rather than items with bases in legal texts. These results provide support for using an analysis approach that is not legal-based when analyzing policy documents for requirements. A significant portion of the items would likely be missed by a legal-based approach because the items with internal rationale would not be captured.

The percentage of actors within classifiable items for each organization and all organizations combined are shown in Figure 4.6. Most of Aetna’s 156 classifiable items (72%) had the `organization` as the actor. The `organization` was the actor in 62% of drugstore.com’s 239
classifiable items. For GSK’s 39 classifiable items, 79% had the organization as the actor. The organization was the actor for 68% of Dossia’s 75 classifiable items. Combined, 67% of the 509 classifiable items were commitments, privileges, or rights of the organization. The fact that the policy documents, across the board, had the organization as the actor in the majority of the classifiable items suggests that the documents express mostly actions that the organization performs. This relates to the fact, stated in Section 1.1.1, that policy documents describe the organizations’ practices, including how users’ personal information will be collected, disclosed, protected, shared, stored, and used.

Figure 4.5: Formative Study - Policy Documents Case - Rationale

Figure 4.6: Formative Study - Policy Documents Case - Actors
DUA Results

We investigate the four DUAs based on several areas that we discussed above—types of items, classifications, rationales, and actors. We examine these areas for each document individually and for all four documents combined. The percentage of items that were classifiable and unclassifiable for each DUA and all DUAs combined are shown in Figure 4.7. For the HS-I DUA, 50% of the 134 items were classifiable, or important for requirements and compliance. Only 39% of the 67 items within the AcademyHealth (AH) Model DUA were classifiable. Of the North Carolina Division of Public Health (NCDPH) DUA’s 65 items, 58% were classifiable. Only 37% of the 60 items in the University of Cincinnati (UC) HIPAA Privacy DUA were classifiable. Examining the data for all DUAs within the formative study, 46% of the 326 items were classifiable. Therefore, less than half of the items were important for requirements and compliance. Although less than half of the items were operationalized as requirements, having classifiable items contained within the DUAs supports the inclusion of these documents in the requirements phase of the software lifecycle. The resulting requirements are important contractual compliance requirements with which the recipient must comply and would likely have been missed if the DUAs were not examined by the requirements engineers.

![Figure 4.7: Formative Study - DUA Case - Types of Items](image)

The percentage of classifications for each DUA and all DUAs combined are shown in Figure 4.8. Most of HS-I DUA’s 67 classifiable items (91%) were commitments, with the remaining items being privileges. Commitments accounted for 65% of AH DUA’s 26 classifiable items; privileges and rights accounted for 31% and 4%, respectively. Most of NCDPH DUA’s 38 classifiable items were commitments (92%). Almost all of UC DUA’s 22 classifiable items were commitments (95%). Combined, most of the 153 classifiable items (87%) were commitments. Therefore, the DUAs expressed mostly pledges from one party to another. Whereas half of the
classifiable items in the policy documents were pledges, we can see that DUAs focused more on pledges. This focus on pledges may relate to the fact that DUAs are legally binding agreements that contain acceptable uses and disclosures of data.

![Figure 4.8: Formative Study - DUA Case - Classifications](image)

The percentage of rationales within classifiable items for each DUA and all DUAs combined are shown in Figure 4.9. Most of HS-I DUA’s 67 classifiable items (54%) had external rationale. The majority of the 26 classifiable items (88%) in AH DUA’s had external rationale. For NCDPH, 84% of the 38 classifiable items had external rationale. The majority of UC DUA’s 22 classifiable items had external rationale (59%). Combined, the majority of the 153 classifiable items (69%) had external rationale. The fact that most of the DUAs’ classifiable items had external rationale may be due to the fact that regulations adopted pursuant to the HIPAA specify that a DUA must exist when certain protected health information is used or disclosed.

The percentage of actors within classifiable items for each DUA and all DUAs combined are shown in Figure 4.10. As we previously discussed, DUAs are established between data custodians and limited data set recipients. These are the two main actors in DUAs; however, other actors are also present within DUAs, which we group together for Figure 4.10. The recipient was the actor in most of HS-I DUA’s 67 classifiable items (72%). The actors in AH DUA’s 26 classifiable items were recipient (79%) and data custodian (21%). In NCDPH DUA, 58% of the 38 classifiable items had the recipient as the actor. The majority of UC DUA’s 22 classifiable items (79%) had the recipient as the actor. Combined, the majority of the 153 classifiable items (72%) had the recipient as the actor. This was not surprising as the recipient...
4.2 Summative Multi-Case Study

We conducted the summative multi-case study to validate the theory and methodology developed during the formative multi-case study. In this section, we discuss the research questions, documents analyzed, and results of the summative multi-case study.

4.2.1 Research Questions

The summative multi-case study was designed to answer the following questions:
RQ_{S1}: Can CPR analysis be applied to documents outside the data set of the formative study?

RQ_{S2}: Does CPR analysis cover everything within the documents that is important for requirements and compliance?

The first question (RQ_{S1}) addresses whether CPR analysis can be validated for documents outside the data set of the formative study. The second question (RQ_{S2}) addresses whether CPR analysis includes heuristics for everything within the documents that is important for requirements and compliance.

### 4.2.2 Documents Analyzed

We examined thirty-five documents for the summative multi-case study. For the policy document case, we analyzed thirty-two policy documents [134]. For the DUA case, we analyzed three DUA.

**Policy Documents Analyzed**

We analyzed the policy documents of eight healthcare organizations. Like with the formative study, the organizations whose policy documents we examined are the same organizations that were analyzed by Antón et al. in their work that mined goals from policy documents of healthcare institutions [24] as well as policy documents of PHR organizations. We analyzed a total of thirty-two policy documents in the four healthcare sectors—health insurance (Aflac\(^6\) and Cigna\(^7\)), online drugstore (DestinationRx\(^8\) and HealthCentral\(^9\)), pharmaceutical (Novartis\(^10\) and Pfizer\(^11\)), and PHR system (Google Health\(^{12,13}\) and Microsoft HealthVault\(^{14}\)). The analyzed documents are listed in Figure 4.11.

For the case study, two analysts familiar with mining policy documents for requirements artifacts and with knowledge of healthcare legal domain performed the analysis. These policy documents include a variety of document types—legal statement, notice of privacy practices, privacy notice, privacy policy, privacy statement, service agreement, terms and conditions, terms of service, and terms of use. The documents varied in format from paragraph to list, or a combination of the two.

\(^6\)http://www.aflac.com/
\(^7\)http://www.cigna.com/
\(^8\)http://www.destinationrx.com/
\(^9\)http://www.healthcentral.com/
\(^10\)http://www.novartis.com/
\(^11\)http://www.pfizer.com/
\(^12\)http://www.google.com/health/
\(^14\)http://www.microsoft.com/en-us/healthvault/
• Health Insurance
  – Aflac
    * Privacy Policy [14]
    * Terms and Conditions [15]
  – Cigna
    * CIGNA Group Insurance Privacy Notice [40]
    * CIGNA International Expatriate Benefits [41]
    * CIGNA International Expatriate Benefits Privacy Forms [42]
    * Gramm-Leach-Bliley Privacy Notice [43]
    * Legal Disclaimer [44]
    * Notice of Privacy Practices
      · CIGNA Medical Group of Arizona [45]
      · CIGNA Medicare Services [46]
      · CIGNATURE Rx [47]
      · Standard Privacy Practices [48]
      · Tel-Drug [49]
    * Privacy Forms [50]
    * Privacy Information [51]
    * Public Online Privacy Statement [52]
    * Reminder Regarding Access to Notice of Privacy Practices [53]

• Drugstore
  – DestinationRx (DRx)
    * Privacy Policy [57]
    * Terms of Use [58]
  – HealthCentral (THCN)
    * Privacy Policy [126]
    * Terms of Use [127]

• Pharmaceutical
  – Novartis
    * Privacy policy [107]
    * Terms of use [108]
  – Pfizer
    * Privacy [111]
    * Privacy Policy [112]
    * Terms of Use [113]

• PHR
  – Google Health (GH)
    * Google Privacy Policy [80]
    * Privacy (FAQ) [81]
    * Privacy Policy [82]
    * Terms of Service [83]
  – Microsoft HealthVault (MHV)
    * Microsoft Online Privacy Statement [99]
    * Information Site Service Agreement [100]
    * Information Site Privacy Statement [101]

**Figure 4.11:** Policy Documents Analyzed in Summative Study
The 32 analyzed documents accounted for 154 pages of text. The health insurance organizations had the most documents (16 documents) and the most pages (69 total pages) as compared to the other three sectors. When comparing organizations, Cigna had the most documents (14 documents) as well as the most pages (54 pages). DestinationRx had the fewest pages with ten pages. There were two Cigna documents that were each a single page in length. The longest document was from Microsoft HealthVault (ten pages).

Similar to Aetna in the formative study, Cigna had five “Notice of Privacy Practices” documents for different plans. Analyzing these five documents, we found that they contained identical commitments, privileges, and rights across the documents. Therefore, when we discuss the results of the study in Section 4.2.3, we only discuss one of the documents, “Notice of Privacy Practices - CIGNATURE Rx.”

**DUAs Analyzed**

For the DUA portion of the summative study, we analyzed three HealthSystem DUAs. The documents are listed in Figure 4.12. These documents contain non-public information about the data custodian and recipient. For this reason, we have assigned document numbers to these documents and the one from the formative study.

- HS-II Data Use Agreement
- HS-III Data Use Agreement
- HS-IV Data Use Agreement

**Figure 4.12: DUAs Analyzed in Summative Study**

### 4.2.3 Results

In this section, we discuss the results of the summative study. Recall that we designed the study to address two questions—\(RQ_{S1}\) and \(RQ_{S2}\). \(RQ_{S1}\) asked whether CPR analysis, the approach developed in the formative study, could be applied to other documents. The documents within the summative study were from the same domain and sectors as the documents within the formative study. We found that CPR analysis could be applied to documents outside of the original (formative) data set. \(RQ_{S2}\) asked whether CPR analysis covered all of the information within the documents that is important for requirements and compliance. We found that CPR analysis did cover all of the information contained within the analyzed documents that was important for requirements and compliance. Given that some of the documents varied in format and wording, the summative study did yield new heuristics.
Policy Document Results

For all of the organizations combined, 51% of the 2875 items were determined to be classifiable, or important for requirements and compliance. Those items that were unclassifiable included headings, definitions, clarifications, and warnings. The percentage of items that were classifiable and unclassifiable for each organization and all organizations combined are shown in Figure 4.13. These results reinforce the results of the formative study, which show that policy documents should be considered for compliance purposes as these documents contain compliance requirements.

![Figure 4.13: Summative Study - Policy Document Case - Types of Items](image)

We found that most of the 1506 classifiable items were privileges and commitments (55% and 44%, respectively). The health insurance and PHR sectors had slightly more privileges than commitments, while online drugstore and pharmaceutical sectors had slightly more commitments than privileges. In all of the documents, fewer classifiable items were classified as rights (less than 1% combined) than privileges or commitments. The percentage of classifications for each organization and all organizations combined as shown in Figure 4.14. Whereas the classifiable items in the formative study were half entitlements and half pledges, based on the summative study we can see that this is not the case for all policy documents as the classifiable items in the summative study were mostly entitlements. Even though there were more entitlements than pledges, the presence of pledges reinforces the fact that these items need to be examined for requirements. An approach that does not actively uncover these pledges would leave a significant portion of the compliance requirements unaddressed.

The majority (93%) of the 1506 classifiable items were internal, which is similar to the results of the formative study. Because the policy documents contained more classifiable items with internal rationale than classifiable items with external rationale, the policy documents expressed
mostly organizational practices or procedures rather than actions required or entitled by legal
texts. The health insurance sector had the smallest percentage of classifiable items with internal
rationale among the sectors with 88%. The other three healthcare sectors each contained at least
95% classifiable items with internal rationale. The percentage of classifiable items’ rationales for
each organization and all organizations combined are shown in Figure 4.15. As in the formative
study, the fact that the classifiable items mostly had internal rationale supports the need for an
approach, like CPR analysis, to analyze the documents rather than a legal-based approach.
Similar to the formative study, we found that a majority (60%) of the classifiable items in policy documents of the summative study expressed actions that are performed by the organization. The percentage of actors within classifiable items for each organization and all organizations combined are shown in Figure 4.16.

![Figure 4.16: Summative Study - Policy Document Case - Actors](image)

**DUA Results**

The percentage of items that were classifiable and unclassifiable for each DUA and the DUAs combined are shown in Figure 4.17. The percentage of the classifiable items within HS-II, HS-III, HS-IV, and all DUAs combined were 54%, 49%, 24%, and 33%, respectively. HS-IV had the smallest percentage of classifiable items (important for requirements and compliance), whereas HS-II had the largest. Similar to our formative study, the presence of classifiable items supports the inclusion of DUAs in the requirements phase of the software lifecycle as the documents contain contractual compliance requirements that are operationalized from the classifiable items. The fact that a large percent of the items were unclassifiable shows that these documents contain many items that are clarifications, definitions, section headings, or warnings, which are not operationalized as requirements.

The percentage of classifications for each DUA and the DUAs combined are shown in Figure 4.18. All of the documents contained mostly commitments with at least 83% of the classifiable items being commitments. There were no rights present in these documents. Similar to our formative study, this focus on pledges may relate to the fact that DUAs are legally binding agreements that contain acceptable uses and disclosures of data. Thus, the documents may express the pledges of the actors based on the acceptable uses and disclosures.
Figure 4.17: Summative Study - DUA Case - Types of Items

Figure 4.18: Summative Study - DUA Case - Classifications
The percentage of rationales within classifiable items for each DUA and the DUAs combined are shown in Figure 4.19. All of the documents contained mostly classifiable items with internal rationale (HS-II 51%, HS-III 87%, and HS-IV 74%). This is the opposite of the formative study where the classifiable items mostly had external rationale. As we discuss in Section 7.2, HS-III and HS-IV were missing the items required by regulations pursuant to HIPAA and the data sets contained data elements not allowed within a limited data set (LDS). The higher percentage of internal rationale in the summative study as compared to the formative study could be due to these missing items.

![Figure 4.19: Summative Study - DUA Case - Rationale](image)

The percentage of actors within classifiable items for each DUA and the DUAs combined are shown in Figure 4.20. All of the documents focused on the actions of the data recipient, which was similar to the results of the formative study. Recall, that the DUAs contain the restrictions on the use of the data by the recipient that relate the actions of the recipient.

4.3 Threats to Validity

We examine the validity of our multi-case studies based on Yin’s validity tests, which include construct validity, internal validity, external validity, and reliability [130]. Because we make no inferences as a result of our case study, internal validity is not applicable [130].

Construct validity establishes “correct operational measures for the concepts being studied” [130]. Yin gives three methods to maintain construct validity: using multiple sources of evidence, establishing chain of evidence, and having key informants review draft case study report [130]. By extracting commitments, privileges, and rights from forty-nine policy documents for twelve
organizations in four different healthcare sectors and seven DUAs from seven groups, we have used multiple sources for our analysis. We establish a chain of evidence by: (1) rigorously following the steps of our methodology when analyzing the documents; (2) maintaining extensive records of all data collected; and (3) maintaining traceability links from documents to the commitments, privileges, and rights and then forward to the requirements. A group of experienced requirements engineering researchers have reviewed our case study reports. In addition, to mitigate this threat, two analysts worked together, debating any issues that were found while analyzing the documents.

External validity is maintained by “establishing the domain to which a study’s findings can be generalized” [130]. We recognize as a threat to external validity the fact that we have only analyzed documents from organizations in the healthcare domain. To mitigate this threat, we attempt to cover a broad range of healthcare organizations. We selected organizations from four different sectors within the industry for policy document case: health insurance companies, online drugstores, pharmaceutical companies, and PHR systems.

Reliability is the ability to repeat a case study with the same results [130]. The analysts’ familiarity with mining policy documents for requirements artifacts and knowledge of healthcare legal domain is a threat to the reliability of our case study [22, 24]. To mitigate this threat, we rigorously followed our case study protocol and our methodology by strictly adhering to the heuristics [18, 20, 30, 85, 93, 95, 98, 130, 132, 134]. The empirical study, discussed in Chapter 6, examines that ability of others to effectively apply our methodology.
4.4 Summary

Our formative study revealed that policy documents and DUAs both contain commitments, privileges, and rights. We can extract these CPRs using natural language patterns and then operationalize them as requirements. Our summative study showed that CPR analysis can be applied to documents outside of those used to develop the approach.

The percentage of items that were classifiable and unclassifiable for each case and study combination along with each case overall are shown in Figure 4.21. The figure shows that a higher percentage of the items in policy documents were classifiable, or translated into requirements, than items in the DUAs. The presence of classifiable items shows that policy documents and DUAs contain compliance requirements and contractual compliance requirements, respectively. For compliance purposes, these requirements should be implemented in systems and organizational practices.

---

Figure 4.21: Multi-Case Studies - Types of Items

The percentage of classifications for each case and study combination along with each case overall are shown in Figure 4.22. A higher percentage of the classifiable items in the DUAs expressed pledges as compared to the classifiable items in the policy documents. DUAs are signed by both parties and either create or extend a contractual relationship. Most of the classifiable items (59%) in the DUAs expressed recipient commitments, relating to the requirements of the recipient’s system. In contrast, policy documents describe an organization’s information practices for collecting, storing, and using users’ personal information as well as what the

---

In Figures 4.21-4.24, the acronyms are as follows: policy document case in formative study (PD-F), policy document case in summative study (PD-S), policy document case in formative and summative studies combined (PD-C), DUA case in formative study (DUA-F), DUA case in summative study (DUA-S), and DUA case in formative and summative studies combined (DUA-C).
user will and can do. The policy documents contain mostly commitments of the organization and privileges of the user, which are requirements for the organization’s system with which both actors interact. Given these results, requirements engineers may expect to have mostly requirements that express actions that the recipient must perform within the system or its practices in order to comply with the contractual compliance requirements of the DUAs. Based on the results for the policy documents, requirements engineers may expect to extract mostly requirements from the policy document that contain actions the organization must perform and actions the user is entitled to perform.

![Figure 4.22: Multi-Case Studies - Classifications](image)

The percentage of rationales within classifiable items for each case and study combination along with each case overall are shown in Figure 4.23. There exists a difference between the types of documents in terms of rationale. Whereas the classifiable items in the policy documents overwhelmingly have internal rationales, the classifiable items in the DUAs are overall more evenly distributed between internal and external rationale. The fact that a higher percentage of the DUAs’ classifiable items had external rationale, as compared to the policy documents’ classifiable items with external rationale, may be due to the fact that regulations adopted pursuant to the HIPAA specify that a DUA must exist when certain protected health information is used or disclosed.

Because policy documents focus primarily on internal rationale rather than external rationale, legal-based approaches do not provide sufficient coverage of compliance requirements. Recall that policy documents contain items that emphasize the commitments that organizations make to their users rather than rights and obligations as found in legal texts. Therefore, legal-based approaches [36] do not provide sufficient coverage of compliance requirements. Our analysis...
revealed that 92% of the classifiable items have internal rationale, whereas only 8% of the classifiable items have external rationale. Thus, within these policy documents, most of the items expressed organizational practices or procedures rather than items that are required or entitled by legal texts. This emphasis on internal actions within policy documents supports the previously mentioned need to analyze policy documents using different approaches, such as CPR analysis, to ensure proper coverage of the compliance requirements in the systems governed by these documents.

Requirements from DUAs are legally binding and thus critical to ensuring compliance. The motivation for compliance with DUAs comes from both the HIPAA and the contractual relationship codified in a DUA. When a covered entity wishes to share an LDS, the HIPAA regulations require that a DUA exist to govern the information exchange (see Section 1.1). Furthermore, breaching a DUA may expose an organization to both regulatory sanctions and legal liability for breach of contract. The situation is different for assessing compliance for policy documents. In the United States, the FTC Act tasks the FTC with protecting users from “unfair or deceptive acts or practices” [70]. The FTC can hold organizations accountable for their policy documents’ items regarding the collection, storage, and use of users’ personal information [86]. There is, however, no contract formed between an organization and its users through the existence of a policy document.

The percentage of actors within classifiable items for each case and study combination along with each case overall are shown in Figure 4.24. Policy documents and DUAs contain different sets of actors. Policy documents focus on two main actors: (1) the organization and (2) the user, where the user interacts with the organization’s system or website. In contrast, DUAs primarily govern two actors: the covered entity (data custodian) and the limited data set recipient. The covered entity shares the LDS with the recipient. DUAs are more targeted
than policy documents as evidenced by the fact that the actors are explicitly listed in the DUA and sign the DUA, whereas in policy documents the users are sometimes implied rather than made explicit. DUAs contain requirements for each of the actors; these requirements may be implemented in two separate software systems. In contrast, requirements from policy documents may be implemented in a single software system by the organization.

Figure 4.24: Multi-Case Studies - Actors
Chapter 5

Examining CPR Analysis for Ensuring Compliance

The truthiness will set you free!
- Stephen Colbert

This chapter discusses a study in which we seek to determine the suitability of CPR analysis for ensuring compliance. The study compares CPR analysis and a goal-based approach by applying each approach to nine policy documents, representing four healthcare organizations from four distinct healthcare sectors: health insurance, online drugstore, pharmaceuticals, and personal health records. The goal-based approach was used for this study as it was the only approach with policy document with which we could compare CPR analysis.

In goal-based analysis, requirements engineers extract goals from a variety of requirements source artifacts, including legal texts and policy documents \([20, 22, 24]\). Goals reflect high-level enterprise objectives or short-term objectives within a system. Goal-based analysis was initially developed as a means to ensure better coverage of software requirements than more traditional requirements engineering approaches. However, previous goal-based studies have not examined goal-based analysis’s suitability for ensuring compliance. We compare the resulting requirements artifacts—CPRs and goals—for compliance based on characteristics that are important for ensuring that requirements comply with governing policy documents. These characteristics are the ability to preserve traceability as well as context and the intended meaning of statements in a policy document. In terms of compliance, preserving traceability by linking the policy documents to compliance requirements—requirements that must be met in order for an organization to comply with its policy documents—helps to demonstrate due diligence by showing where specific policies are operationalized in corresponding software \([33, 94, 109]\). Herein, context refers
to the specificity and language of a given policy statement for the purpose of determining
the statement’s intended meaning. We address the following research question using these characteristics:

\[ RQ_1: \text{Which approach, CPR analysis or goal-based analysis, is better at ensuring compliance?} \]

The remainder of this chapter is structured as follows: Section 5.1 describes our study methodology. Section 5.2 describes characteristics that are important for ensuring compliance between requirements and governing policies. Section 5.3 provides a comparison of the two analysis approaches based on the characteristics discussed in Section 5.2. Section 5.4 discusses threats to validity. Section 5.5 provides a discussion of the comparison.

## 5.1 Methodology

We analyzed a total of nine policy documents from four organizations representing four healthcare sectors: health insurance, online drugstores, pharmaceuticals, and personal health records (PHRs). These organizations were chosen because they had been successfully used in previous studies, including our multi-case studies (Chapter 4), to identify requirements artifacts [24, 132, 134].

The documents analyzed for this study are listed in Figure 5.1. The only “Notice of Privacy Practices” document we analyzed for Aetna was for Student Health even though there are seven different Aetna “Notice of Privacy Practices” documents for the different healthcare benefits plans. Because these documents were identical with the exception of the plan name, we only analyzed one representative document. Aetna also had two “Notice of Information Practices” documents; however, these documents were structured differently and contained different information. Thus, we analyzed both of these documents.

For this study, we examined the results of the two requirements analysis approaches—CPR analysis and goal-based analysis. A CPR analysis expert (the author) performed the CPR analysis. Two goal-based analysis experts\(^1\) performed the goal-based analysis. Both approaches produce artifacts—CPRs and goals—that can be operationalized as requirements. For CPR analysis, we used the methodology and heuristics discussed in Chapters 2 and 3, respectively. We now give an overview of the goal-based methodology and discuss the differences between the two approaches.

\(^1\)Drs. Antón and Earp were the goal-based analysis experts.
### 5.1.1 Goal-based Analysis

As previously mentioned, goal-based analysis is a content analysis technique in which pre-requirements goals are extracted from post-requirements text artifacts [20, 22, 24]. In goal-based analysis, requirements engineers identify, document, and annotate strategic and tactical goals by exploring privacy policies. Strategic goals are those that reflect high-level enterprise goals, whereas tactical goals involve short-term goal achievements [18, 19]. The analysts organize the goals based on a taxonomy of privacy protections and vulnerabilities. Then the analysts elaborate the goals. During the goal refinement step, the analysts remove synonymous and redundant goals, resolve inconsistencies, and operationalize goals into requirements. It is important to note that the goals were never operationalized into software requirements and thus not fully elaborated.

Antón and Earp codified domain specific heuristics for applying Goal-Based Requirements Analysis Method (GBRAM) for goal-mining policy documents. The following is a heuristic for identifying goals [20]:

To identify goals, each statement in a privacy policy is analyzed by asking, “What goal does this statement or fragment exemplify?” and/or “What goal does this statement obstruct or thwart?” The identified goals are worded to express a state that is true, or the condition that holds true, when the goal is realized.

The following is a heuristic for classifying goals [20]:

---

<table>
<thead>
<tr>
<th>Company</th>
<th>Policy Documents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aetna (Health Insurance)</td>
<td>Notice of Information Practices - Large Case Pension</td>
</tr>
<tr>
<td></td>
<td>Notice of Information Practices - Life &amp; Disability</td>
</tr>
<tr>
<td></td>
<td>Notice of Privacy Practices - Student Health</td>
</tr>
<tr>
<td></td>
<td>Web Privacy Statement</td>
</tr>
<tr>
<td>DestinationRx (Drugstore)</td>
<td>Privacy Policy</td>
</tr>
<tr>
<td></td>
<td>Terms of Use</td>
</tr>
<tr>
<td>GlaxoSmithKline (Pharmaceutical)</td>
<td>Internet Privacy Statement</td>
</tr>
<tr>
<td></td>
<td>Terms and Conditions</td>
</tr>
<tr>
<td>Dossia (PHR)</td>
<td>Privacy Statement</td>
</tr>
</tbody>
</table>

*Figure 5.1: Analyzed Policy Documents*
Classification of goals involves differentiating goals according to goal class (e.g., protection vs. vulnerability) and subject matter. Protection goals are classified by analyzing each goal and asking, “Does this goal protect one’s private information?” Whereas, vulnerabilities are classified by considering each goal and asking “Does this goal potentially compromise the privacy and/or security of one’s private information?”

The following is a goal refinement heuristic [20]:

Organization of goals entails eliminating redundancies and reconciling synonymous goals. Goals are considered synonymous if their intended states are equivalent or if they mean the same thing to different stakeholders who simply express the goal using different terminology. It is up to the analyst to identify these instances.

The objective of goal-based analysis is to extract goals from policy documents in such a manner that they may be reused via the use of a goal repository. These goals are codified in a web-based Privacy Goal Management Tool (PGMT) developed at North Carolina State University [90]. The PGMT assists analysts in the goal-mining, reconciliation, and management processes [22]. The tool stores several values for each goal—ID, actor, keyword, description, subject, taxonomy, and goal creator. The actor element within the goals is one of the following: affiliate, customer, institution, third party advertiser, or unclassified. The keyword element is one of 59 keywords. Antón et al. originally identified 57 keywords that are commonly found in Internet privacy policies [22]; later research efforts resulted in two additional keywords—GUARD and PERMIT—being added.

5.1.2 Differences Between Analysis Approaches

CPR analysis was specifically developed to ensure compliance between software requirements and statements in policy documents and DUAs. In contrast, the goal-based analysis is broadly applicable to many different kinds of source materials, including existing requirements and design specifications, stakeholder interview transcripts, and legacy code. In addition, the two approaches differ with respect to their objectives. The primary objective of CPR analysis is to ensure compliance, whereas goal-based analysis within the scope of policy analysis has two objectives—trying to ensure requirements coverage of security and privacy considerations and developing a reusable corpus of generalizable privacy- and security-related software requirements. It was never intended to support compliance checking per se. The objectives of generalizability and broad coverage are somewhat at odds with the objective of strict compliance. The latter, as discussed in this chapter, requires stricter adherence to language and specificity because compliance is the primary objective.
The two approaches also produce different artifacts. CPR analysis ultimately produces a set of requirements, whereas goal-based analysis produces a set of goals and associated goal occurrences, where a goal occurrence helps maintain traceability to the source policy statement. In order to compare artifacts that are pre-requirements, we will be comparing the CPRs, which are the output of Step 2 of CPR analysis (Section 2.2.2), and goals, which are the output of goal-based analysis. Both artifacts are represented using structured natural language. The two approaches also differ based on the naming of the parties in the artifacts. In CPR analysis, the two primary parties are the *organization* and the *user*, whereas in goal-based analysis, the two primary parties are the *institution* and the *customer*.

### 5.2 Preserving Context and Intended Meaning

To address *RQ*$_1$, we compare CPR analysis and goal-based analysis by examining characteristics that are important for ensuring compliance with governing policy documents. Herein, we examine the ability to preserve context and intended meaning. For requirements to accurately reflect what is stated in governing policy documents, we must preserve language, specificity, and modality. The *language* of policy documents refers to the terminology or wording. The *specificity* of the policy documents refers to the degree of detail. The *modality* of the actions contained in policy documents expresses whether the action is pledged or entitled.

As a generally understood principle of law, it is essential to preserve the precise language of the policy (e.g., [106, 117]). When courts seek to interpret a statute, contract, will, or other legal document, they invariably begin with the plain language of the original document. As a result, preserving the original language is an advantage because courts will not give equal weight to subsequent interpretations of the original document. There is a clear risk of distorting the original meaning with every transformation or interpretation of the original text. This possibility for distortion is illustrated by the telephone game, where a message is passed from person to person; typically by the time the message has reached the final person, the message has changed dramatically. Otto and Antón note the paramount importance of consistent definitions and terminology for regulatory compliance [109]. By preserving the original language of the policy document, any subsequent interpretation can be verified by legal experts.

Policy statements contain varying specificities, or varying degrees of details, regarding actions, purposes, objects, and conditions. Some policy statements are general [61]:

*We protect the privacy of your health information.*

Other statements are extremely detailed [60]:

*Unless you explicitly and specifically consent, Dossia will not disclose your health information or contact information to third parties for them to use for marketing purposes.*
When extracting requirements from policy documents for compliance purposes, we need to preserve the same specificity—neither more general nor more specific than the policy document—for the same reasons that we must preserve the document’s original language.

Each action, or capability, in a policy document has a modality—pledged or entitled—that we preserve. An action’s modality determines whether it should be required or allowed by the system.

5.3 Case Study Results

This section employs actual examples taken from our CPR analysis and goal-based analysis to illustrate whether the approaches ensure compliance by preserving the language, specificity, and modality in the original policy statements.

5.3.1 Language

Preserving the original language in a policy document ensures compliance because legal experts can verify the interpretations made by requirements engineers. It is important to maintain the original language in the policy document so that the intended meaning of the document can be preserved. CPR analysis preserves the language of policy statements by employing the terminology, or wording, of the policy statements within the associated commitments, privileges, and rights. Given that the objectives of goal-based analysis were generalizability and reuse, goals do not always preserve the language of the associated policy statements.

Maintaining the language preserves the meaning

Example 1 shows a commitment and a goal that both preserve the original language of a statement from the Dossia Privacy Statement (DPS). CPR analysis resulted in Commitment_{DPS40}^{2} (comply with applicable laws) because the action was pledged by the organization, which was determined by the use of CH_{PD}^{1} and the modal verb must. Goal-based analysis yielded Goal_{1947}^{3} (COMPLY with applicable laws), which is identical to the extracted commitment. Both the commitment and the goal are expressed using the same language as the policy statement in order to ensure compliance. The commitment and goal refer to Dossia, the actor within the policy statement, using different terms—organization and institution, respectively.

---

2^The document code and number are the ID for the CPR within the given document.
3^The number is the goal ID from the PGMT.
Dossia Privacy Statement: Although Dossia is structured to give participants control over their PCHRs, Dossia must also comply with applicable laws.

**Commitment**_{DPS40} (organization) comply with applicable laws

**Goal**_{1947} (institution) COMPLY with applicable laws

---

**Example 1:** Commitment and goal that maintain the language of policy statement

---

**A change in language can lead to a change in meaning**

Example 2 shows Commitment_{DPS61} that preserves the meaning of a policy statement by maintaining the original language of the policy statement. In contrast, the corresponding goal (Goal_{102}) does not preserve the meaning because it does not preserve the original language and wording of the policy statement. The policy statement from the Dossia Privacy Statement states that the organization sends an email in order to notify the user of changes to the policy.

CPR analysis yielded Commitment_{DPS61} because the statement expresses a pledge that the organization is making to its users. This statement is classified as a commitment using CH_{PD} 1: the presence of the modal verb will and the organization as the actor. Commitment_{DPS61} conveys that the organization pledges to send an e-mail about the changes; thus, preserving the language and intended meaning of the policy statement and ensuring compliance.

In contrast, goal-analysis yielded Goal_{102}, which does not preserve the original language and intended meaning. Goal-based analysis explicitly tries to eliminate design details because it favors delaying design decisions until software design rather than during requirements engineering. Thus, the goal does not explicitly preserve how the notification will be delivered. However, if the organization does not notify users about the changes via email, they would not be in compliance with the policy. This noncompliance would likely lead to FTC action against the organization, as it could be considered a deceptive practice.

---

Dossia Privacy Statement: If, however, Dossia makes changes to the Privacy Statement that would materially affect your protections or choices explained here, Dossia will, at least 30 days in advance, take the following steps: Send an e-mail about the change to your e-mail address in our records.

**Commitment**_{DPS61} (organization) send an e-mail about the changes to the user’s e-mail address in organization’s records, at least 30 days in advance of changes if organization makes changes to the Privacy Statement that would materially affect the user’s protections or choices explained within privacy statement

**Goal**_{102} (institution) NOTIFY customer of change to PP

---

**Example 2:** Commitment that maintains language and conditions; goal that does not...
5.3.2 Specificity

Specificity is important for ensuring compliance because systems that are governed by policies should be as specific or detailed as the policies—neither more general nor more specific than the policy document—for the same reasons that preserving the original language is essential (e.g., [106, 117]). CPR analysis preserves the specificity of policy statements by documenting statements’ attributes. These attributes maintain the same level of detail as the policy statements. Given that the objectives of goal-based analysis were generalizability and reuse, goals do not always preserve the specificity of the associated policy statements.

Maintaining the specificity of purposes

Organizations often explicitly state reasons for performing actions; we refer to these reasons as purposes. We must preserve the purposes in the requirements artifacts in order to maintain the context of the policy statements. When purposes are given for an action, the action should be performed only for those purposes.

Consider the policy statement in Example 3, which contains two purposes: (1) create a data connection between the user’s computer and the organization’s servers, and (2) maintain a data connection between the user’s computer and the organization’s servers. These purposes constrain the reasons why the organization collects and stores web usage information.

Dossia Privacy Statement 3: Dossia may also collect and store web usage information, including IP address and related information, needed to create and maintain a data connection between your computer and Dossia servers.

Privilege $DPS_{7a}$: (organization) collect web usage information in order to create a data connection between the user’s computer and the organization’s servers

Privilege $DPS_{7b}$: (organization) collect web usage information in order to maintain a data connection between the user’s computer and the organization’s servers

Privilege $DPS_{7c}$: (organization) store web usage information in order to create a data connection between the user’s computer and the organization’s servers

Privilege $DPS_{7d}$: (organization) store web usage information in order to maintain a data connection between the user’s computer and the organization’s servers

Goal $G_{26}$: (institution) COLLECT site usage information

Goal $G_{1938}$: (institution) STORE site usage information

Example 3: Privileges that maintain the specificity of purposes and goals that do not

CPR analysis yielded four privileges based on the presence of two actions (HH 1) and two purposes (DAH 8), where a privilege exists for each possible action-purpose combination. The items were classified using CH$_{PD}$ 1 and CH 1 because the actions were independent. These four
privileges preserve the specificity of the original policy statement, including the purposes, and, thus, ensure compliance.

Goal-based analysis yielded two goals—Goal26 and Goal1938. These two goals omit the purposes for which the actions (collect and store) occur. The multiple actions of the policy statement are addressed by having multiple goals that preserve the two actions through the keywords COLLECT (Goal26) and STORE (Goal1938). However, neither goal specifies the purpose for the action. These purposes are needed to preserve the specificity of this policy statement. If the organization collects or stores web usage information for any reason other than to create or maintain a connection, it would be considered noncompliant and invite FTC action.

**Maintaining the specificity of objects**

We preserve the objects of the actions in the requirements artifacts in order to maintain the policy statements’ context. Consider the policy statement in Example 4 that contains the following objects: administrative, physical, and technical security technologies and internal controls. To ensure compliance, we must include all of the objects in the requirements artifacts.

CPR analysis yielded CommitmentDPS32 through the use of CHPD 2. The objects of the statement are documented as administrative, physical, and technical security technologies and internal controls. These maintain the specificity of the policy statement’s objects.

Goal-based analysis yielded Goal1175, which expresses the following objects: security measures (technical, contractual, administrative steps). These objects are misaligned with the objects of the policy statement. Goal1175 includes a new security measure—contractual—and fails to include physical security technologies and internal controls. Because the goal does not preserve the specificity of the object, it does not ensure compliance. Moreover, Goal1175 fails to preserve the purposes for using the technologies and controls. The policy statement has the purpose of protect user’s information from unauthorized access, use, and disclosure, whereas Goal1175 had the purpose of protect against loss/misuse/alteration of information under institutions control. These purposes do not align, resulting in noncompliance; for example, preventing unauthorized disclosure is not a purpose in Goal1175, but it is a permissible purpose pursuant to the policy statement.

**Maintaining the specificity of conditions**

Organizations often express conditions or restrictions for when actions will be performed in their policies. These conditions are important for preserving the specificity of policy statements. When a condition exists for an action, the action should be performed if and only if the condition is met.

Consider Example 2, above, in which the organization pledges to send an email about changes
Dossia Privacy Statement: Dossia uses administrative, physical, and technical security technologies and internal controls, including encryption of health information, to protect your information from unauthorized access, use, and disclosure.

Commitment_DPS32: (organization) use administrative, physical, and technical security technologies and internal controls in order to protect user’s information from unauthorized access, use, and disclosure.

Goal_1175: (institution) USE security measures (technical, contractual, administrative steps) to protect against loss/misuse/alteration of information under institutions control.

Example 4: Commitment that maintains the specificity of object and goal that does not.

5.3.3 Modality

Maintaining modality is important for ensuring compliance because it preserves whether an action is pledged or entitled. An action’s modality determines whether it should be required or allowed within the system.

CPR analysis preserves the modality through its classifications—commitments, privileges, and rights. These classifications determine how CPRs are operationalized into requirements. There were 191 pledges (commitments) and 141 entitlements (privileges or rights) extracted from the nine analyzed document.

In a study of goals from Internet privacy policies in the finance and healthcare domains, Breaux found that the goals did not preserve the modality of the policy statements [30]. Examples 1-4 show that the goal-based approach did not preserve modality, leaving ambiguity as to whether an action is pledged or entitled. However, we did observe that the keywords ALLOW and RESERVE can be used to preserve the entitlement modality. Consider the goals in Example 5, they preserve the modality with ALLOW user to. In this example, the goals also preserve the language and specificity of the policy statement. Our CPR analysis similarly preserves the modality because privileges are entitlements. The policy statement was split into multiple items.
using HH 1; the items were classified as privileges using CH$_{PD}$ 3 and CH 1.

| Dossia Privacy Statement$_5$: You can choose to directly enter your own health information, authorize your health care entities to submit your information electronically, or import information from independent websites. |
| Privilege$_{DPS9a}$: (user) enter user’s health information |
| Privilege$_{DPS9b}$: (user) authorize user’s health care entities to submit user’s health information electronically |
| Privilege$_{DPS9c}$: (user) import health information from independent websites |
| Goal$_{1935}$: (Institution) ALLOW user to manually enter their own health information into your PCHR |
| Goal$_{1936}$: (Institution) ALLOW user to authorize other individuals/entities to enter/send PHI in/to PCHR |
| Goal$_{1954}$: (Institution) ALLOW user to import information from another website |

**Example 5:** Privileges and goals that maintain modality of policy statement

Consider Goal$_{2254}$ in Example 6. This goal preserves the original entitlement with the phrase *RESERVE the right*. It also preserves the original language, specificity, and modality. The privilege that results from CPR analysis similarly preserves the modality of the statement along with the language and specificity. The policy statement was classified as a privilege using CH$_{PD}$ 1 and CH 1.

| Dossia Privacy Statement$_6$: If Dossia were to transfer assets or operations in connection with a merger, sale, bankruptcy, or other transaction, Dossia might transfer PCHR information to the acquiring or merging entity. |
| Privilege$_{DPS4}$: (organization) transfer PCHR information to acquiring or merging entity, if organization were to transfer assets or operations in connection with a merger, sale, bankruptcy, or other transaction |
| Goal$_{2254}$: (institution) RESERVE the right to transfer assets (including PCHR info) or operations in connection with a merger, sale, bankruptcy, or other transaction to the acquiring or merging entity |

**Example 6:** Privileges and goals that maintain modality of policy statement

### 5.4 Threats to Validity

We examine the validity of our case study based upon Yin’s validity tests, which include construct validity, internal validity, external validity, and reliability [130]. We maintain construct
validity by using multiple sources for analysis (nine policy documents from four organizations), establishing a chain of evidence, and having others review our report. Because we make inferences about which approach performs better at ensuring compliance as a result of our case study, internal validity is a threat to this study. To address this threat, the analysts performing each approach were experts in the respective, given approach; thus, the differences in the results were based on the difference in the approaches, not the differences in the analysts’ knowledge. There is a possible threat to external validity as well, because we have analyzed policy documents solely from the healthcare domain. Another possible threat to validity is the use of the PGMT. Because the PGMT is a repository of goals, the goals that can be expressed may be limited by what is in the repository, but the analysts did create new goals whenever a synonymous goal did not exist. Reliability refers to repeating the study with the same result; we maintain a strict process and instructions within the study.

A limitation of both approaches is that they are both manual processes. Although goal-based analysis uses the PGMT to promote goal reuse, the process of extracting goals from policy documents is still manual. A limitation of CPR analysis is that the approach has been applied to policy documents only within the healthcare domain. Goal-based analysis, however, has been applied to policy documents from the e-commerce, financial, and healthcare domains. A limitation of our comparison of the two approaches is that the comparison only includes nine policy documents from a single domain—healthcare.

5.5 Discussion

In this case study, we observe that CPR analysis ensures compliance more directly than goal-based analysis by maintaining language, specificity, and modality. Another important way to support compliance is via traceability. To establish software compliance with all governing legal texts and policies, legal traceability from these documents to requirements specifications is essential [33, 109]. Legal traceability helps to demonstrate due diligence in a court of law [94]. Although not the major focus of this chapter, we note that both approaches preserve traceability by linking requirements artifacts to the policy documents and policy statements.

CPR analysis consistently preserves the original policy statement language. In contrast, goal-based analysis does not always preserve the original policy statement language, which may lead to interpretations that stray from the original intended meaning. Goal keywords do not consistently maintain the language of the action. Although CPRs may be also considered slight transformations of the policy statement, they nonetheless preserve the original language and thus facilitate legal experts verifying the representation.

Preserving the specificity of the original policy statements—neither more general nor more specific than the policy document—is important for compliance. Any compliance requirements
analysis approach must thus include everything contained in the policy document while not inferring artifacts that are not strictly derived from the policy document. CPR analysis preserves the specificity of the policy statements by being as general or as specific as the statements. Goal-based analysis artifacts often fail to maintain specificity and may include synthesized details beyond what is stated in the original policy statement. Because goal-based analysis generalizes policy statements in an attempt to be reusable, the purposes and conditions are sometimes not preserved by the goals.

Preserving an action’s modality—whether an action is pledged or entitled—aids in ensuring compliance. CPR analysis preserves modality via the specification of commitments, privileges, and rights. More often than not, goals do not preserve modality. Consequently, whether a goal expresses desired or required achievement remains unclear unless the keywords ALLOW and RESERVE are used to preserve the modality for entitlements.

Recall that CPR analysis and goal-based analysis were developed with different objectives. The FTC Act [70] and the need for organizations’ systems to comply with their policies motivated the development of CPR analysis. Designed with compliance in mind, CPR analysis preserves language, specificity, and modality. In contrast, the goal-based approach was designed to ensure broad requirements coverage via generalizability and goal reuse; these objectives are at odds with the need to ensure compliance. On the other hand, CPRs are institution-specific and thus not reusable unless different institutions’ policy statements are identical—an unlikely scenario, given our prior experience.

The goal-based analysis effort yielded goals with no corresponding CPRs in the CPR analysis. Generally speaking, these goals related to warnings, disclaimers, and clarifications—items that remain unclassified with CPR analysis. Therefore, the goal-based approach is advantageous when it is important to capture the warnings and disclaimers expressed in a policy statement. This is consistent with results of prior studies, in which goal-based analysis was especially well suited for identifying messaging requirements [18].

Although goal-based analysis does not always preserve language, specificity, and modality, organizations can use goal-based analysis to gain an understanding of high-level objectives that software must satisfy. From the perspective of privacy and security requirements, the taxonomy of privacy protections and vulnerabilities [20] does aid in uncovering commonly overlooked requirements by forcing analysts to mitigate those risks introduced by vulnerabilities and exceptions that can compromise a system.
In this chapter, we discuss an empirical user study to measure and compare the ability of requirements engineers to effectively extract compliance requirements—requirements that must be met in order for an organization to comply with its policies—from a policy document using one of three analysis approaches—CPR analysis, goal-based analysis, and non-method-assisted analysis. As discussed throughout this dissertation, CPR analysis employs commitments, privileges, and rights to extract compliance requirements from policy documents, where commitments are pledges and privileges and rights are entitlements. Goal-based analysis identifies and refines goals, which reflect high-level enterprise objectives or short-term objectives within a system. Non-method-assisted analysis serves as the control condition within the study and allows subjects to use any analysis method of their choosing.

Our study shows that requirements engineers can benefit from applying CPR analysis rather than goal-based analysis or non-method-assisted analysis to produce compliance requirements. The CPR subjects produced a higher median number of expected compliance requirements—requirements derived from the policy by the experimenter (the author) and two other requirements engineers against which we compare the subject-produced requirements. The requirements produced by the CPR subjects had better correctness and completeness with respect to expected compliance requirements than those requirements produced by goal-based and control subjects. Although the goal-based and control subjects did not produce as many expected compliance requirements, they did produce more synthesized requirements—requirements not directly derived from the policy but identified through inquiry [18]. These synthesized requirements were not compliance requirements but instead discussed aspects of the system not described in
the policy. This suggests that CPR analysis is more focused on promoting compliance whereas other analysis methods may provide better coverage of system requirements overall without concern for compliance.

The remainder of this chapter is as follows: Section 6.1 discusses the experimental design of the study; Section 6.2 briefly discusses the pilot study; Section 6.3 contains data analysis for the study; Section 6.4 discusses threats to validity; and Section 6.5 provides a discussion of the study in terms of requirements engineering.

6.1 Experimental Design

Our experimental design enabled us to compare CPR analysis to goal-based (GB) analysis and non-method-assisted analysis approach, which served as our control, to determine the ability of requirements engineers to effectively extract compliance requirements from policy documents. In this section, we discuss the design, subjects, materials, procedure, measurements for the study, and analysis tests.

6.1.1 Design

In our experimental design the subjects were divided into three groups, each of which was asked to use one of the following analysis methods (conditions): CPR analysis, goal-based analysis [17, 18], and an analysis method of their choosing (control). The independent measures were the three condition groups. The subjects in each group performed two related analysis tasks—a problem analysis task and a requirements analysis task—for the given condition. Each of the three groups was given an analogous set of documents, containing instructions for both tasks but differing in the assigned analysis method. All three groups were given the exact same policy document to analyze.

6.1.2 Subjects

The subjects in the study were students in the graduate software engineering course (CSC 510) at North Carolina State University (NCSU) during the Fall 2011 semester. The CSC 510 students had the opportunity to volunteer to participate in the study and were compensated with extra credit points on their final project for the course if they completed both analysis tasks. The subjects understood (a) the purpose of the study and (b) that their identities would remain confidential. In compliance with NCSU Institutional Review Board rules, the students were given the option to complete a pair of alternative tasks; however, no students opted to complete the alternative tasks.
Thirty-four subjects completed both analysis tasks for the study. Three additional subjects began the study but did not complete both tasks; these results were not included in our analysis. In order to ensure that experts were not overrepresented in any one condition or group, the subjects were assigned to the three groups in a balanced fashion based on the subjects’ background and software engineering experience—modeling, database modeling, and requirements engineering.

The demographics for the subjects that we considered when assigning the subjects to groups are shown in Table 6.1. The table includes the total number of subjects who completed the study; the number of subjects whose major is computer science; the number of subjects pursuing B.S., M.S., M.C.S, and Ph.D. degrees; median age of the subjects; and the ratio of females and males. There were twelve subjects in the CPR condition group and eleven subjects in the goal-based and control condition groups, respectively. The majority of the subjects (30) were computer science majors. Most of the subjects (24) were pursuing Masters degree. The youngest subject was 19 years old, and the oldest subject was 49 years old. The median ages for the CPR, goal-based, and control subjects were 23 years old, 23 years old, and 22 years old, respectively. The majority of the subjects (23) were male.

<table>
<thead>
<tr>
<th>Number of Subjects</th>
<th>CPR</th>
<th>GB</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer Science</td>
<td>12</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Pursuing B.S.</td>
<td>11</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>Pursuing M.S.</td>
<td>11</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>Pursuing M.C.S.</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Pursuing Ph.D.</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Median Age (Years)</td>
<td>23</td>
<td>23</td>
<td>22</td>
</tr>
<tr>
<td>Female/Male</td>
<td>4/8</td>
<td>5/6</td>
<td>2/9</td>
</tr>
</tbody>
</table>

The software engineering experience of the subjects by condition is presented in Table 6.2. When assigning subjects to condition groups we attempted to have the groups as balanced as possible. The majority of the subjects (24) had taken another software engineering course, but no subjects had taken a requirements engineering course. Two subjects had taken an information/telecom law course. Sixteen of the thirty-four subjects had database modeling experience within a course. Ten subjects had industry experience with database modeling, and seven subjects had industry experience with modeling. Eight subjects had industry experience with requirements engineering.
Table 6.2: Subject Software Engineering Experience

<table>
<thead>
<tr>
<th>Experience</th>
<th>CPR</th>
<th>GB</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taken another (other than CSC 510) software engineering course</td>
<td>9</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Taken a requirements engineering course</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Taken a information/telecom law course</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Experience using database modeling within a course</td>
<td>5</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>Experience using database modeling within industry</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Experience using modeling within industry</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Requirements engineering experience within industry</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

6.1.3 Materials

We provided the subjects with the following materials for the study: an informed consent form, a questionnaire\(^1\) about subjects’ background and software engineering experience, as well as two sets of detailed instructions for the tasks. The first instruction set describes the problem analysis task (Part 1). These instructions describe the problem, present a step-by-step overview of the assigned analysis method, and provide a detailed example of how to construct a model of a policy for each of the respective methods. The second instruction set describes the requirements analysis task (Part 2). These instructions describe how to translate the results of the problem analysis task into software requirements and provide guidance about how to write a requirements document. Both instruction sets also included a checklist of artifacts that subjects need to submit for the given task and a set of open-ended questions\(^2\) about their experiences in completing the task.

Along with the instructions, the subjects were given worksheets to use for their completion of the problem analysis task. The CPR subjects were given a CPR-tailored Microsoft (MS) Excel worksheet that they were to use to capture the results of their CPR problem analysis efforts. The goal-based subjects were given a goal-based-tailored MS Excel worksheet that they were to use to capture the results of their goal-based problem analysis efforts. The control subjects were asked to submit the results of their problem analysis in an MS Word document because they could choose the approach they wanted to use.

The policy document analyzed by all subjects during the problem analysis and requirements analysis tasks was included within the problem analysis instructions. It was a portion\(^3\) of a legitimate policy document from a popular social networking site, Facebook. To minimize bias due to name recognition, we replaced all references to Facebook with SocialNet. Our rationale for selecting the given portion of the Facebook privacy policy was twofold: (a) it contained a reasonable number of system requirements; and (b) it could be analyzed within a reasonable

\(^{1}\)The questionnaire can be seen in Figure B.1 on page 152.
\(^{2}\)The open-ended questions can be seen in Figures B.2 and B.3 on page 153.
\(^{3}\)The policy document can be seen in Figure B.4 on page 154.
amount of time. This portion of the policy discussed the information that the organization receives, including: information about the user; content; transactional information; site activity information; access device and browser information; and cookie information. An additional material that was instrumental for our study was the set of expected compliance requirements against which subject-produced requirements were compared.

### 6.1.4 Procedure

Prior to administering the study, we produced a set of expert-produced expected compliance requirements, derived from the same SocialNet policy that the subjects were given, to which we compare the subject-produced requirements. For the study, three requirements engineers—the experimenter (the author) and two other requirements engineers, who are experienced with extracting requirements from documents—produced the set of expert-produced expected compliance requirements, which contained seventy-three requirements. Using Wideband Delphi [123], we came to a consensus on the expected set of compliance requirements. First, each expert was given the SocialNet policy. Second, each expert produced a set of requirements from the policy. The first author analyzed the policy three times—one for each condition using the problem and requirement analysis instructions—and combined the results. The other two experts analyzed the policy using primarily goal-based approaches. Third, we met for two hours to discuss our sets of requirements. Because we were not able to make it through all of the requirements in two hours, we also met for an additional one and half hours to complete our consensus for the expected set of compliance requirements. Antón and Breaux employed a similar approach in which they relied on expected sets in their empirical user studies in which subjects apply requirements engineering approaches [18, 30].

The experimenter introduced the study and analysis tasks to the subjects in class. Subjects were asked to work independently and not consult others about the study. The experimenter requested that the subjects submit the results of their analysis tasks (Part 1 and Part 2) online via Moodle, a course management system. The two tasks were self-paced with separate deadlines. For the study, the subjects had one week to complete each of the task. The subjects received reminders two days prior to each deadline.

During Part 1, the subjects performed problem analysis on the SocialNet policy document based on their respective conditions. CPR subjects performed the first two steps of CPR analysis—(1) parse and (2) classify—to produce a set of commitments, privileges, and rights. The subjects submitted their completed CPR analysis worksheet, which contained the commitments, privileges, and rights they extracted from the SocialNet policy document. The worksheet also included links between each commitment, privilege, or right and the policy statement from which it was extracted.
Goal-based subjects performed six steps: (1) identify goals, (2) identify agents and agent responsibilities, (3) eliminate redundancies, (4) analyze obstacles, (5) identify scenarios, and (6) construct goal hierarchy. The subjects submitted their completed goal-based analysis worksheet, three elaborated goal scenarios, and a goal hierarchy.

Control subjects used any modeling technique with which they were familiar to analyze the policy document; none of the control subjects used CPR analysis or goal-based analysis. Instead, most used a combination of approaches for their analysis: four subjects used definitions, three subjects used diagrams, two subjects used key functions, one subject used constraints, and one subject employed used cases. Definitions, diagrams, key functions, and constraints were all discussed as examples of possible approaches within the condition task instructions. The control subjects submitted a two-page statement of the properties based on modeling the policy document using their techniques of choice.

During Part 2, the subjects produced requirements specifications based on their artifacts from Part 1. The translation from the artifacts in Part 1 to the requirements depended on the subjects’ condition group. CPR subjects translated, or operationalized, the commitments, privileges, and rights into requirements using natural language templates. The goal-based subjects translated the goals into requirements based on how the goals should be accomplished by the system. The control subjects used any technique with which they were familiar to perform the translation. Along with requirements specification document, each subject submitted how they translated the artifacts from Part 1.

The artifacts that the subjects produced for each task based on their respective conditions are outlined in Table 6.3.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Problem Analysis Task (Part 1)</th>
<th>Requirements Analysis Task (Part 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPR</td>
<td>- CPR analysis worksheet</td>
<td>- Requirements Specification Document</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Updated CPR analysis worksheet</td>
</tr>
<tr>
<td>GB</td>
<td>- Goal-based analysis worksheet</td>
<td>- Requirements Specification Document</td>
</tr>
<tr>
<td></td>
<td>- Three elaborated scenarios</td>
<td>- Updated goal-based analysis worksheet</td>
</tr>
<tr>
<td></td>
<td>- Goal hierarchy</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>- Two page statement of the properties</td>
<td>- Requirements Specification Document</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Description of translation</td>
</tr>
</tbody>
</table>

6.1.5 Measurements

The primary measurements for this study are correctness and completeness, which test the performance of the subjects extracting compliance requirements from a policy document. Correctness
measures whether a subject’s requirements relate to the expected compliance requirements. Completeness measures whether a set of subject-produced requirements covers every requirement within the expected compliance requirements set.

First we compare the subject-produced requirements to the expert-produced expected compliance requirements in order to examine whether the subject found expected compliance requirements. Given a subject-produced requirement and an expected compliance requirement, we determine whether the subject-produced requirement relates to the expected compliance requirement by asking, “Does the subject-produced requirement contain the same action and specificity as the expected compliance requirement?” If the answer is yes, then we note that the subject-produced requirement and the expected compliance requirement are a related pair. For compliance purposes as discussed in Chapter 5, it is essential to preserve the same specificity as the policy document—neither more general nor more specific than the policy (e.g., [106, 117]).

Consider the following related pair, where both the expected compliance requirement (ER) and the subject-produced requirement (SPR) contain the same action (ask) and the same specificity:

\[
\text{ER}_1: \text{SocialNet shall be able to ask the user for additional information for security reasons.}\\
\text{SPR}_{647}: \text{SocialNet may ask the user additional information for security reasons.}
\]

**Codings for Related Pairs**

We examine whether a subject-produced requirement that is contained in a related pair has any of the following qualities—differing modality than the expected compliance requirement, differing point of view than the expected compliance requirement, and relating to multiple expected compliance requirements. Each quality corresponds to a coding, which we document for the related pair. Each related pair may have zero or more codings. These codings refer to unwanted qualities with the subject-produced requirement because, ideally, a subject-produced requirement would have the same modality and point of view as the expected compliance requirement and would be self-contained, only relating to a single expected compliance requirement.

**Modality**

We examine whether the subject-produced requirement in a related pair has the same modality—pledged or allowed—as the expected compliance requirement. Because we are concerned with compliance, a subject-produced requirement should ideally have the same modality as the expected compliance requirement to ensure that if an action is pledged it is required within the software system and if an action is entitled it is allowed within the software system. We determine this by asking, “Does the subject-produced requirement have the same modality as the expected compliance requirement?” If not, we note this with the \textit{Mod} coding. Consider the
following related pair in which the expected compliance requirement (ER_{27}) is pledged whereas the subject-produced requirement (SPR_{100}) is allowed:

\textbf{ER}_{27}: SocialNet shall share user’s photos with others.

\textbf{SPR}_{100}: A user shall be able to share photos

**Point of View**

We examine whether the subject-produced requirement in a related pair has the same point of view as the expected compliance requirement. A related pair has different points of view if the actions expressed in the two requirements are from the view of different actors. We note the difference in point of view because we want to ensure that the actor in the policy can act within the system in a manner consistent with the policy. We determine this by asking, “Does the subject-produced requirement have the same actor as the expected compliance requirement?” If not, we note this with the PoV coding. Consider the related pair above in which the expected compliance requirement (ER_{27}) is from the point of view of SocialNet whereas the subject-produced requirement (SPR_{100}) is from the point of view of the user.

**Relating to Multiple Expected Compliance Requirements**

We examine whether a subject-produced requirement relates to multiple expected compliance requirements. The expected compliance requirements were broken down to each include a single, testable item by avoiding the use of ‘and’ [29, 72, 121]. If the subject did not perform the same type of requirement splitting, then a single subject-produced requirement may be contained in related pairs with multiple expected compliance requirements. We determine this by asking, “Does the subject-produced requirement contain the same action and specificity as multiple expected compliance requirements?” If so, we note this with the Mul coding. Consider the following requirements, where the subject-produced requirement (SPR_{559}) relates to multiple expected compliance requirements (ER_{46} and ER_{47}):  

\textbf{ER}_{46}: SocialNet shall be able to retain details of the users transaction on SocialNet.

\textbf{ER}_{47}: SocialNet shall be able to retain details of the user’s payments.

\textbf{SPR}_{559}: SocialNet shall be able to retain the details of transactions or payments user makes on SocialNet.

**Possible Coding Combinations for Related Pair**

There are three possible codings—differing modality than the expected compliance requirement (Mod), differing point of view than the expected compliance requirement (PoV), and relating
to multiple expected compliance requirements \((Mul)\). Because a related pair may have zero or more codings, there are eight possible coding combinations. Table 6.4 provides each possible coding combination. The first column states the possible coding combinations for related pairs. The second column provides a description of each coding combination; these descriptions are based on the meaning of each coding. For example, if the related pair does not have any coding, then the subject-produced requirement contains the same action and specificity as the expected compliance requirement. These coding combinations allow us to examine the qualities of the subject-produced requirements with respect to the expected compliance requirements.

<table>
<thead>
<tr>
<th>Coding</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(none)</td>
<td>The subject-produced requirement contains the same action and specificity as the expected compliance requirement.</td>
</tr>
<tr>
<td>(Mod)</td>
<td>The subject-produced requirement contains the same action and specificity as the expected compliance requirement, but the subject-produced requirement has different modality than the expected compliance requirement.</td>
</tr>
<tr>
<td>(PoV)</td>
<td>The subject-produced requirement contains the same action and specificity as the expected compliance requirement, but the subject-produced requirement has different point of view than the expected compliance requirement.</td>
</tr>
<tr>
<td>(Mul)</td>
<td>The subject-produced requirement contains the same action and specificity as the expected compliance requirement, but the subject-produced requirement relates to additional expected compliance requirement(s).</td>
</tr>
<tr>
<td>(Mod/PoV)</td>
<td>The subject-produced requirement contains the same action and specificity as the expected compliance requirement, but the subject-produced requirement has different modality and different point of view than expected compliance requirement.</td>
</tr>
<tr>
<td>(Mod/Mul)</td>
<td>The subject-produced requirement contains the same action and specificity as the expected compliance requirement, but the subject-produced requirement has different modality than the expected compliance requirement and relates to additional expected compliance requirement(s).</td>
</tr>
<tr>
<td>(PoV/Mul)</td>
<td>The subject-produced requirement contains the same action and specificity as the expected compliance requirement, but the subject-produced requirement has different point of view than the expected compliance requirement and relates to additional expected compliance requirement(s).</td>
</tr>
<tr>
<td>(Mod/PoV/Mul)</td>
<td>The subject-produced requirement contains the same action and specificity as the expected compliance requirement, but the subject-produced requirement has different modality and different point of view than the expected compliance requirement and relates to additional expected compliance requirement(s).</td>
</tr>
</tbody>
</table>
Codings for Subject-Produced Requirement Not in a Related Pair

For a subject-produced requirement that is not contained in any related pairs, we examine the reason why it did not relate to any expected compliance requirements. Subject-produced requirements do not relate to expected compliance requirements if they do not contain the same action or the same specificity as the expected compliance requirements. Subject-produced requirements also do not relate to expected compliance requirements if the subject-produced requirements are not requirements.

Action Not in Policy

Because we are concerned with extracting compliance requirements, a subject-produced requirement will not relate to any expected compliance requirement if the subject-produced requirement is not derived from the policy. We determine this by asking, “Is the action of the subject-produced requirement contained within the policy?” If not, we note this with the NP coding.

Differing Specificity

Subject-produced requirements do not relate to expected compliance requirements if they do not have the same specificity for a particular action. For compliance purposes the subject-produced requirements should contain the same specificity as the expected compliance requirements and the policy; otherwise, the subject-produced requirement is more or less specific than the original policy. If the subject-produced requirement is more specific than the policy, then it adds constraints to the system. We determine this by asking, “Does the subject-produced requirement place additional constraints on the system in comparison to the policy?” If so, we note this with the MS coding. If the subject-produced requirement is less specific than the policy, then it is not policy-compliant because it is missing details. We determine this by asking, “Is the subject-produced requirement less specific than the policy?” If so, we note this with the LS coding.

Not a Requirement

Other subject-produced requirements do not relate to expected compliance requirements because they are not actual requirements. We determine this by asking, “Does the subject-produced requirement actually express a requirement?” If not, we note this with the NR coding.

Possible Codings for Subject-Produced Requirements Not in Related Pair

Each subject-produced requirement that does not relate to any expected compliance requirements will have one of the codings listed in Table 6.5. The first column contains the possible codings.
for subject-produced requirements. The second column contains the description of the subject-produced requirement with the given coding.

Table 6.5: Possible Codings for Subject-Produced Requirements Not in Related Pair

<table>
<thead>
<tr>
<th>Coding</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP</td>
<td>The subject-produced requirement is not in the policy.</td>
</tr>
<tr>
<td>MS</td>
<td>The subject-produced requirement is more specific than the policy.</td>
</tr>
<tr>
<td>LS</td>
<td>The subject-produced requirement is less specific than the policy.</td>
</tr>
<tr>
<td>NR</td>
<td>The subject-produced requirement is not actually a requirement.</td>
</tr>
</tbody>
</table>

Correctness and Completeness

To test the performance of requirements engineers extracting compliance requirements from a policy, we examine the correctness and completeness of each subject-produced set of requirements. Correctness and completeness measure two different parts of effectiveness. Correctness measures whether the requirements produced by a subject relate to the expected compliance requirements. The correctness measure associated with a subject-produced set of requirements corresponds to the number of subject-produced requirements that are related to expected compliance requirements divided by the total number of requirements produced by the subject. Completeness measures whether a set of subject-produced requirements covers every requirement within the expected compliance requirements set. The completeness measure associated with a subject-produced requirements set is the number of expected compliance requirements that are contained within related pairs for the subject divided by the total number of expected compliance requirements. Both correctness and completeness examine related pairs, or subject-produced requirements that contain the same action and specificity as expected compliance requirements, regardless of related pair codings.

Our definitions of correctness and completeness differ slightly from precision and recall [125] because (a) a subject-produced requirement can relate to multiple expected compliance requirements and (b) subject-produced requirements can be repetitive, meaning that multiple subject-produced requirements can relate to a single expected compliance requirement. Therefore, we calculate the number of true positives (related pairs) from different perspectives for correctness and completeness. For correctness, we examine the number of subject-produced requirements that are contained in related pairs; whereas, for completeness, we examine the number of expected compliance requirements that are contained within related pairs for the subject. Using these correctness and completeness measures, we test the following two hypotheses to examine whether there are significant differences in terms of correctness and completeness between the
three conditions:

\( H_{1.0} \): The correctness values of requirements engineers performing CPR analysis, goal-based analysis, and non-method-assisted analysis come from identical distributions having the same median.

\( H_{1.1} \): The correctness values of requirements engineers performing CPR analysis, goal-based analysis, and non-method-assisted analysis do not come from identical distributions having the same median.

\( H_{2.0} \): The completeness values of requirements engineers performing CPR analysis, goal-based analysis, and non-method-assisted analysis come from identical distributions having the same median.

\( H_{2.1} \): The completeness values of requirements engineers performing CPR analysis, goal-based analysis, and non-method-assisted analysis do not come from identical distributions having the same median.

6.1.6 Analysis Tests

To test our hypotheses, we use the Kruskal-Wallis test, which is a nonparametric rank-sum test for three or more samples. For a \( p \)-value less than .05, we reject the null hypothesis because at least one distribution and median significantly differs from the other two. When a significant difference between the distributions exists, we conduct follow-up tests to evaluate pairwise differences between the three groups using the Wilcoxon Rank Sum test. In order to control for Type I errors (rejecting true null hypothesis) across tests, we use the Bonferroni correction approach for the Wilcoxon Rank Sum test. We also use median and range to compare the three condition groups.

6.2 Pilot Study

We conducted a pilot study in the CSC 510 course at NCSU during the Spring 2011 semester. Eighteen subjects completed both analysis tasks for the pilot study. Four additional subjects began the study but did not complete both tasks; these results were not included in our analysis. In order to ensure that experts were not overrepresented in any one condition or group, the subjects were assigned to the three groups in a balanced fashion based on the subjects’ background and software engineering experience—modeling, database modeling, and requirements engineering.

The eighteen subjects produced a total of 370 requirements. Although the CPR subjects did not produce as many requirements as the other subjects, their requirements did relate to
more expected compliance requirements than the requirements produced by the goal-based and control subjects. We reject our null hypothesis \((H_{1,0})\) that the correctness values of requirements engineers in the three conditions come from identical distributions having the same median \((p = .010)\). Through pairwise comparisons, we found that the subjects performing CPR analysis had significantly greater correctness than the subjects performing goal-based analysis \((p = .002)\).

We weakly reject our null hypothesis \((H_{2,0})\) that the completeness values of requirements engineers in the three conditions come from identical distributions having the same median \((p = .065)\). Through pairwise comparisons, we found the subjects performing CPR analysis had slightly greater completeness than the subjects performing goal-based analysis \((p = .018)\).

Based on the results of this pilot study, we adapted parts of the study and repeated it on a different set of subjects for the final study. We added emphasis on requirements splitting within the CPR problem analysis instructions. We also added to the requirements analysis task instructions and asked all subjects to report how they translated artifacts from Part 1 to requirements in Part 2.

6.3 Data Analysis

This section discusses the results of the final study (Fall 2011), including responses to open-ended questions for the analysis tasks and results of empirical evaluation.

6.3.1 Subject Responses to Open-ended Questions for Analysis Tasks

As part of the two analysis tasks, the 34 subjects answered questions about completing each task. First we examine the subjects’ responses to the questions for the problem analysis task. The majority of the subjects (24) stated that the analysis enabled them to identify, uncover, or understand an aspect of the problem that they did not acquire when they first read the policy document (CPR: 5, GB: 11, control: 8). The majority of the subjects (22) stated that they did not have prior experience performing software problem analysis (CPR: 8, GB: 9, control: 5). The majority of the subjects (17) that found the analysis helpful in understanding the policy also did not have prior experience with software problem analysis (CPR: 5, GB: 9, control: 3).

The subjects’ responses to the questions for the requirements analysis task revealed that the majority of the subjects (25) believed the analysis enabled them to identify, uncover, or understand an aspect of the problem that they did not acquire during their initial problem analysis exercise (CPR: 7, GB: 10, control: 8). The majority of the subjects (30) stated they found the analysis method they used useful for clearly identifying requirements (CPR: 11, GB: 10, control: 9).

The subjects also provided the amount of time they spent performing their analysis (both tasks). Three subjects—one CPR subject and two goal-based subjects—did not provide their
time. For the remaining 31 subjects, the control subjects stated a lower median number of hours performing the analysis (both tasks) (Median - CPR: 5, GB: 5, control: 4). These differences may relate to the differences in the amount of guidance given for each analysis approach. The control subjects were also the least consistent in terms of time spent as the number of hours spent on the analysis had the most variation between subjects (Range - CPR: 5.5, GB: 4, control: 10.95).

6.3.2 Results of Empirical Evaluation

This section discusses the results of our empirical evaluation for the study. The thirty-four subjects produced a total of 652 requirements. The total number of requirements produced by each subject is shown in Figure 6.1. The control subjects produced a higher median number of requirements in total than the CPR and goal-based subjects (Median - CPR: 18, GB: 15, control: 19). However, the control subjects were also the least consistent as the number of subject-produced requirements had the most variation between subjects (Range - CPR: 19, GB: 20, control: 47). The control subjects had the most variation in terms of the number of subject-produced requirements identified, which is not surprising because they received the least amount of guidance for the analysis.

Figure 6.1: Total Number of Subject-Produced Requirements by Condition and Subject

The total number of expected compliance requirements to which the subject-produced...

4The raw data for all figures in this section can be seen in Table B.1.
requirements related is shown in Figure 6.2. The CPR subjects produced significantly more expected compliance requirements to which the subject-produced requirements related than goal-based subjects ($p < .001$) and control subjects ($p = .019$). The CPR subjects produced a higher median number of expected compliance requirements to which the subject-produced requirements related (Median - CPR: 23.5, GB: 7, control: 16). The CPR subjects were more consistent than the goal-based and control subjects as the number of expected compliance requirements to which the subject-produced requirements related had the least variation between subjects (Range - CPR: 14, GB: 22, control: 29).

CPR subjects produced significantly more expected compliance requirements to which the subject-produced requirements related than:

- goal-based subjects ($p < .001$)
- control subjects ($p = .019$)

**Figure 6.2:** Total Number of Expected Requirements to which Subject-Produced Requirements Related

**Correctness and Completeness**

The correctness values for the subjects in the three conditions are shown in Figure 6.3. We reject our null hypothesis ($H_{1.0}$) that the correctness values of requirements engineers in the three conditions come from identical distributions having the same median ($p < .001$). Through pairwise comparisons, we found that the subjects performing CPR analysis had significantly greater correctness, or produced significantly more subject-produced requirements related to
expected compliance requirements, than the subjects performing goal-based analysis \((p < .001)\) and non-method-assisted analysis \((p < .001)\). There were no significant differences between subjects with modeling experience and subjects with no modeling experience.

CPR subjects yielded significantly greater correctness than:
- goal-based subjects \((p < .001)\)
- control subjects \((p < .001)\)

**Figure 6.3:** Correctness by Condition and Subject

The completeness values for the subjects in the three conditions are shown in Figure 6.4. We reject our null hypothesis \((H_{2.0})\) that the completeness values of requirements engineers in the three conditions come from identical distributions having the same median \((p = .002)\). Through pairwise comparisons, we found the subjects performing CPR analysis had significantly greater completeness, or significantly more expected compliance requirements were contained in related pairs, than the subjects performing goal-based analysis \((p < .001)\) and non-method-assisted analysis \((p = .019)\). There were no significant differences between subjects with modeling experience and subjects with no modeling experience.

The medians and ranges for the correctness and completeness values by condition are provided in Table 6.6. The CPR subjects had higher median correctness and completeness values than the goal-based and control subjects. A higher correctness value is desirable because it expresses a higher percentage of subject-produced requirements that relate to expected compliance requirements. A higher completeness value is desirable because it expresses a higher percentage of subject-identified expected compliance requirements. The CPR subjects had
CPR subjects yielded significantly greater completeness than:

- goal-based subjects \((p < .001)\)
- control subjects \((p = .019)\)

**Figure 6.4:** Completeness by Condition and Subject

The highest median percentage of their subject-produced requirements relate to the expected compliance requirements. In addition, they identified a higher median percentage of the expected compliance requirements than the goal-based and control subjects. The CPR subjects also had a smaller range for their correctness values than the goal-based and control subjects. Therefore, the CPR subjects were the most consistent because the CPR subjects produced similar percentages of subject-produced requirements that related to expected compliance requirements, whereas the goal-based and control subjects produced unequal percentages of subject-produced requirements that related to expected compliance requirements. CPR subjects also had smaller ranges for their completeness values than the goal-based and control subjects. Therefore, CPR subjects were more consistent than the goal-based and control subjects because the CPR subjects developed similar numbers of expected compliance requirements, whereas the goal-based and control subjects developed unequal numbers of expected compliance requirements. Based on correctness and completeness values, CPR subjects were more effective than goal-based and control subjects at extracting compliance requirements from the policy.

**Coding Combinations with Significant Differences between Conditions**

The number of related pairs with no coding for the subjects in each condition is shown in Figure 6.5. Each subject-produced requirement in these related pairs contained the same action
Table 6.6: Correctness and Completeness for Conditions

| Condition | Correctness | | | Completeness | | |
|-----------|-------------|-----------------|-------------|
|           | Median      | Range           | Median      | Range           |
| CPR       | 0.788       | 0.476           | 0.322       | 0.205           |
| GB        | 0.316       | 0.733           | 0.096       | 0.301           |
| Control   | 0.400       | 0.533           | 0.219       | 0.384           |

and specificity as an expected compliance requirement, and the related pair did not have the unwanted qualities—differing modality than the expected compliance requirement, differing point of view than the expected compliance requirement, and relating to multiple expected compliance requirements. The CPR subjects produced significantly more related pairs with no coding than goal-based subjects \( (p < .001) \) and control subjects \( (p < .001) \). The CPR subjects produced the highest median number of related pairs with no coding (Median - CPR: 8.5, GB: 2.0, control: 3.0), but the goal-based subjects were the most consistent as a group (Range - CPR: 12, GB: 8, control: 10).

Figure 6.5: Total Number of Related Pairs with No Coding

CPR subjects produced significantly more related pairs with no coding than:

- goal-based subjects \( (p < .001) \)
- control subjects \( (p < .001) \)

The number of related pairs with Mul coding is shown in Figure 6.6. The subject-produced
requirements in these related pairs each related to multiple expected compliance requirements. Subjects performing CPR analysis produced significantly more related pairs with Mul coding than subjects performing goal-based analysis ($p = .008$). The goal-based subjects produced the lowest median number of related pairs with Mul coding (Median - CPR: 10.5, GB: 0.0, control: 6.0) and were the most consistent (Range - CPR: 19, GB: 11, control: 14).

The number of related pairs with PoV/Mul coding is shown in Figure 6.7. Each subject-produced requirement in these related pairs contains the same action and specificity as multiple expected compliance requirements, and the subject-produced requirement has different point of view than the expected compliance requirements. Most of the subjects did not produce related pairs with PoV/Mul coding. Subjects performing non-method-assisted analysis produced significantly more related pairs with PoV/Mul coding than subjects performing CPR analysis ($p = .005$). The CPR and goal-based subjects produced the lowest median number of related pairs with PoV/Mul coding (Median - CPR: 0, GB: 0, control: 2), and CPR subjects were the most consistent (Range - CPR: 0, GB: 3, control: 5).

The number of subject-produced requirements with NP coding is shown in Figure 6.8. These subject-produced requirements are not in the policy. The CPR subjects produced a significantly different distribution of subject-produced requirements with NP coding than goal-based subjects ($p = .003$) and control subjects ($p = .022$). The CPR subjects produced the lowest median
Control subjects produced significantly more related pairs with PoV/Mul coding than CPR subjects \((p = .005)\) and were also the most consistent \((\text{Range} - \text{CPR}: 8, \text{GB}: 23, \text{control}: 12)\).

**Derived or Synthesized Requirements**

We also examined whether the subject-produced requirements were derived or synthesized. Derived requirements are identified directly from policies, and synthesized requirements are identified through additional inquiry \([18]\). Derived requirements included subject-produced requirements that were related to expected compliance requirements and subject-produced requirements that were less specific than policy \((LS\ coding), which are shown in Figure 6.9. The CPR subjects produced significantly more derived requirements than goal-based subjects \((p < .001)\) and control subjects \((p = .002)\). Therefore, CPR subjects performed better at extracting requirements directly from the policy. The CPR subjects produced the highest median number of derived requirements \((\text{Median - CPR: 15, GB: 5, control: 9})\), but the goal-based subjects were the most consistent as a group \((\text{Range - CPR: 16, GB: 11, control: 28})\).

Synthesized requirements included subject-produced requirements that contained actions that were not contained within the policy \((NP\ coding), subject-produced requirements that place additional constraints on the system in comparison to the policy \((MS\ coding), and subject-produced requirements that do not actually express requirements \((NR\ coding), which are shown in Figure 6.10. The CPR subjects produced significantly fewer synthesized requirements than goal-based subjects \((p = .003)\) and control subjects \((p = .003)\). Goal-based and control subject
CPR subjects produced a significantly different distribution of subject-produced requirements with \textit{NP} coding than:

- goal-based subjects ($p = .003$)
- control subjects ($p = .022$)

\textbf{Figure 6.8: Total Number of Subject-Produced Requirements with \textit{NP} Coding}

CPR subjects produced significantly more derived requirements than:

- goal-based subjects ($p < .001$)
- control subjects ($p = .002$)

\textbf{Figure 6.9: Derived Requirements}
performed better at extracting requirements of the system not discussed within the policy document. The CPR subjects produced the lowest median number of synthesized requirements (Median - CPR: 2, GB: 11, control: 12) and were the most consistent (Range - CPR: 10, GB: 24, control: 21).

CPR subjects produced significantly fewer synthesized requirements than:

- goal-based subjects ($p = .003$)
- control subjects ($p = .003$)

**Figure 6.10: Synthesized Requirements**

**Inter-rater Reliability**

We examined the inter-rater reliability, or agreement, of the raters—the CPR subjects and the CPR expert (author)—when determining whether an item is a commitment, privilege, right, or unclassifiable. We measured this agreement through the use of Fleiss’ Kappa [71]. The Fleiss’ Kappa was 0.493. Based on Landis and Koch’s levels of agreement [91], this kappa represents a moderate level of agreement between the raters.

**6.4 Threats to Validity**

We examine the validity of this study based on Yin’s validity tests, which include construct validity, internal validity, external validity, and reliability [130]. We maintain construct validity
by establishing a chain of evidence through subject artifacts and having a group of experienced requirements engineering researchers review our study report. A possible threat to construct validity is that a single source for analysis—portion of Facebook policy document—was used. Possible threats to internal validity include the inferences made based on the results. To mitigate these threats, we assigned subjects to three condition groups in a balanced fashion based on their background and software engineering experience. A possible threat to external validity, or being able to generalize results, is that subjects only examined a portion of a single policy document. There is also a threat that subjects were students rather than practitioners. While this is a threat, others have examined that in certain circumstances there can be both experimental and pedagogical value to a empirical study completed by students [39, 88, 116, 124]. Reliability refers to repeating the study with the same result; we maintain a strict process and instructions within the study.

6.5 Summary

In the United States, the Federal Trade Commission can hold organizations liable for what they say within their policies; thus, it is essential for organizations to include policy documents as a source of requirements when building systems that will be governed by these policies. It can also be the case that systems may have to be retrofitted based on policy development or changes. This study demonstrates requirements engineers’ effectiveness in extracting compliance requirements from a policy document using CPR analysis.

Our study shows that requirements engineers can benefit from using CPR analysis to produce compliance requirements rather than goal-based analysis or non-method-assisted analysis. CPR subjects performed better than the goal-based and control subjects in terms of correctness and completeness. The CPR subjects had a higher median percentage of their subject-produced requirements relate to the expected compliance requirements and identified a higher median percentage of the total expected compliance requirements than the goal-based and control subjects, both of which are important for compliance. This result was not surprising as CPR analysis was specifically developed to aid in extracting compliance requirements from policy documents. CPR subjects also produced the most related pairs with no codings. This means that they were better at extracting requirements that were related to the expected compliance requirements that did not have the unwanted qualities—differing modality than the expected compliance requirement, differing point of view than the expected compliance requirement, and relating to multiple expected compliance requirements.

When extracting requirements, it is important to split the requirements in order to include a single, testable item by avoiding the use of ‘and’ [29, 72, 121]. We were surprised to find that CPR subjects had a higher median number of related pairs with the Mul coding than
the other groups. This suggests that we need to improve the CPR heuristics for splitting the requirements and/or further highlight these heuristics to ensure that requirements engineers apply them correctly.

If requirements engineers are not concerned with extracting compliance requirements, goal-based analysis and non-method-assisted analysis may be better approaches than CPR analysis. CPR analysis focuses on derived requirements, whereas goal-based analysis and non-method-assisted analysis focus on synthesized requirements. Recall that derived requirements are identified directly from policies, and synthesized requirements are identified through additional inquiry [18]. The goal-based and control subjects produced a higher median number of synthesized requirements than CPR analysis subjects. This may be attributed to the fact that these two subject groups employed a more inquiry-driven approach than the CPR subjects. The inquiry-driven approach entails asking the following types of questions: what-is, how-to, who, what-kinds-of, when, relationship, what-if, follow-on [18, 114]. Our study suggests that using multiple approaches together would yield a better result by having better coverage of derived and synthesized requirements.
Chapter 7

CPR Analysis within the Development of a Health System Prototype

Don’t let schooling interfere with your education.
- Mark Twain

This chapter discusses the application of CPR analysis to data use agreements (DUAs) within the development of a health system prototype, hereafter referred to as HealthSystem (HS). In our work with policy documents, we applied CPR analysis to documents posted online with no immediate effect on an actual software system. In contrast, HealthSystem is a healthcare organization with which we were able to aid in their efforts to develop a software prototype for predicting potential outbreaks of diseases.

The two CPR-related objectives for the HealthSystem project are: (1) demonstrate that CPR analysis can be applied to the organization’s DUAs thereby aiding the development of its prototype and (2) examine the effects of CPR analysis on the organization’s information practices and the resulting system prototype.

The remainder of this chapter discusses the HealthSystem project, HealthSystem’s software and operational requirements, the standardized DUA template developed for HealthSystem, and our lessons learned.
7.1 Overview of HealthSystem

The HealthSystem prototype is being developed to manage health data from a number of sources (e.g., government agencies, emergency management services data sources, and poison control sources). Upon realizing the importance of proceeding with a better understanding of the requirements as well as the implications for regulatory and contractual compliance, HealthSystem brought us on board to assist in specifying requirements for a prototype that was already in development.

The HealthSystem prototype accesses health data in accordance with governing DUAs that have been signed by participating parties (e.g., government agencies, emergency management services data sources, and poison control sources). Our task was to determine how the governing DUAs would affect the in-development prototype. We identified requirements expressed in the DUAs to specify previously overlooked software and operational requirements that must govern the prototype and HealthSystem’s information practices, respectively.

7.2 Requirements

We applied CPR analysis to four DUAs, governing the HealthSystem prototype, to identify the requirements contained therein. A signed DUA is a legally binding agreement between a covered entity and a limited data set recipient that specifies how the limited data set (LDS)\(^1\) will be used and to whom it will be disclosed [104]. Our work with the HealthSystem prototype led to our formative and summative DUA case studies. As discussed in Chapter 4, we developed CPR analysis for DUAs and validated that it could be applied to DUAs not in the formative data set.

HealthSystem currently has four data sources (from different kinds of covered entities), each relating to one of its four DUAs. Herein, we refer to these four DUAs as: HS-I DUA (government agency), HS-II DUA (poison control source), HS-III DUA (emergency management services data source A), and HS-IV DUA (emergency management services data source B). As discussed in Chapter 1, regulations adopted pursuant to the United States Health Insurance Portability and Accountability Act (HIPAA) specify that a DUA must exist when certain protected health information is used or disclosed. Herein, we focus on a part of the HIPAA Privacy Rule that addresses DUAs: 45 C.F.R. §164.514(e). Within the context of the HIPAA, DUAs are critical, legally binding agreements that must be examined by requirements engineers in order to develop compliant software. Under §164.514(e), DUAs must be established between covered entities.

\(^1\)Recall that a limited data set (LDS) “refers to [protected health information] that excludes 16 categories of direct identifiers and may be used or disclosed, for purposes of research, public health, or health care operations, without obtaining either an individual’s Authorization or a waiver or an alteration of Authorization for its use and disclosure, with a data use agreement” [104]. See Figure 1.1 on page 4 for a summary of what information DUAs must contain.
and limited data set recipients if an LDS is being used or disclosed. The LDS can be used or disclosed only for research, public health, or healthcare operation. DUAs must contain certain information, such as the permitted uses and disclosures, permitted recipients, and safeguards (see Figure 1.1 on page 4 for full list of what must be included in DUAs based on §164.514(e)).

Using CPR analysis, we examined the four DUAs to ensure their compliance with the requirements of §164.514(e). Our analysis revealed that HS-I and HS-II were written in accordance with the requirements expressed in HIPAA §164.514(e). In contrast, HS-III and HS-IV were missing many of the items required by §164.514(e), and their data set contained data elements not allowed within an LDS. HS-III and HS-IV failed to comply with §164.514(e)(3)(i) as the DUAs fail to address the fact that the LDS can be used or disclosed only for research, public health, or healthcare operation purposes. The DUAs also did not address the fact that disclosures in violation of the DUA must be reported and that agents will be held accountable to the standards of the recipient—both of which are required by §164.514(e)(4)(ii)(C)(3) and §164.514(e)(4)(ii)(C)(4), respectively. We also identified non-compliance with §164.514(e)(4)(ii)(C)(5) because the DUAs do not specify that the recipient must not identify the individuals contained within the data. Additionally, the data set corresponding to HS-III DUA contained data elements that should not have been present in a limited data set. These elements included: First Name, Last Name, Incident Address, and Patient Home Address. Similarly, the data set corresponding to HS-IV DUA contained the following elements that should not have been present in it: First Name, Last Name, Incident Address, Patient Home Address, and Social Security Number.

This kind of analysis proved very helpful because applying a CPR-based approach to the DUAs enabled us to perform a gap analysis on DUAs for the purpose of ensuring requirements compliance. Prior to our analysis, the HealthSystem project team was not aware that the DUAs needed modification in order to comply with the DUA requirements specified in §164.514(e). In addition, the non-compliant identifiers contained within the HS-III and HS-IV data sets have been removed from the prototype as a result of our analysis. Based on our analysis, we developed a DUA template that complies with §164.514(e) to aid in establishing future HIPAA-compliant DUAs; we discuss this template in Section 7.3.

CPR analysis revealed that at least 83% of the classifiable items in each of the four DUAs were commitments. As discussed in Chapter 4, the dominant focus on pledges is likely due to DUAs being legally binding agreements that contain acceptable uses and disclosures of data. These acceptable uses and disclosures result in commitments about how the data will be used and disclosed. Figure 7.1² portrays the distribution of classifications among classifiable items for each DUA.

The most prominent actor in the DUAs was the data recipient, HealthSystem; the data recipient was the actor in 71% of the 238 classifiable items. This is somewhat expected given that

²None of the DUAs contained rights, which is why the rights classification is not included in the Figure 7.1.
the recipient is the party receiving and using the data based on the DUA. The distribution of actors among classifiable items for each DUA is shown in Figure 7.2. Whereas the recipient and data custodian are the two main actors in the HealthSystem DUAs, we did observe that other actors—such as employees, agents, and subcontractors—are present in contractual compliance requirements; we group actors other than the recipient and data custodian together for Figure 7.2. Examining the actor in the classifiable items, we found that HealthSystem does not implement CPRs of the data custodian because these CPRs should be implemented within the data custodian’s system or practices, as discussed in Section 7.2.3.

Most, 63%, of the classifiable items in the DUAs expressed commitments from HealthSystem.
(the recipient) to the data custodian. This makes sense given that DUAs express the restrictions on the use of data that the recipient pledges to abide by when it signs the DUAs. The distribution of classification/actor combinations among classifiable items for each DUA is shown in Figure 7.3.³

![Figure 7.3: Classification/Actor Combinations of Classifiable Items in HealthSystem DUAs](image)

The CPRs extracted from the four DUAs were operationalized into 238 contractual compliance requirements. Recall from Chapter 1 that a contractual compliance requirement is a requirement that is agreed upon by two (or more) parties and is enforceable by law; all of the contractual compliance requirements discussed herein were extracted from HealthSystem’s DUAs using CPR analysis. The contractual compliance requirements reflected both system requirements and operational requirements. As discussed in Chapter 2, system requirements and operational requirements are defined as follows:

- **System requirements** specify the capabilities of the system or software. We also refer to these requirements as software requirements.

- **Operational requirements** specify “business rules or operational procedures” [20] outside of the system. We also refer to these requirements as requirements of the organization’s practices.

In collaboration with other HealthSystem prototype project team members, we specified a set of software and operational requirements for the prototype and the organization’s practices,

³For Figure 7.3, we only include classification/actor combinations that existed in the HealthSystem DUAs.
respectively, that reflected the pledges and entitlements expressed in the governing DUAs. The process of merging and reconciling the contractual compliance requirements from the DUAs with broader requirements for the prototype and the organization practices required us to ensure that the contractual compliance requirements were addressed by the prototype and/or the organization’s practices.

The final set of requirements for the HealthSystem prototype and organizational practices is comprised of 189 requirements—89 software requirements and 100 operational requirements. In merging and reconciling requirements, we ensured that each contractual compliance requirement extracted from the DUAs was addressed by at least one software or operational requirement. Of the 189 requirements, 127 addressed contractual compliance requirements, which must be met in order to comply with the DUAs. Most of these contractual compliance requirements would not have been included, risking non-compliance, were it not for our CPR analysis effort.

Sections 7.2.1 and 7.2.2 discuss our software and operational requirements for the HealthSystem prototype. We discuss each category of software and operational requirements within the set. We include the total number of requirements of each category as well as the total number of these requirements that address contractual compliance requirements extracted from the DUAs through CPR analysis. A summary of this information is given in Table 7.1. For each category, we also discuss how CPR analysis aided in the specification of these requirements.

<table>
<thead>
<tr>
<th>Requirement Type</th>
<th>Requirement Category</th>
<th>Number of Requirements</th>
<th>Number of Requirements that Address Contractual Compliance Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Software</td>
<td>Access Control</td>
<td>37</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Auditing and Logging</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Acceptable Use</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Encryption</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>General</td>
<td>23</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td><strong>89</strong></td>
<td><strong>36</strong></td>
</tr>
<tr>
<td>Operational</td>
<td>Acceptable Use</td>
<td>15</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Privacy-Related</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>General</td>
<td>75</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td><strong>100</strong></td>
<td><strong>91</strong></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>189</strong></td>
<td><strong>127</strong></td>
</tr>
</tbody>
</table>
7.2.1 Software Requirements

Thirty-six of the 89 software requirements addressed contractual compliance requirements expressed in the DUAs. The software requirements are organized according to five categories: (1) access control, (2) auditing and logging, (3) acceptable use, (4) encryption, and (5) general.

Access Control

The category of software requirements with the largest number of requirements is access control with 37 of the 89 software requirements. Access control requirements specify who can access data and how that access should be restricted. Without these access control requirements, the prototype could potentially allow unauthorized users to access data.

Six of the 37 access control requirements address contractual compliance requirements extracted from the DUAs. Consider the following access control requirement for HealthSystem’s prototype that addresses contractual compliance requirements:

\[ R_{14}: \text{The prototype shall provide the ability to restrict access to all data elements from, aggregated by, or derived from a particular data source.} \]

\( R_{14} \) describes data access restrictions based upon a particular data source. Because the HealthSystem prototype has multiple data sources or data sets, each governed by a specific DUA, the prototype must be able to restrict access to data elements accordingly. Each DUA specifies the individuals or groups of individuals who can access the data set. In terms of compliance, these are the only individuals who should be able to access the particular data set. Prior to applying CPR analysis to the DUAs, the access control requirements did not include restrictions based on the data source; thus, a prototype user may have been able to access any data set regardless of the set’s data source. This would have resulted in a DUA violation due to not providing access to the same individuals or groups stated in the DUA. Here, CPR analysis led to the inclusion of important access control requirements that were necessary for complying with the DUAs.

Auditing and Logging

Auditing and logging requirements describe both the data that needs to be logged and how the data should be logged to support auditing. We identified seven auditing and logging requirements for the HealthSystem prototype; three of which address contractual compliance requirements yielded from our CPR analysis.

The HS-I and HS-IV DUAs explicitly require the HealthSystem prototype to comply with auditing and logging procedures; for compliance purposes, these procedures must be implemented...
in the final software prototype. Consider the following example of an auditing and logging requirement that addresses contractual compliance requirements yielded from our CPR analysis:

\[ R_{40}: \text{The prototype shall have the capability to generate audit reports in which the purpose of use conflicts with the purposes allowed within the DUA.} \]

Requirement \( R_{40} \) explicitly includes “purposes allowed within the DUA;” this requirement was not surfaced until we applied CPR analysis to the DUAs. Without auditing and logging requirements like \( R_{40} \), the prototype would not have supported the logging necessary to report data use conflicts as described in the HS-I and HS-II DUAs.

**Acceptable Use**

Acceptable use requirements describe how the data set can or cannot be used by system users. We identified 16 acceptable use requirements—all of which address contractual compliance requirements; this makes sense given that §164.514(e) in the HIPAA Privacy Rule specifies that the main purpose of DUAs is to describe acceptable uses and disclosures of healthcare information. In the HealthSystem prototype, data can only be used or disclosed as specified in the DUAs. Without these acceptable use requirements extracted using CPR analysis, it would be difficult (if not impossible) to verify whether or not the system was using the data sources in compliance with the DUAs.

**Encryption**

We identified six encryption software requirements; two of these address contractual compliance requirements that were yielded from CPR analysis. Encryption requirements describe how the data is protected through the use of encryption. Consider the following encryption requirement:

\[ R_{63}: \text{The prototype shall encrypt all data at rest and in transit using the standard algorithms.} \]

Two of the HealthSystem prototype DUAs—HS-III DUA and HS-IV DUA—specifically require all shared data governed by the respective DUA to be encrypted. The other DUAs were more general about the HealthSystem prototype’s safeguards to protect the data. Encryption was included within the requirements prior to CPR analysis being applied to the DUAs. However, our CPR analysis enabled us to confirm the need for encrypting data for particular data sets in order to comply with the commitments of the corresponding DUAs.

**General**

General requirements are system requirements that do not fall into one of the previously described categories; they address a variety of topics including, how the system is implemented, how the
data is tracked from data source, and the use of notifications within the system. We identified 23 general requirements; nine of which address contractual compliance requirements. Consider the following general requirement about notifications:

\[ R_{76} : \text{The prototype shall notify project manager when date is within 30 days of the end of the DUA that the DUA is coming up for expiration.} \]

This requirement, \( R_{76} \), is inferred from contractual compliance requirements, which discussed the termination or expiration of HS-I and HS-II DUAs. \( R_{76} \) was inferred because the HealthSystem prototype project manager must know when each DUA is set to expire in order to ensure that the data is returned or destroyed in compliance with the DUA. Prior to performing CPR analysis on the DUAs, there was no evidence that the requirements included this notion that a data set must be returned or destroyed based on the terms and effective dates specified in the DUAs. Although none of the DUAs' effective dates lapsed prior to our application of CPR analysis, there was a risk of non-compliance due to a lack of awareness of the effective dates specified therein and the need to act according to reaching the end of the effective dates.

### 7.2.2 Operational Requirements

Operational requirements describe an organization’s operations or procedures. These requirements describe, for example, how HealthSystem returns or destroys the data when the DUAs expire or terminate; restrictions on publishing results from the analysis of the data; and compliance with legal texts. Although not actually implemented in the system, these requirements are essential to ensure that HealthSystem complies with all governing DUAs. Prior to our application of CPR analysis to DUAs, there is no evidence that a set of operational requirements existed that described the HealthSystem’s practices. Ninety-one of the 100 final operational requirements address contractual compliance requirements expressed in the DUAs. We discuss three categories of operational requirements: acceptable use, privacy-related, and general.

#### Acceptable Use

The acceptable use requirements comprised both software requirements and operational requirements. The operational requirements specify acceptable use of the data outside of the prototype. Fifteen of the 100 operational requirements are acceptable use requirements; thirteen of these address contractual compliance requirements. All of the DUAs describe the sharing and publication of information and analysis results. HS-I, HS-II, and HS-IV DUAs require the data custodians’ approval prior to any publications based on the data sets. Without addressing the operational contractual compliance requirements relating to acceptable uses, it would be difficult (if not impossible) to verify whether or not the organization was using the data sources in a way that complied with the DUAs.
Privacy-Related

Ten of the operational requirements are privacy-related; nine of these address contractual compliance requirements. These requirements specify what the HealthSystem prototype must do to protect the privacy of the individuals whose health information is contained within the prototype. Consider the following privacy-related requirement, which addresses contractual compliance requirements operationalized via CPR analysis of HS-I and HS-II DUAs:

\[ R_{184} \text{: The recipient shall make no attempt to re-identify any individual from any one or combined data source.} \]

When performing CPR analysis, we documented that these items had external rationale, specifically §164.514(e)(4)(ii)(C)(5) of the HIPAA Privacy Rule that states the recipient will “not identify the information or contact the individuals.” This section of the HIPAA Privacy Rule applies to DUAs in general, and therefore, the requirement that addresses it also applies to all of the data sets obtained through DUAs. This includes the data sets corresponding to HS-III and HS-IV DUAs, despite the fact that these DUAs lacked any statements to this effect. This example demonstrates the importance of identifying the rationale in the CPR analysis methodology. Had we not understood the external rationale for \( R_{184} \), the requirement may have been implemented such that it only applied to datasets whose DUAs explicitly stated the requirement.

General

A general operational requirement is a requirement about an organization’s operations or procedures not related to acceptable use or privacy. Seventy-five of the operational requirements are general operational requirements; 69 of these requirements address contractual compliance requirements extracted from the DUAs. Consider the following general operational requirement that addresses contractual compliance requirements:

\[ R_{106} \text{: The recipient shall report any identified data breaches, inappropriate uses, or inappropriate disclosure as well the associated impact to data custodians.} \]

This requirement has an external rationale in the HIPAA Privacy Rule §164.514(e)(4)(ii)(C)(3): “Report to the covered entity any use or disclosure of the information not provided for by its data use agreement of which it becomes aware.” Requirement \( R_{106} \) exemplifies why general operational requirements need to be identified. If requirement \( R_{106} \) is not included within HealthSystem’s practices, HealthSystem would not comply with the DUAs or the particular section of the HIPAA Rule from which it is derived.
7.2.3 Data Custodian Requirements

Data custodian requirements are requirements that contain the data custodian as the actor. These requirements are not implemented by the recipient, HealthSystem, but instead are requirements of the data custodian’s system or organizational practices. Consider the following data custodian requirement from HS-II:

\[ \text{HS-II}_1: \text{The data custodian shall collect information on all calls received by the data custodian related to known or suspected poisoning exposures as well as informational requests, some of which may contain IIHI as defined by HIPAA or PHI.} \]

This collection of information by the data custodian is not an action that occurs within HealthSystem’s prototype or practices. Instead, the collection occurs prior to the data custodian sharing the data with HealthSystem’s prototype. Therefore, this requirement should be implemented within the data custodian’s system or practices, and we did not include these data custodian requirements in our final set of HealthSystem requirements.

7.3 DUA Template

CPR analysis on HealthSystem’s DUAs revealed the need for a standard DUA template to govern all future DUAs. A template will aid HealthSystem prototype project in standardizing how it receives data from multiple data sources. We created a DUA template based on our CPR analysis and our understanding of DUAs as specified in the HIPAA Privacy Rule and related regulations. Guidelines about DUAs, which were provided by the project’s legal counsel, were helpful in structuring and wording the template. The template shown in Appendix C will be finalized with the help of the project’s legal counsel.

The new HealthSystem prototype DUA template will help ensure that future DUAs have matching formats, contain the same basic requirements, and comply with the HIPAA Privacy Rule.

7.4 Summary

We have shown that by applying CPR analysis to the HealthSystem DUAs we were able to identify previously overlooked, yet critical, contractual compliance requirements. We addressed these contractual compliance requirements by incorporating them into a set of software and operational requirements. We now discuss lessons learned and limitations of our work on the HealthSystem prototype.

CPR analysis can be applied to aid in the efforts to develop a software prototype. Our work with the HealthSystem prototype shows that CPR analysis can be applied to DUAs to aid the
development of a software prototype. Although we have yet to validate that CPR analysis can be used within the context of a real system, our experience with this prototype suggests that CPR analysis is scalable for application within the context of a real production system. By applying CPR analysis to DUAs, we aided the development of the prototype in two primary ways, by: (1) identifying contractual compliance requirements and (2) developing a DUA template. We have demonstrated that many of the contractual compliance requirements would not have been addressed by the system or the operational requirements had CPR analysis not been performed on the DUAs. Without these requirements, the HealthSystem prototype would have been exposed to both regulatory sanctions and legal liability for breach of contract. CPR analysis also revealed the need for a standard DUA template. The DUA template we developed will help ensure that future DUAs will have matching formats, contain the same basic requirements, and comply with the HIPAA Privacy Rule.

All project members need to understand the importance of complying with DUAs. The HealthSystem prototype was already in development when we joined the project team. Many of the project members viewed the DUAs as documents that authorized them to obtain data, but they did not understand the effects that compliance with the agreements would have on the prototype development and the organization. By applying CPR analysis to the DUAs, we produced sets of CPRs and requirements. The CPRs, and in particular the commitments, provided an unexpected benefit of serving as talking points during project meetings. As talking points, the commitments allowed us to reiterate to the project members the pledges that the HealthSystem project team made in each DUAs. We were also able to stress the fact that after making the pledges we must ensure that the prototype and/or the organization comply with them. Having a list of commitments helped the project members to understand how the DUAs would affect the prototype and the organization, as well as why it was important to comply with the DUAs. Once the members had this understanding, they were more willing to accept the inclusion of the documents and requirements within the prototype and the project team. Because the project is ongoing and simply a prototype, a limitation of our work is that we are unable to draw any conclusions about effects on a final production system.
Chapter 8

Conclusion

The only true way is to marshal the ferocity of your ambition over the course of many days, weeks, months, and (if you could finally come to accept it) years.

_The Trial of Miles; Miles of Trials._
- John L. Parker, Jr, *Once A Runner*

The research presented in this dissertation addresses the problem of developing compliant software systems. Specifically, _given that organizations’ policy documents and data use agreements express commitments, privileges, and rights that can be legally binding, requirements engineers must specify these commitments, privileges, and rights as software requirements_. To address this problem, we developed CPR (commitment, privilege, and right) analysis, which is an approach to derive requirements from policy documents and data use agreements (DUAs). During CPR analysis, requirements engineers extract commitments, privileges, and rights from the documents and operationalize them as requirements.

This chapter provides summaries of the previous chapters, summarizes the contributions of the work, and discusses potential future work.

### 8.1 Chapter Summaries

Chapter 1 introduced the problem addressed by this work. We described the motivation for examining and complying with policy documents and DUAs. In the United States, the Federal Trade Commission (FTC) is empowered to monitor organizations’ compliance with policy documents. Thus, organizations need to ensure that they are doing what they say they do within their policy documents. A DUA is a legally binding agreement between a covered entity and a limited data set recipient that specifies how the limited data set will be used and to whom it will
be disclosed [104]. Based on regulations adopted pursuant to the Health Insurance Portability and Accountability Act (HIPAA), DUAs must contain certain information and restrictions. We also discussed several areas of related work.

Chapter 2 presented the theory and methodology for CPR analysis. The theory of commitments, privileges, and rights discussion included a comparison of our classifications to concepts in the following areas: legal concepts, legal-based requirements engineering, and multi-agent agreements. We described the three steps that comprise the CPR analysis methodology—parse, classify, and operationalize.

Chapter 3 introduced the heuristics used to perform CPR analysis on policy documents and DUAs. We provided heuristics for each analysis step. Some of the heuristics apply to both types of documents—policy documents and DUAs, other heuristics are document type-specific. The heuristics for classifying items as commitments, privileges, and rights are based on natural language phrases present within the documents.

Chapter 4 discussed our formative and summative multi-case studies. The two cases were based on the two types of documents—policy documents and DUAs. As a result of these studies, we developed CPR analysis and validated it by applying it to additional documents. We observed that overall the policy documents produced more privileges than commitments, and the DUAs consisted primarily of commitments, which may relate to the fact that DUAs are legally binding agreements that contain acceptable uses and disclosures of data. The CPRs from policy documents overwhelmingly had internal rationale, whereas the CPRs from DUAs had more external items, which may be due to the fact that regulations adopted pursuant to the HIPAA specify that a DUA must exist when certain protected health information is used or disclosed.

Chapter 5 presented our comparison study in which we compared CPR analysis to goal-based analysis based on each approach’s ability to ensure compliance. We observed that by maintaining the language, specificity, and modality of the policy documents CPR analysis ensured compliance more directly than the goal-based approach.

Chapter 6 discussed our empirical user study. In this study subjects were divided into three groups, each of which performed a different analysis method in order to extract requirements from a policy document. We examined whether the requirements produced by the subjects were expected requirements and what percentage of the expected requirements each subject produced. We found that requirements engineers can benefit from using CPR analysis to produce compliance requirements rather than goal-based analysis or non-method-assisted analysis. The CPR subjects had a higher median percentage of their subject-produced requirements relate to the expected compliance requirements and identified a higher median percentage of the total expected compliance requirements than the goal-based and control subjects, both of which are important for compliance.
Chapter 7 discussed our work within the development of a health system prototype to manage health data from a number of sources to support analytical decision-making. We describe our experiences applying CPR analysis to the organization’s DUAs. We showed that by applying CPR analysis to the DUAs we were able to identify previously overlooked, yet critical, contractual compliance requirements.

8.2 Contributions

The following subsections outline the primary contributions of this work and the demonstrated experiences from the studies performed to validate the work.

8.2.1 Primary Contributions

The primary contributions of this dissertation are:

- A theory of commitments, privileges, and rights to describe the items of importance for requirements and compliance within policy documents and DUAs (Chapter 2);

- A validated CPR analysis methodology for extracting commitments, privileges, and rights from policy documents and DUAs and operationalizing them as requirements (Chapter 2); and

- A set of heuristics to guide analysts through each step of the methodology, which is summarized in Appendix A (Chapter 3).

8.2.2 Demonstrated Experiences

The experiences reported in this dissertation demonstrate that:

- Because policy documents focus primarily on internal rationale rather than external rationale, legal-based approaches do not provide sufficient coverage of compliance requirements. (Chapter 4)

- CPR analysis ensures compliance more directly than goal-based analysis by maintaining language, specificity, and modality. (Chapter 5)

- Requirements engineers can effectively apply CPR analysis to produce compliance requirements. (Chapter 6)

- CPR analysis identified contractual compliance requirements that would have be overlooked in a system. (Chapter 7)
Commitments, privileges, and rights can be used by project members to discuss DUAs and the need for compliance. (Chapter 7)

8.3 Future Work

The work we have presented addresses the need for an approach that includes policy documents and DUAs in the requirements engineering process for compliance reasons. However, natural extensions to this work provide a rich space for us to continue our research, such as automating the CPR analysis process, applying CPR analysis within other domains, and improving organizations’ documents through CPR analysis.

CPR analysis is currently a manual process that is applied by the requirements engineers. We plan to transition the approach from a purely manual process to at least a partially automated process. This extension will require natural language processing and machine learning research.

To this point we have focused on policy documents and DUAs within the healthcare domain. Further studies are needed to determine how the approach is generalizable to domains other than healthcare. Along with generalizing to other domains, we plan to examine other types of documents, such as business associate agreements.

CPR analysis focuses on an organization complying with its policy documents and DUAs by doing what it states within its policy documents and DUAs, respectively. We plan to examine how CPR analysis can be employed to assist organizations in improving their documents. Another extension of the work is examining the effects CPRs have on privacy as related to both the organizations and the user.
REFERENCES


135


APPENDICES
Appendix A

CPR Analysis Heuristics

A.1 Parsing Heuristics

Table A.1: Parsing Heuristics (PH)

<table>
<thead>
<tr>
<th>Name</th>
<th>Heuristic</th>
<th>Page</th>
</tr>
</thead>
</table>
| PH 1 | Given a paragraph of the document, follow the following steps to parse it into sentences:  
1. Tentatively place sentence boundaries after all periods and question marks.  
2. Disqualify boundaries if the period follows an abbreviation.  
3. Disqualify boundaries if the period or question mark is part of a URL or email address.  
The result will be a set of statements based on sentences. | 25 |
| PH 2 | - For lists that do not contain continuations, parse each list item using PH 1.  
- For lists that contain continuations, complete the following steps:  
1. Prepend the fragment before the list to each list item.  
2. Use PH 1 to parse the resulting list items. | 25 |
| PH 3 | Parse each section heading as a separate statement. | 26 |
A.2 Classification Heuristics

Table A.2: Item Heuristic (IH)

<table>
<thead>
<tr>
<th>Name</th>
<th>Heuristic</th>
<th>Page</th>
</tr>
</thead>
</table>
| IH 1 | If the item is one of the following:  
- **Clarification**: item clarifies or explains another item.  
- **Definition**: item defines a key word or phrase.  
- **Section heading**: item denotes the start of a section.  
- **Warning**: item provides a warning, advisory, or suggestion.  
Note that the item is **unclassifiable**. Otherwise, note that the item is **classifiable**. | 27 |

Table A.3: Helper Heuristic (HH)

<table>
<thead>
<tr>
<th>Name</th>
<th>Heuristic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>HH 1</td>
<td>If an item contains multiple actions, split the item into multiple items—each of which contains a single action.</td>
<td>28</td>
</tr>
<tr>
<td>HH 2</td>
<td>When a pronoun is present in an item, change the pronoun to the antecedent that the noun references.</td>
<td>29</td>
</tr>
</tbody>
</table>

Table A.4: Classifying Heuristic (CH)

<table>
<thead>
<tr>
<th>Name</th>
<th>Heuristic</th>
<th>Page</th>
</tr>
</thead>
</table>
| CH 1 | - If the entitlement contains an independent action (no complementary action exists), classify it as a **privilege**.  
- If the entitlement contains a dependent action (a complementary action from another party that must take place exists), classify it as a **right**. | 29 |
### Table A.5: Classifying Heuristic for Policy Documents (CH\textsubscript{PD}) - Part 1

<table>
<thead>
<tr>
<th>Name</th>
<th>Heuristic</th>
<th>Page</th>
</tr>
</thead>
</table>
| CH\textsubscript{PD} 1 | If the item is in the format \[actor\] \[modal verb\] \[action\], classify the item based on the modal verb and possibly the actor.  
- If the item contains the modal verb \textit{must}, classify it as a \textbf{commitment} because the item expresses necessity through a pledge.  
- If the item contains the modal verb \textit{can, could, may, or might}, then it expresses a possibility through an action that the actor is entitled to perform. Classify the entitlement as a \textbf{privilege} or a \textbf{right} using CH 1.  
- If the item contains the modal verb \textit{do, shall, should, will, or would}, then it is actor-dependent.  
  - If the \textit{organization} is the \textit{actor}, classify the item as a \textbf{commitment} because the item expresses what the organization pledges to the user.  
  - If the \textit{user} is the \textit{actor}, the item is an entitlement. Classify the entitlement as a \textbf{privilege} or a \textbf{right} using CH 1. | 30   |
| CH\textsubscript{PD} 2 | If the item is in format \[actor\] \[action\], classify it based on the actor.  
- If the \textit{organization} is the \textit{actor}, classify the item as a \textbf{commitment} because it expresses an action the organization pledges to the user.  
- If the \textit{user} is the \textit{actor}, then the item expresses an action the user is entitled to perform and is an entitlement. Classify the entitlement as a \textbf{privilege} or a \textbf{right} using CH 1. | 31   |
| CH\textsubscript{PD} 3 | If the item is in one of the following formats:  
- \[actor\] \textit{is/are able to [action]}  
- \[actor\] \textit{may be able to [action]}  
- \[actor\] \textit{should be able to [action]}  
- \[actor\] \textit{will be able to [action]}  
- \[actor\] \textit{is/are entitled to [action]}  
- \[actor\] \textit{can choose to [action]}  
- \[actor\] \textit{may choose to [action]}  
- \[actor\] \textit{will have the authority to [action]}  
- \[actor\] \textit{has/have the ability to [action]}  
The item is an entitlement. Classify the entitlement as a \textbf{privilege} or a \textbf{right} using CH 1. | 31   |
| CH\textsubscript{PD} 4 | If the item is in one of the following formats:  
- \[actor\] \textit{is/are required by law to [action]}  
- \[actor\] \textit{is/are required to [action]}  
- \[legal text\] \textit{requires [actor] to [action]}  
- \[legal texts\] \textit{require [actor] to [action]}  
- \[system\] \textit{requires [actor] to [action]}  
- \[system\] \textit{will require [actor] to [action]}  
Classify the item as a \textbf{commitment}. | 32   |
| CH\textsubscript{PD} 5 | If the item is in the following format:  
\[actor\] \textit{is/are committed to [present participle of action]}  
Classify the item as a \textbf{commitment}. | 32   |
Table A.6: Classifying Heuristic for Policy Documents (CH\textsubscript{PD}) - Part 2

<table>
<thead>
<tr>
<th>Name</th>
<th>Heuristic</th>
<th>Page</th>
</tr>
</thead>
</table>
| CH\textsubscript{PD} 6 | If the item is in one of the following formats:  
  - [legal text] affords [actor] the right to [action]  
  - [legal texts] afford [actor] the right to [action]  
  - [legal text] allows [actor] the right to [action]  
  - [legal texts] allow [actor] the right to [action]  
  - [legal text] gives [actor] the right  
  - [legal texts] give [actor] the right  
  - [actor] has/have the right to [action]  
  - [actor] reserve/s the right to [action]  
  The item is an entitlement. Classify the entitlement as a privilege or a right using CH 1. | 33 |
| CH\textsubscript{PD} 7 | If the item is in one of the following formats:  
  - [actor] agree to [action]  
  - [actor] consent to [action]  
  - [actor] has/have agreed to [action]  
  Classify the item as a commitment. | 34 |
| CH\textsubscript{PD} 8 | If the item is in one of the following formats:  
  - [actor] agree to [object]  
  - [actor] consent to [object]  
  Classify the item as a commitment. The commitment has agree or consent as the action. | 34 |
| CH\textsubscript{PD} 9 | If the item is in one of the following formats:  
  - If [condition], please [action]  
  - If [condition], [action]  
  - When [condition], [action]  
  - [purpose], please [action]  
  - [purpose], [action]  
  - Please [action]  
  The item is an entitlement of the user. Classify the entitlement as a privilege or a right using CH 1. | 35 |
| CH\textsubscript{PD} 10 | If the item is in one of the following formats:  
  - [object] [action]  
  - [objects] are [past participle of action]  
  - [object] is [past participle of action]  
  - [object] do not [action]  
  - [object] shall be [past participle of action]  
  - [object] will be [past participle of action]  
  Classify the item as a commitment of the organization. | 35 |
<table>
<thead>
<tr>
<th>Name</th>
<th>Heuristic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH$_{DU,A}$ 1</td>
<td>If the item is in the format [actor] [modal verb] [action], classify the item based on the modal verb.</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>- If the item contains the modal verb <em>must</em>, <em>shall</em>, or <em>will</em>, classify it as a <strong>commitment</strong> because the item expresses necessity through a pledge.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- If the item contains the modal verb <em>may</em>, then it expresses a possibility through an action that the actor is entitled to perform. Classify the entitlement as a <strong>privilege</strong> or a <strong>right</strong> using CH 1.</td>
<td></td>
</tr>
<tr>
<td>CH$_{DU,A}$ 2</td>
<td>If the item is in format of [actor] [action], classify the item as a <strong>commitment</strong>.</td>
<td>36</td>
</tr>
<tr>
<td>CH$_{DU,A}$ 3</td>
<td>If the item is in one of the following formats:</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>- [actor] has the authority to [action]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- [actor] shall have the right to [action]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>The item is an entitlement. Classify the entitlement as a <strong>privilege</strong> or a <strong>right</strong> using CH 1.</td>
<td></td>
</tr>
<tr>
<td>CH$_{DU,A}$ 4</td>
<td>If the item is in the following format:</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>[actor] agrees to [action]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Classify the item as a <strong>commitment</strong>.</td>
<td></td>
</tr>
<tr>
<td>CH$_{DU,A}$ 5</td>
<td>If the item is in one of the following format:</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>- [actor] agrees to [object]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- [actors] agree to [object]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Classify the item as a <strong>commitment</strong>. The commitment has agree as the action.</td>
<td></td>
</tr>
<tr>
<td>CH$_{DU,A}$ 6</td>
<td>- If the item is in one of the following formats:</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>- [object] shall [action]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- [object] is [past participle of action]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- [object] must be [past participle of action]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- [object] will be [past participle of action]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- [object] shall be [past participle of action]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- [object] should be [past participle of action]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Classify the item as a <strong>commitment</strong> of the <strong>recipient</strong>.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- If the item is in the following format:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[object] may be [past participle of action]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>The item is an entitlement of the <strong>recipient</strong>. Classify the entitlement as a <strong>privilege</strong> or a <strong>right</strong> using CH 1.</td>
<td></td>
</tr>
<tr>
<td>CH$_{DU,A}$ 7</td>
<td>If the item is in one of the following format:</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>- [actor] agrees that [parties] shall [action]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- [actor] agrees that [parties] will [action]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>The item contains multiple commitments:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- a commitment from the actor with the action agree, and</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- commitments from each of the parties with the given action.</td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Heuristic</td>
<td>Page</td>
</tr>
<tr>
<td>------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
</tbody>
</table>
| DAH 1| Find the actor, or responsible stakeholder, by asking:  
- Who is the subject of this item?  
- Who is performing this action?  
The answer to these questions will be the actor.                                                                                       | 40   |
| DAH 2| Find the action by asking:  
What action is the actor performing?  
The answer to this question will be the action.                                                                                           | 41   |
| DAH 3| For negative actions (actions containing not), there are two separate cases—those with conditions and those without conditions.  
- If the item contains a negative action and a condition, use the contrapositive. Document the negated action (resulting in a positive action after removing the double negative) and the negated condition with only added to it.  
- If the item contains a negative action and no condition, document the negative action.                                                    | 41   |
| DAH 4| Find the object by asking:  
What is the object of the actor’s action?  
The answer to this question will be the object.                                                                                           | 42   |
| DAH 5| Find the object’s source by asking:  
- Who is the originator of the object?  
- What is the originator of the object?  
The answer to these questions will be the object’s source.                                                                                  | 42   |
| DAH 6| Find the target by asking:  
Who is the intended recipient of the actor’s actions?  
The answer to this question will be the target.                                                                                          | 42   |
| DAH 7| Find the purpose by asking:  
What is the reason for performing this action?  
The answer to this question will be the purpose.                                                                                          | 43   |
| DAH 8| If an item contains multiple purposes joined by the logical OR, then split the item into multiple items—each of which contains a single purpose.                                                       | 43   |
| DAH 9| Find the condition by asking:  
What is the restriction on performing the action?  
The answer to this question will be the condition.  
Additionally, the condition is often found within an item by identifying the following terms:  
- if  
- unless  
- when  
Additionally, the condition is often found within an item by identifying the following terms:  
- if  
- unless  
- when                                                                 | 43   |
| DAH 10| If an item contains multiple conditions joined by the logical OR, then the item is split into multiple items—each of which contains a single condition.                                                             | 44   |
| DAH 11| Examples are indicated by one of the following formats:  
- including [...]  
- For example, [...]  
- such as [...]  
- e.g., [...]  
Where [...] is the example.                                                                                                               | 44   |
<table>
<thead>
<tr>
<th>Name</th>
<th>Heuristic</th>
<th>Page</th>
</tr>
</thead>
</table>
| DRH\textsubscript{PD} 1 | If the classifiable item includes one of the following phrases:  
  - in accordance with law  
  - in accordance with regulation  
  - permitted by law  
  - permitted by regulation  
  - required by law  
  - required by regulation  
  - with applicable laws and regulations  
  - [legal text/s]  
  Document the rationale as \textbf{external} because a legal text is the basis of the action. | 45 |
| DRH\textsubscript{PD} 2 | If the classifiable item refers to children under the age of 13, then document the rationale of the classifiable item as \textbf{external} based on the Child Online Privacy Protection Act (COPPA). | 46 |
| DRH\textsubscript{PD} 3 | Document the rationale of the classifiable item as \textbf{external} when the analyst knows that the classifiable item is based on legal texts, even though the classifiable item does not explicitly state legal texts as the rationale of the classifiable item. | 46 |
| DRH\textsubscript{PD} 4 | If the rationale of the classifiable item is not documented as external using DRH\textsubscript{PD} 1, DRH\textsubscript{PD} 2, or DRH\textsubscript{PD} 3, document the rationale as \textbf{internal} because organizational practice or procedure is the basis of the action. | 47 |
Table A.10: Documenting Rationale Heuristics for DUAs (DRH$_{DU\text{A}}$)

<table>
<thead>
<tr>
<th>Name</th>
<th>Heuristic</th>
<th>Page</th>
</tr>
</thead>
</table>
| DRH$_{DU\text{A}}$ 1 | If the classifiable item includes one of the following phrases:  
  - in accordance with law  
  - in accordance with regulation  
  - permitted by law  
  - permitted by regulation  
  - required by law  
  - required by regulation  
  - specified by law  
  - specified by regulation  
  - with applicable law  
  - with applicable regulation  
  Document the rationale as **external** because a legal text is the basis of the action. | 47 |
| DRH$_{DU\text{A}}$ 2 | If the classifiable item includes one of the following phrases:  
  - requirements of HIPAA  
  - under HIPAA  
  - with HIPAA  
  Document the rationale as **external** because legal text, specifically the HIPAA, is the basis of the action. | 48 |
| DRH$_{DU\text{A}}$ 3 | Document the rationale of the classifiable item as **external** when the analyst knows that the classifiable item is based on 45 C.F.R. §164.514(e), even though the classifiable item does not explicitly state 45 C.F.R. §164.514(e) as the rationale of the classifiable item. | 48 |
| DRH$_{DU\text{A}}$ 4 | If the rationale of the classifiable item is not documented as external using DRH$_{DU\text{A}}$ 1, DRH$_{DU\text{A}}$ 2, or DRH$_{DU\text{A}}$ 3, document the rationale as **internal** because organizational practice or procedures is the basis of the action. | 49 |

Table A.11: Forming CPR Heuristic (FH)

<table>
<thead>
<tr>
<th>Name</th>
<th>Heuristic</th>
<th>Page</th>
</tr>
</thead>
</table>
| FH 1 | Form the commitment, privilege, or right using the following template:  
  [action] [object] from [object’s source] to/with [target] for/in order to [purpose] given/if [conditions] | 49 |
A.3 Operationalization Heuristics

Table A.12: Requirement Type Heuristic (RTH)

<table>
<thead>
<tr>
<th>Name</th>
<th>Heuristic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>RTH 1</td>
<td>If the CPR is a capability of the system or software, then operationalize it as a system requirement. Otherwise, operationalize it as a operational requirement.</td>
<td>50</td>
</tr>
</tbody>
</table>

Table A.13: Operationalizing Heuristics (OH)

<table>
<thead>
<tr>
<th>Name</th>
<th>Heuristic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>OH 1</td>
<td>Given a commitment that should be operationalized as a system requirement based on RTH 1, operationalize it as a system requirement using the following template: The system shall require the [actor] to [commitment].</td>
<td>50</td>
</tr>
<tr>
<td>OH 2</td>
<td>Given a commitment that should be operationalized as a system requirement based on RTH 1, operationalize it as a operational requirement using the following template: The [actor] shall [commitment].</td>
<td>51</td>
</tr>
<tr>
<td>OH 3</td>
<td>Given a privilege or a right that should be operationalized as a system requirement based on RTH 1, operationalize it as a system requirement using the following template: The system shall allow the [actor] to [privilege/right].</td>
<td>51</td>
</tr>
<tr>
<td>OH 4</td>
<td>Given a privilege or a right that should be operationalized as a system requirement based on RTH 1, operationalize it as a operational requirement using the following template: The [actor] shall be able to [privilege/right].</td>
<td>52</td>
</tr>
</tbody>
</table>
Appendix B

User Study Materials & Data
Figure B.1: User Study Questionnaire
1. Did your analysis enable you to identify, uncover, or understand an aspect of the problem that you did not acquire when you first read the policy document in Section 3? If so, please explain.

2. Have you had prior experience performing a software problem analysis? Which methods of analysis have you used? How much did you draw upon from your prior experience in solving this problem?

3. How much time did you spend performing your analysis? (in hours or portions of hours)

**Figure B.2: Problem Analysis Questions**

1. Did writing the requirements enable you to identify, uncover, or understand an aspect of the problem that you did not acquire during your initial problem analysis exercise? If so, please explain.

2. Did you find the analysis method you used useful for clearly identifying requirements? If so, how? If not, what aspect was difficult?

3. How much did you draw upon from your problem analysis in writing your requirements specification document? What aspects were straightforward? What aspects were confusing (perhaps requiring a bit of guess work)?

4. How much time did you spend performing your analysis? (in hours or portions of hours)

**Figure B.3: Requirements Analysis Questions**
### Information We Receive

**Information you provide to us:**

**Information About Yourself.** When you sign up for SocialNet, you provide us with your name, email, gender, and birth date. During the registration process we give you the opportunity to connect with your friends, schools, and employers. You will also be able to add a picture of yourself. In some cases we may ask for additional information for security reasons or to provide specific services to you. Once you register you can provide other information about yourself by connecting with, for example, your current city, hometown, family, relationships, networks, activities, interests, and places. You can also provide personal information about yourself, such as your political and religious views.

**Content.** One of the primary reasons people use SocialNet is to share content with others. Examples include when you update your status, upload or take a photo, upload or record a video, share a link, create an event or a group, make a comment, write something on someone’s Wall, write a note, or send someone a message. If you do not want us to store metadata associated with content you share on SocialNet (such as photos), please remove the metadata before uploading the content.

**Transactional Information.** We may retain the details of transactions or payments you make on SocialNet. If you do not want us to store your payment source account number, you can remove it using your payments page. Information we collect when you interact with SocialNet:

**Site activity information.** We keep track of some of the actions you take on SocialNet, such as adding connections (including joining a group or adding a friend), creating a photo album, sending a gift, poking another user, indicating you “like” a post, attending an event, or connecting with an application. In some cases you are also taking an action when you provide information or content to us. For example, if you share a video, in addition to storing the actual content you uploaded, we might log the fact that you shared it.

**Access Device and Browser Information.** When you access SocialNet from a computer, mobile phone, or other device, we may collect information from that device about your browser type, location, and IP address, as well as the pages you visit.

**Cookie Information.** We use “cookies” (small pieces of data we store for an extended period of time on your computer, mobile phone, or other device) to make SocialNet easier to use, to make our advertising better, and to protect both you and SocialNet. For example, we use them to store your login ID (but never your password) to make it easier for you to login whenever you come back to SocialNet. We also use them to confirm that you are logged into SocialNet, and to know when you are interacting with SocialNet Platform applications and websites, our widgets and Share buttons, and our advertisements. You can remove or block cookies using the settings in your browser, but in some cases that may impact your ability to use SocialNet.

---

*SocialNet is a social networking site*

**Figure B.4:** SocialNet Policy Document
Table B.1: User Study Data Discussed in Section 6.3

<table>
<thead>
<tr>
<th>Subject</th>
<th>SPR</th>
<th>ER</th>
<th>Cor</th>
<th>Com</th>
<th>NC</th>
<th>Mul</th>
<th>PoV/Mul</th>
<th>NP</th>
<th>D</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPR-A</td>
<td>31</td>
<td>21</td>
<td>0.71</td>
<td>0.288</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>22</td>
<td>9</td>
</tr>
<tr>
<td>CPR-B</td>
<td>28</td>
<td>21</td>
<td>0.786</td>
<td>0.288</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>26</td>
<td>2</td>
</tr>
<tr>
<td>CPR-C</td>
<td>24</td>
<td>24</td>
<td>0.583</td>
<td>0.329</td>
<td>6</td>
<td>17</td>
<td>0</td>
<td>8</td>
<td>15</td>
<td>9</td>
</tr>
<tr>
<td>CPR-D</td>
<td>21</td>
<td>16</td>
<td>0.524</td>
<td>0.219</td>
<td>6</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>CPR-E</td>
<td>19</td>
<td>17</td>
<td>0.789</td>
<td>0.233</td>
<td>12</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>2</td>
</tr>
<tr>
<td>CPR-F</td>
<td>19</td>
<td>26</td>
<td>1</td>
<td>0.356</td>
<td>9</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>19</td>
<td>0</td>
</tr>
<tr>
<td>CPR-G</td>
<td>17</td>
<td>24</td>
<td>1</td>
<td>0.329</td>
<td>12</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>CPR-H</td>
<td>16</td>
<td>12</td>
<td>0.563</td>
<td>0.164</td>
<td>6</td>
<td>7</td>
<td>0</td>
<td>1</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>CPR-I</td>
<td>15</td>
<td>23</td>
<td>1</td>
<td>0.315</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>CPR-J</td>
<td>15</td>
<td>26</td>
<td>0.933</td>
<td>0.356</td>
<td>6</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>CPR-K</td>
<td>14</td>
<td>27</td>
<td>0.929</td>
<td>0.37</td>
<td>8</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>CPR-L</td>
<td>12</td>
<td>24</td>
<td>0.75</td>
<td>0.329</td>
<td>4</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>GB-A</td>
<td>27</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>23</td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>GB-B</td>
<td>26</td>
<td>21</td>
<td>0.308</td>
<td>0.288</td>
<td>1</td>
<td>11</td>
<td>0</td>
<td>7</td>
<td>8</td>
<td>18</td>
</tr>
<tr>
<td>GB-C</td>
<td>22</td>
<td>2</td>
<td>0.136</td>
<td>0.027</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>8</td>
<td>14</td>
</tr>
<tr>
<td>GB-D</td>
<td>22</td>
<td>3</td>
<td>0.136</td>
<td>0.041</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>3</td>
<td>19</td>
</tr>
<tr>
<td>GB-E</td>
<td>19</td>
<td>9</td>
<td>0.316</td>
<td>0.123</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>GB-F</td>
<td>15</td>
<td>22</td>
<td>0.6</td>
<td>0.301</td>
<td>6</td>
<td>10</td>
<td>0</td>
<td>3</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>GB-G</td>
<td>15</td>
<td>21</td>
<td>0.733</td>
<td>0.288</td>
<td>8</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>4</td>
</tr>
<tr>
<td>GB-H</td>
<td>14</td>
<td>4</td>
<td>0.214</td>
<td>0.055</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>8</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>GB-I</td>
<td>10</td>
<td>7</td>
<td>0.4</td>
<td>0.096</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>GB-J</td>
<td>8</td>
<td>5</td>
<td>0.625</td>
<td>0.068</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>GB-K</td>
<td>7</td>
<td>11</td>
<td>0.429</td>
<td>0.151</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>NMA-A</td>
<td>53</td>
<td>34</td>
<td>0.509</td>
<td>0.466</td>
<td>10</td>
<td>7</td>
<td>2</td>
<td>6</td>
<td>31</td>
<td>22</td>
</tr>
<tr>
<td>NMA-B</td>
<td>28</td>
<td>25</td>
<td>0.321</td>
<td>0.342</td>
<td>3</td>
<td>13</td>
<td>0</td>
<td>5</td>
<td>9</td>
<td>19</td>
</tr>
<tr>
<td>NMA-C</td>
<td>27</td>
<td>16</td>
<td>0.333</td>
<td>0.219</td>
<td>4</td>
<td>8</td>
<td>0</td>
<td>12</td>
<td>10</td>
<td>17</td>
</tr>
<tr>
<td>NMA-D</td>
<td>26</td>
<td>21</td>
<td>0.423</td>
<td>0.288</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>10</td>
<td>11</td>
<td>15</td>
</tr>
<tr>
<td>NMA-E</td>
<td>22</td>
<td>16</td>
<td>0.455</td>
<td>0.219</td>
<td>6</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>NMA-F</td>
<td>19</td>
<td>8</td>
<td>0.368</td>
<td>0.11</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>7</td>
<td>12</td>
</tr>
<tr>
<td>NMA-G</td>
<td>15</td>
<td>18</td>
<td>0.4</td>
<td>0.247</td>
<td>3</td>
<td>14</td>
<td>0</td>
<td>5</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>NMA-H</td>
<td>15</td>
<td>6</td>
<td>0.333</td>
<td>0.082</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>6</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>NMA-I</td>
<td>15</td>
<td>13</td>
<td>0.6</td>
<td>0.178</td>
<td>8</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>NMA-J</td>
<td>10</td>
<td>6</td>
<td>0.3</td>
<td>0.082</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>NMA-K</td>
<td>6</td>
<td>9</td>
<td>0.833</td>
<td>0.123</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>

Key - SPR: Subject-Produced Requirements, ER: Expected Requirements to which Subject-Produced Requirements Related, Cor: Correctness, Com: Completeness, NC: Related Pairs with No Coding, Mul: Related Pairs with Mul Coding, PoV/Mul: Related Pairs with PoV/Mul Coding, NP: Subject-Produced Requirements with NP Coding, D: Derived Requirements, S: Synthesized Requirements
Appendix C

DUA Template

This appendix provides the DUA template created based on our CPR analysis and understanding of DUAs as specified in the HIPAA Privacy Rule and related regulations. Guidelines about DUAs, which were provided by the HealthSystem project’s legal counsel, were helpful in structuring and wording the template. The bolded, italicized items within the template represent the items that the recipient and/or data custodian will replace with the specifics of the agreement.

Data Use Agreement (DUA)

This Data Use Agreement, hereinafter DUA, establishes the terms and conditions under which (RECIPIENT) (Data Recipient or simply Recipient) will obtain and use a limited data set (Limited Data Set or LDS) from (DATA CUSTODIAN) (Data Custodian). This DUA is effective on (DATE).

1. LIMITED DATA SET

The LDS will include the following elements: (LIST OF ELEMENTS).

The LDS does not include the following sixteen identifiers: (i) Names; (ii) Postal address information, other than town or city, State, and zip code; (iii) Telephone numbers; (iv) Fax numbers; (v) Electronic mail addresses; (vi) Social security numbers; (vii) Medical record numbers; (viii) Health plan beneficiary numbers; (ix) Account numbers; (x) Certificate/license numbers; (xi) Vehicle identifiers and serial numbers, including license plate numbers;(xii) Device identifiers and serial numbers; (xiii) Web Universal Resource Locators (URLs); (xiv) Internet Protocol (IP) address numbers; (xv) Biometric identifiers, including finger and voice prints; and (xvi) Full face photographic images and any comparable images.

2. USE OR DISCLOSURE
(a) The Data Custodian may use or disclose the LDS only for purposes of research, public health, or health care operations.

(b) The Recipient may use or disclose the LDS for the following purposes: (PURPOSES). The Recipient may not use or further disclose the LDS in a manner that would violate this DUA if done by the Data Custodian.

(c) The following individuals may use or receive the LDS: (INDIVIDUALS). No portion of the LDS shall be disclosed to any person other than those who are under the direct supervision of the investigator who has signed the investigator statement on the signature page of this DUA.

3. RESTRICTIONS ON USE OR DISCLOSURE

The Recipient agrees that it, its employees and any other entities and individuals to whom it discloses the LDS, shall be bound to these terms and conditions:

(a) The Recipient will only use or further disclose LDS for limited purposes.

(b) The Data Custodian may use PHI to create an LDS, whether or not the LDS is to be used by the Data Custodian. The Data Custodian may disclose PHI only to a business associate to create an LDS, whether or not the LDS is to be used by the Data Custodian.

(c) The Recipient will not use or further disclose the LDS other than as listed within this DUA or as required by law.

(d) The Recipient will use appropriate safeguards to prevent use or disclosure of LDS other than as provided by this DUA. ((OPTIONAL) LIST OF SAFEGUARDS)

(e) The Recipient will report to Data Custodian if it becomes aware of any use or disclosure of LDS not provided by this DUA.

(f) The Recipient will not identify or contact the individuals within the LDS.

(g) The Data Custodian will discontinue disclosure of LDS to Recipient and report problem to the Secretary, Department of Health and Human Services, if the Recipient breaches the DUA.

4. TERM AND TERMINATION

(a) This DUA shall become effective on the date fully executed by both parties below and shall continue as long as Data Recipient retains any copies of the LDS, unless otherwise terminated pursuant to the termination provisions below or by applicable law or regulation.
(b) Upon expiration or termination of this DUA, Data Recipient shall return or destroy all copies of the LDS maintained by Data Recipient and shall provide written notice to Data Custodian.

(c) Should Data Recipient commit a material breach of this DUA that is not cured within thirty (30) days after Data Recipient receives notice of such breach from Data Custodian, Data Custodian will discontinue disclosure of LDS and may report the breach to the Secretary, Department of Health and Human Services, if appropriate pursuant to federal regulation.

(d) The obligations of Data Recipient under Section 3 of this DUA shall survive termination of this DUA.

Data Custodian Name

Data Custodian Address

Data Recipient Name

Data Recipient Address

Data Custodian Signature

Data Recipient Signature