

A Comprehension Approach for Formalizing Legal Text: A Decision Tree Model Approach for Privacy Rules of HIPAA

Imran Khan, Moheeb Alwarsh & Javed I. Khan

*Department of Computer Science
Kent State University, Kent, Ohio 44240, USA*

imran.khan@iu.edu.pk / malwarsh@kent.edu / javed@kent.edu

Key Words—HIPAA, Privacy Rules, Formalization, Logical rules set

Abstract

Use of an information system to assist and guide practitioners to achieve HIPAA compliance can greatly benefit the US health care system. However, a critical step in it is how to formalize HIPAA legal text to help machine processing. It is not a trivial task. Previous approaches attempted to form rules from clause and have seen very limited success- often showing ambiguity and imprecision because of the complexity of HIPAA text structure. We propose a novel approach based on deeper modelling of HIPAA world. The technique is based on one of the first of its kind- a model of the complete conceptual space (actors/action/decision/constraints) in which the original HIPAA Privacy Acts has been defined in terms of an Entity Relation Action (ERA) model. The clauses of HIPAA legal text is then converted into a logical rule set involving only the elements from this ERA model. These rules are then integrated into a disambiguated decision tree (DDT) precisely identifying the allowed and prescribed actions. Given any EMR query the DT then enables one not only to verify the compliance as well as provide complete release guidance as prescribed by HIPAA, generate explanation and audit. The technique can usher a whole new range of associated benefits for large scale Healthcare System

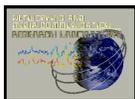
Index Terms—HIPAA Act, Privacy Rules, Formalization, Logical rules set

1. INTRODUCTION

The U.S. Department of Health and Human Services (“HHS”) in 1996 created the Health Insurance Portability and Accountability Act (HIPAA) as a means of providing a mechanism to protect civil rights when sharing patients’ medical health information and we will refer to this information as protected health information. Failing in conforming to the HIPAA Act may result in a fine up to \$25,000 per year and between 1 to 5 years in prison [4, 5, 6]. In 2009, Congress further amended HIPAA with the HITECH Act further addressing the concerns associated with the electronic transmission of health information. It was passed as a part of the American Recovery and Reinvestment Act, President Obama’s first major legislative initiative upon taking office. [7]. HIPAA Administrative Simplification, Regulation Text: 45 CFR Parts 160, 162, and 164 [8] regulate the use and disclosure of personal health information.

HIPAA defines how a *covered entity*- which includes Health Plans, Health Care Clearinghouse, or a Health Care Provider, Hospital, etc. who can share the protected health information in under various circumstances meeting the often conflicting needs of doctors, hospitals, patients, insurers, employers, researchers, and other myriads of health and medical service providers. The law covers *protected health information* that includes all individually identifiable health information that can be transmitted or maintained in electronic or any other kind of media.

The length of law is quite extensive and delves into finance, accounting, amendment rights, and even standards and specification of service such as how the information to be handed over. The complexity of the act itself and the organization of the legal text often make it very difficult for practitioners to determine whether they are in compliance or not [9]. The scope of HIPAA is also remarkable. Unlike any other mass databases the growths of medical record databases are phenomenal. Almost every citizen in developed world today has active medical records and with the emergence of electronic systems these records are exponentially growing. These also need to be routinely exchanged between myriads of entities. The sheer scale and complexity calls for increased automation



that can provide practitioners guidance for HIPAA compliance. Today HIPAA requires experts with deep familiarity with various intricate provisions of HIPAA to verify compliance. Often, compliance is managed by placing grossly simplified administrative process flow for set cases. Unfortunately, none of these practices are scalable or cost competitive. The existing release practices based on pre-set flow tends to be overly restrictive than actual HIPAA would allow due to the cautious implementations. Often it requires patients to sign-off (and waive) broader rights than required for their treatment- essentially defeating the very purpose of HIPAA.

Particularly if we regularized the HIPAA Act it looks very difficult and complex for the inexperienced person due to several reasons. For example the law generally allows protected information to be shared between appropriate entities for the purpose of treatment. However, clause 164.508.a.2 [8], seems to contradict this by stating that “if the protected information is a psychotherapy note then a covered entity, i.e., a health plan, a health care provider or a clearinghouse, must obtain an authorization before disclosure”. Thus simple reasoning based on actions allowed by one portion of the law, without accounting for prohibitions in other portions of the law, might provide inaccurate result [10].

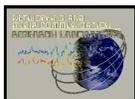
The complexity of HIPAA, combined with potentially stiff penalties for violators, has lead physicians and medical centres to withhold information from those who may have a right to it. A review of the implementation of the HIPAA Privacy Rule by the U.S. Government Accountability Office found that health care providers were “uncertain about their legal privacy responsibilities and often responded with an overly guarded approach to disclosing information than necessary to ensure compliance with the Privacy rule [17].

Complying with laws and regulations is challenging, because legal texts contain ambiguities, cross-references to sections of the same or different legal texts, and possibly conflicting definitions and domain-specific terminology [13]. In addition, laws and regulations undergo updates and amendments, requiring software engineers to manage and track these changes [13]. Also, in legal systems implementation of the acts gets refined gradually as its various provisions are tested in contests and courts provides case specific clarifications.

Cross-references to external legal texts would be explored to obtain additional software requirements. Unfamiliar engineers with laws that are governing a domain would require some tools and techniques to identify compliance requirements [11]. As a result, rules would be used with logical operators to make a relationship between the requirements and the output [14]. Each rule is represented as an if-then statement. Many of these rules are combined to create a complete result for a query.

Several studies proposed solutions to formalize HIPAA legal text into some form of logic rule set. In last decades, general attempts have been made to convert legal text into logic rules [18, 19]. More recently there is renewed interest to tackle HIPAA. In [10], the authors examined sections of HIPAA and investigated if Datalog like stratified first order system of logic can be instituted to verify compliance of a medical information release request messages sent by providers. In the process of interpreting the legal text they also observed extensive “conflicts” as well as “anomalies” regarding lack of regulation in HIPAA. The proposed stratified Datalog with limited use of negation technique for ensuring termination and efficiency. Their proposed mechanism combines associated rules in the form of “permitted by” and “forbidden by” where the later has precedence for making a decision. In the cases of perceived ambiguity the system has been biased where prohibition takes precedence over permission. The resolution process seems –such as use of negation seems to inject additional semantics not explicit in HIPAA¹. We found that this method doesn’t produce precise results in all cases as discussed later on. In [1], authors use production rule model to verify HIPAA compliance. They have classified rules to four types; rights, obligations, permissions and definition. The problem with this approach is its deficiencies in resolving overlapping conditions between two obligations. In [16], the authors presented the concept of positive and negative norms to take a decision. The first means a transmission that might occur, where the second means a transition that must occur. All negative norms must be satisfied to release protected health information but precision of this solution is not

¹ Its support is more based on contextual integrity theory of Nissenbaum [20].



accurate because we found cases that contradict this theory. We propose a new methodology to formalize legal texts and eventually facilitate algorithmic HIPAA conformant sharing of medical information.

It seems one the basic problem with all the previous approaches is the lack of a clearly defined overall context in which the HIPAA legal Act has been framed. HIPAA- defined in 1996 did not anticipate machine processing and has been defined on the assumption of a domain expert who will be familiar with the general context of the rule. Subsequently the first generation attempts to formalize the rules also depended on surface semantic structure of the legal text. In fact this lack of general HIPAA model has also created appearance of some of the ambiguity cited in earlier literature during machine processing. Some of these ambiguities are not ambiguity when examined by a human expert.

Our Approach:

We are different from the other because in this research we present an alternate approach that instead of *declarative translation* of HIPAA text emphasizes *semantic comprehension* of HIPAA before logical rule generation. We attempt to capture and accommodate deeper underlying semantics of the complex aspects of health information sharing, for that approach we have to start one step back. Unlike others we first construct the Entity Relationship Model (ERM) and it includes the entities (actors, and their relationships)- medical entities, records, actions, rights- etc., that defines the semantics of the domain on which the HIPAA Act and their provisions have been laid and structured. Based on the HIPAA World ERM and generated concept categories we convert the corpus of legal texts into a set of logical constraints and actions. These rules are then integrated into a disambiguated decision tree (DDT) precisely identifying the allowed and prescribed actions. Given any EMR query the DDT then enables one not only to verify the compliance as well as provide complete release guidance as prescribed by HIPAA, generate explanation and audit. The overall process is explained in Fig.1.

The resulting system generates much precise decision and detailed guidance. This is no surprise. Because our model requires the designer to explicitly comprehend and extract the connections (with HIPAA experts) and summarize the overall behaviour as a set of constructed rules and subsequent DDT. The system pre-resolves the semantic long connections, and as we will see thus generates much precise resolutions. Give an EMR transaction request thus the *decision trees* much precisely resolves them. Also, the resulting system can much better articulate other intents of HIPAA such specify how to release particular piece of information, if denied what are the alternate options, generate logically coherent explanation supporting the decisions conforming, etc. conforming to the original expectations of HIPAA. Of course the entire process can be subsequently automated, edited and evolved. While in other above mentioned approaches have the lack of understanding, which only shows the result for deny or disclose with rule referenced without any guide line or explanation.

We demonstrate our overall approach by modelling provisions of section 164 of HIPAA [12], which is related to the security and privacy issues of health care. Section 164 covers the general provisions rules and security related standards for exchanging PHI. It consists of 683 non repeated clauses and we covered all clauses starting from 164.502 up to 164.530 in this study.

2. WORLD RULES MODEL OVERVIEW

We present a new methodology of formalizing legal text in order to use it information system. In an effort to advance health care privacy and exchange of protected health information, we have created a road map of how we could formalize privacy rules of HIPAA Act to logical rules in order to facilitate the implementation of exchanging protected health information between different entities. The first part of this process requires human pre-processing of legal text where the second part requires computer based processing.

The first part consists of several steps and it begins by generating different concept classes from privacy rules of HIPAA Act, extract information and distribute them among these concept classes. Each rule or clause in these concept classes will be identified by a tag and they will be connected together based on how a request of disclosing protected health information is processed in privacy rules of HIPAA Act. Nevertheless, different sections



of privacy rules might consist of related information needs to be considered in this process. At the end, we assemble information together in a rule set to create logical rules that govern the implementation of input and output, see the human based process in Figure 1. Generated information in the Rule Set will be used to process requests in the second part of the diagram, see below part of Figure 1. This part is an automated part where information system can be used to process requests.

Transponder evaluates information from requesters, which includes but not limited to, identity, purpose, type of information and authorization with generated Rule Set from the first part of Figure 1. Whereas, the transponder will work as a compliance checking between requested information from one side and available patient medical data from the other side. Furthermore, five outputs will be generated as a result of the decision made by transponder. The first output is “decision” which is used to decide if the request would be denied or not. Then, how information is disclosed with special instruction, for example protected health information will be disclosed by email, format of data, summary of record, fee, time etc. Also, what information will be disclosed with a list of deliverables, like de-identified patients records. In addition, we would be able to produce audit which indicates which rules triggered and explanation of how these rules evaluated.

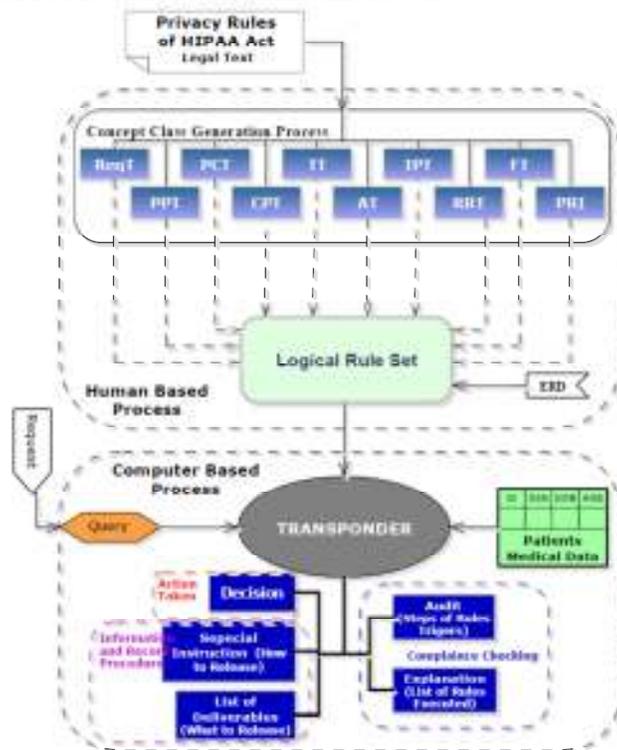
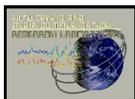


Fig. 1. HIPAA Privacy Detailed World Rules Model

3. LEGAL TEXT TO LOGICAL RULE SET PROCESS

Converting legal text to logics rules set, requires a full understand of how information is processed logically. This would require a conceptual view of how privacy rules of HIPAA Act consists of different data types that need to be integrated in certain way to comply with proper implementation of these rules. To understand how rules are formed logically, pre-processing of legal text for different data types will be discussed in the following sub-sections.

We need to understand how privacy rules structured, interrelated and overlapped. For example, why do we need a law to govern the release, exchange and use of information? What is the purpose of the law? Who is



responsible for implementing it? In what conditions can this law be used? What will be the action taken? How to respond to requesters? We can conclude that there are some reasons or purposes for laws and there are some conditions for these purposes. Also, for each condition there is a response and action. In other words, we need to cover all aspects of legal text of privacy rules and create Concept Classes. Each Concept Class will contain related information. Privacy rules of HIPAA Act are divided into different sections and each section contains clauses. For example, clause 164.506.C.1 of privacy rules that belong to section 164.506 states

“A covered entity may use or disclose protected health information for its own treatment, payment, or health care operations”.

We could extract several pieces of information from this clause. For example, a “covered entity” is a requester of information, “treatment, payment or health care operations” are purposes for disclosing protected health information, and “its own” is a pre-condition for using or disclosing protected health information. All requesters are grouped under one Concept Class for this section, conditions and pre-conditions are also grouped in separate Concept Classes.

Based on our understanding of the privacy rules, we found 10 types of Concept Classes and some of these classes available only in certain sections. As a result, we created a generalized version to be used in all privacy rules sections. Whereas, each section of privacy rules of HIPAA act will generate 10 concept classes. Each concept class consist of legal text from different clauses. To distinguish between these clauses in each concept class, we have assigned a tag for each clause, see Figure 1.1.

Tag	Description of Concept Classes
ReqT	Request Class: This class contains tags used to identify requesters of information role. For example researcher
PCT	Pre-Condition Class: all prerequisites that need to be satisfied before evaluating requests are collected under this class. For example, if authorization is available or not.
PPT	Purpose Class: Purposes for disclosing protected health information.
CPT	Conditional Purpose Class: All rules for evaluating privacy rules of HIPAA Act with PPT, PCT and ReqT will be under this class.
AT	Action Class: Atomic action that is produced as a result of evaluating each request.
TT	Time Class: Time Required for processing a request. For example, protected health information will be released after 30 days to de-identify this information
RRT	Record Release Class: Information that will be released as a result of a request.
IPT	Information Procedure Class: Rule to Inform how information will be release. For example information will be released with a fee that needs to be paid.
FT	Fee Class: Rules that identify non-free to release protected health information.
PRI	Patient Record Item Class: Medical and none medical records related to patients.

Fig. 1.1. Concept class description

3.1. ER Model

To make a relationship between tags in concept classes for each section, we need to create entity relationship diagram to connect these concept classes together based on how information logically flow. Each request for disclosing protected health information must conform to this diagram, see Figure 2. For example, if a researcher wants to disclose protected health information, he/she would initiate a *request* (1) to a *covered entity* (2) that manages *patient information* (3), *patients record items* (4), *what* will be released (9) and *how* they will be released (12). This request will include *requester* information (1) and the *purpose* of this request (5). Covered entity (2) will take an *action* (6) based on the purpose of researcher (5). Then include the *time* (10) and *fee* (11) for releasing information. There are two types of *conditions* (7 and 8) must be considered when evaluating a requests (1). The



first (7) is to make sure that a researcher conforms to privacy rules of HIPAA Act for information requester before moving further in the evaluation process (like having a proper and valid authorization). The second (8) is to make sure that requester, purpose of request and requested information in compliance with privacy rules of HIPAA Act.

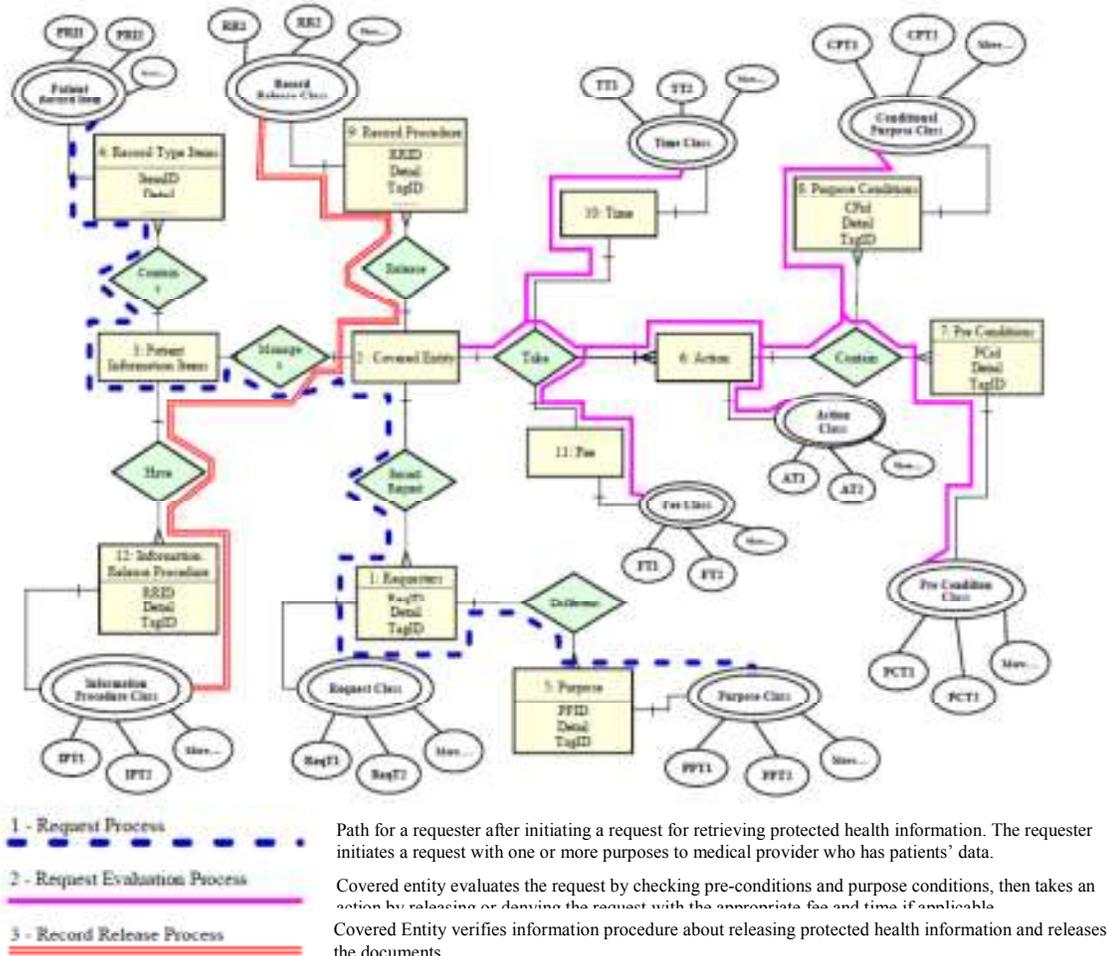
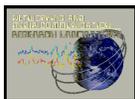


Fig. 2. Entity Relation Diagram for HIPAA privacy rules

3.2. Legal text to Tags Generation

These Concept Classes are organized as follow; First Concept Class is related to requesters. For example, we search in clauses for requesters of information. Then, each requester will be assigned a tag (ReqT1, ReqT2 ...etc) and added under requesters' concept class. For example, clause 164.512.(H).1 has one requester which is a researcher, see Table 1. Conditions that need to be satisfied before sending requests are placed under pre-conditions concept class and each pre-condition rule is assigned a (PCT) tag. For example, a requester of information need to have a role of researcher or be a part of a covered entity before sending a request, see Table 3. Meanwhile, all record items in each section of privacy rules need to be evaluated for dependency. For example psychotherapy notes cannot be disclosed without an authorization from individual as stated in clause 164.508.a.2. This indicates that a condition needs to be satisfied for this type of record item. We mark all record items that have dependencies and put them in one table, see Table 10.

All purposes of requests, which are related to requesters, indicate reasons for disclosing protected health information are listed under purpose concept class and each rule is assigned a (PT) tag, see Table 2. Also,



conditions for denying or disclosing protected health information referred to as conditional purpose. All rules are placed under conditional purpose concept class with (CPT) tag. An example is when a requester wants to disclose protected health information, we check if he/she has a proper authorization, see Table 4. Once a condition evaluated, an action needs to be taken to whether deny or disclose protected health information. We collect these actions under actions concept class and each action assigned a (AT) tag, see Table 6. Time class indicates the time needed to release protected health information. Rules are assigned (TG) tags, see Table 5. Record Release concept class is related to the rules that indicate what type of information will be released. For example psychotherapy notes. (RRT) tag is used to distinguish between rules in this class, see Table 9. Information Procedure concept class represents how information will be released and (IPT) tag used in this class, see Table 8. Finally Fee concept class represents the fee required to release protected health information and (FT) tag is used for rules in this concept class, see Table 7. As a result, extracted information from sections will be distributed among 10 concept classes and decision will be based on combinations of tags from these concept classes. Note; fee class and time class might not be available in all privacy rules sections of HIPAA Act.

HIPAA Legal Text For Researcher	Clause Ref. #	Tag ID
Is the request from a researcher?	(164.512) i. 1	ReqT1

Table-1. Request Tags. ReqT Example (Researcher)

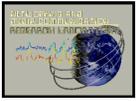
HIPAA Legal Text For Researcher	Clause Ref. #	Tag ID
Is there an authorization for this researcher?	(164.508) (b).3.i	PCT1
Is the authorization expired?	(164.508) (c).1.v	PCT2
Is there any condition placed by covered entity on this research?	(164.508) (b).4.i	PCT3
Does the research meets conditions in PCT3?	(164.508) (b).4.i	PCT4

Table-3 Pre Conditions Tags. PCT Example (Researcher)

Conditional Purpose Tags	Clause Ref. #	Tag ID
Is waiver available?	(164.512) i. 1. i	CPT1
Is there a brief description from the researcher about this research?	(164.512) (i). 2. iii	CPT2
Are the minimum requirements for documents related to research satisfied?	164.514 (d).iii.D	CPT3
Is the protected health information necessary for the research purposes?	(164.512) i. 1. iii.C	CPT7
Did the covered entity obtain consents for CPT9, CPT10 and CPT11?	(164.512) i.1.ii	CPT8
Is the use or disclosure sought solely to review protected health information as necessary to prepare a research protocol or for similar purposes preparatory to research?	(164.512) i.1.ii.A	CPT9
Does the research intend to remove protected health information from the covered entity?	(164.512) i.1.ii.B	CPT10
Is the use of protected health information necessary for the research purposes?	(164.512) i.1.ii.C	CPT11

Table-4 Conditional Purpose Tags. CPT Example (Researcher)

Time Tags	Tag ID
Within 3 days	TT1
Within 7 days	TT2
Within 30 days	TT3
Within 60 days	TT4



Extension of Time up to 10 days	TT5
Extension of Time up to 15 days	TT6
Extension of Time up to 30 days	TT7
Time Extension Reason in Written	TT8
Less than 6 year	TT9
Prior of 6 year	TT10

Table- 5 Time Related Tags. TT Example (Researcher)

Actions Tags	Tag ID
Denial	AT1
Unreviewable Denial	AT2
Reviewable Denial	AT3
Release	AT4
Review on Denial	AT5
Update	AT6
Temporarily Suspend	AT7

Table-6 Action Tags – AT Example

Fee Tags	Clause Ref. #	Tag ID
Cost of document preparation Fee	164.524 (c) 4	FT1
Copying of Document	164.524 (c) 4.i	FT2
Postage Fee	164.524 (c) 4.ii	FT3
No Fee	164.524 (c) 4	FT4

Table-7 Fee Related Tags FT Example

HIPAA Legal Text For Researcher	Clause Ref. #	Tag ID
Release protected health information as mentioned in the waiver.	(164.512) i. 2	IPT1
Disclose protected health information based on Individual preferences	164.532 (a)	IPT2
Restrict disclosing information for users who specifically restricted disclosing protected health information for research.	164.532 (b)	IPT3

Table-8 Information Procedure Tags - IPT Example (Researcher)

HIPAA Legal Text For Researcher	Clause Ref. #	Tag ID
Limited data set for the purposes of research, public health, or health care operations must be disclosed as a default.	164.514 (e).3.i	RRT1
Protected health information needs to be delivered in a media based on request		RRT2

Table-9 Record Release Tags - RR Tags Example (Researcher)

Record Items – PRI	PRI-Tag	PRI Status
HIV	PRI1	Unconditional
Blood Cancer	PRI2	Unconditional
Chest Cancer	PRI3	Unconditional



Billing Information	PRI4	Conditional
Psychotherapy Notes	PRI5	Conditional

Table-10 Record Items – Example

3.3. Explanation of World Rule Model

By refereeing to Figure 1, transponder consists of several steps to process information in a certain sequence. These steps consist of five stages. The first stage is to receive a request. The second stage is to process information from the request. Then, conditions are evaluated and tested to take the proper decision. Once a decision is made, information will be released with the required release time and fee if applicable, see Figure 3.

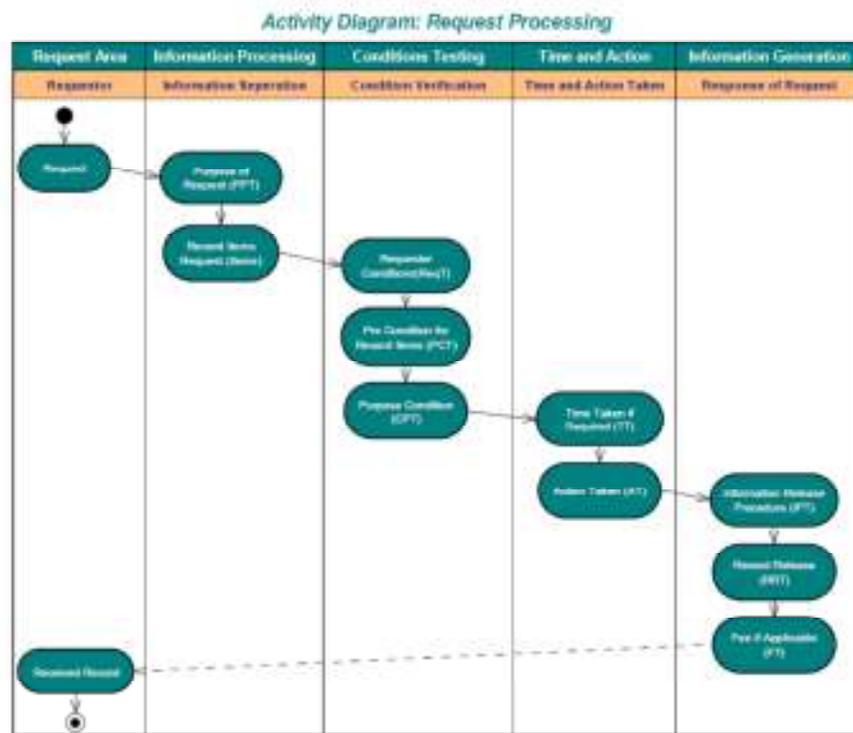


Fig 3. Request Processing by Transponder

We focused on certain clauses related to researchers from section 164.508, 164.512, 164.514 and 164.532 to explain the process of converting legal text to logical rules. We will use tables 1 to 10 which represent the outcome of combining rules from concept classes that belong to these sections. Figure 4 shows how different tags are combined. Rules will be generated based on all possible combinations between requesters (ReqT from Table 1), purposes (PPT from Table 2) and record items (PRI) which come with requests (Table 10). The result of this process will be associated with the appropriate pre-condition tags (PCT from Table 3) and all possible combination of conditional purpose tags (CPT from Table 4). Nevertheless, each pre-condition (PCT) must be evaluated with three pre-conditions (PCT for requesters, PCT for Purposes and PCT for Record Items) as shown in Figure 5. If requester's pre-conditions satisfied, then purposes pre-condition need to be satisfied. Once both pre-conditions are satisfied, record item pre-conditions (Table 10) need to be satisfied. For example, a PCT for researchers would an authorization, if purpose is research then there are some PCT this purpose and if record items is psychotherapy, then a PCT (specific authorization) need also to be satisfied.

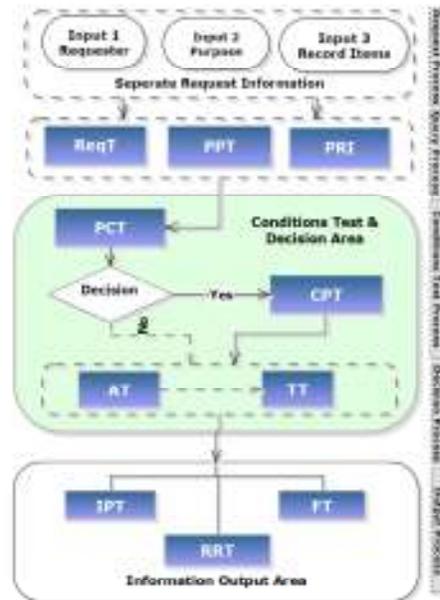
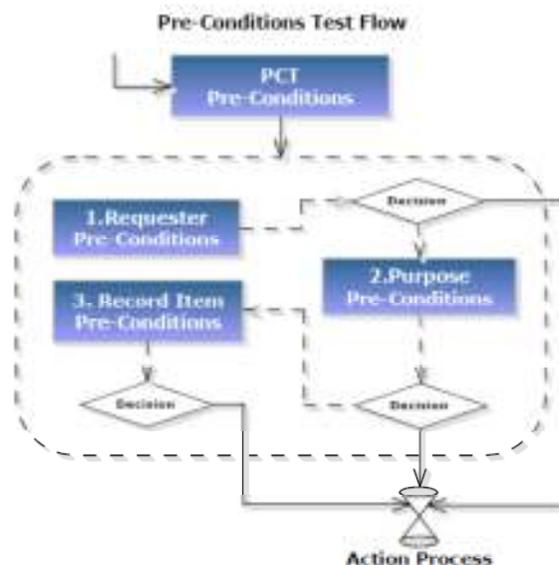


Fig.4 Combination of Rules

At this stage, a decision can be made based on the combination of the previous tags. The right action (AT) from Table 6 and time tag (TT from Table 5) will be associated with the generated rules. The final step is to combine tags related to how information will be released (IPT from Table 8), what information will be released (RRT from Table 9) and if there is any fee (FT from Table 7), see Figure 4. The outcome of this process can be seen in Figure 6. Since there isn't a fee or time associated with the selected sections of privacy rules, they will not be included in the Figure 6. Each line in Figure 6 represents one possible logical rule set generated by combining different tags as explained earlier. Patient Record Items (PRI) from Table 10 have been included in Figure 6 to represent different types of protected health information. All Patient Record items (PRI) in Table 10 can be used with generated rules set in Figure 6.



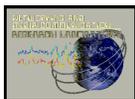


Fig.5 Combination of Pre-condition process

Generated Rules for Researcher From All Sections						
S. No	Requestor	Purpose	PRI	Condition	Action	Information Release Procedure
1	Researcher	PPT1	PRI1 PRI5	" ReqT1 " & " PCT1 ^ PCT2 ^ PCT3 " & " CPT1 ^ CPT2 ^ CPT3 ^ ~CPT4 (CPT5 ^ CPT6 ^ CPT7) ^ CPT8 (CPT9 ^ CPT10 ^ CPT11)"	AT4	" RRT1 RRT2 " & "~IPT1 ^ IPT2 ^ IPT3) "
2	Researcher	PPT1	PRI1 PRI5	" ReqT1 " & " !(PCT1 ^ PCT2 ^ PCT3) " & " CPT1 ^ !CPT2 ^ CPT3 ^ ~CPT4 (CPT5 ^ CPT6 ^ CPT7) ^ CPT8 (CPT9 ^ CPT10 ^ CPT11)"	AT4	" RRT1 RRT2 " & "~IPT1 ^ IPT2 ^ IPT3) "
3	Researcher	PPT1	PRI1 PRI5	" ReqT1 " & " PCT1 ^ PCT2 ^ PCT3 " & " !CPT1 ^ CPT2 ^ CPT3 ^ ~CPT4 (CPT5 ^ CPT6 ^ CPT7) ^ CPT8 (CPT9 ^ CPT10 ^ CPT11)"	AT1	No Waiver (CPT1)
4	Researcher	PPT1	PRI1 PRI5	" ReqT1 " & " PCT1 ^ PCT2 ^ PCT3 " & " CPT1 ^ CPT2 ^ !CPT3 ^ ~CPT4 (CPT5 ^ CPT6 ^ CPT7) ^ CPT8 (CPT9 ^ CPT10 ^ CPT11)"	AT1	Required Document are not Completed (CPT3) –
5	Researcher	PPT1	PRI1 PRI5	" ReqT1 " & " PCT1 ^ PCT2 ^ PCT3 " & " CPT1 ^ CPT2 ^ CPT3 ^ ~CPT4 (CPT5 ^ CPT6 ^ CPT7) ^ !CPT8 (CPT9 ^ CPT10 ^ CPT11)"	AT3	The request is about Decedent and no information is provided (CPT9 CPT10 CPT11)
6	!Researcher	PPT1	PRI1 PRI5	" !ReqT1 " & " PCT1 ^ PCT2 ^ PCT3 " & " CPT1 ^ CPT2 ^ CPT3 ^ ~CPT4 (CPT5 ^ CPT6 ^ CPT7) ^ CPT8 (CPT9 ^ CPT10 ^ CPT11)"	AT1	Request from Different Entity Not the Researcher

Fig. 6. Generated Rules for Researcher

Symbols used in Figure 6 are explained as follow; '^' means "and within a rule", '&' means "and between rules", '||' is used as "or", '!' for "not", '~' means "may", '()' means sub conditions. The way how a query will be processed is in this format: **If (Requestor = "Researcher" & Purpose="Purpose Tags", & Items = "Privacy Record Items", & Pre Condition = "Pre-Conditions Tags" & ReqT Condition = "Requester Tags" & Purpose Condition="Conditional Purpose Tags") Then (Action="Action Tags", & Record Release="Record Release Tags" & Information Procedure="Information Procedure Tags", & Time Taken="Time Tags", & Fee ="Fee Tags");**

First logical rule in Figure 6 means, a requester (researcher) can disclose (AT4) all protected health information (PRI) if the purpose (PPT1) is for research, he/she must meets all preconditions " PCT1 ^ PCT2 ^ PCT3 ", must meet all condition purpose tags "CPT1 ^ CPT3 ^ CPT8", must meet all sub conditions of CPT8 (CPT9 ^ CPT10 ^ CPT11), might meet CPT4, must meets (CPT5 ^ CPT6 ^ CPT7) if CPT4 is applied. The released information will be based on "RRT1 || RRT2" (one of them must be satisfied) and "~IPT1 might satisfied ^ IPT2 must satisfied ^ IPT3 must satisfied) "

3.4. Referenced and Unreferenced Relations

One of the main important processes of our approach is how to link external unreferenced clauses between sections together in order to provide a more precise decision in denying or disclosing protected health information. For example, each request will be evaluated with all rules in all sections even if there is no direct cross reference, see Figure 7. For example, if there is information that is related to disclosing protected health information in one section but there is no cross reference rule indicating that this information shouldn't be disclosed in another section, then by implementing this approach, we will cover all referenced and unreferenced clauses in all section.



Fig. 7. Evaluation of Requests between different sections.

4. QUERY / RESULT

In this section we are presenting two examples to provide a practical explanation of how a request is processed with our World Rule Model approach.

4.1. Researcher 1st Example

In this example, we used a query (same query in two different formats; SQL and as a form in Table 11) to disclose protected health information for a researcher based on a research request. We assume that medical provider has Table 12 which consists of patients protected health information. Researcher will send a request and this request will be converted to query as follow:

Researcher request in natural language: "A researcher wants to disclose protected Health Information from Table 12 regarding HIV patients who were admitted in current year to improve public health care. He/she does not have authorization to access protected health information but granted a waiver from the medical provider to fulfil this job".

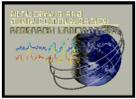
Requester Role:	Researcher	ID:	3456
Authorized By:	None	Waiver Rule:	Yes - Research
Purpose 1:	Research	Single Record Items:	HIV
Purpose 2:	None	Multi Record Items:	None
Multi Purpose:	None		
Date From:	1/1/2011	Date To:	12/12/2011
Record Format:	Default		

Table 11. Researcher Query

Query in SQL Format

Select "Patients Data" from Table13 where "Patients Data Type" = "RRI5" and Date between 20110101 and 20111212 && Requester= "Researcher" && Purpose = "Research" groupby Requester && Purpose having Waiverrule= Research_waiver && Authorization = "No"

Getting a decision after extracting the required information from the request is the first step by the transponder to process this query. To find the result of this query and what will be the generated result based on Figure 6 rule set and Figure 4 mechanism, we have 5 generated rules for researcher in Figure 6 and we will evaluate them all. If



any logical rule is satisfied, then protected information will be processed to check the instruction for disclosing protected health information and the deliverables that identify the final output. First rule (1) in Figure 6 will not be satisfied because the researcher doesn't have an authorization. In this case he/she doesn't meet the preconditions " PCT1 ^ PCT2 ^ PCT3". In second logical rule, the researcher doesn't have authorization but has a waiver (CPT1). Based on this rule, the researcher will be granted a decision to disclose protected health information but the way and type of output will be decided later based on information release procedure. Third rule will be denied because he/she doesn't have authorization or waiver. Fourth and fifth rules will be denied because CPT3 and CPT8 are not satisfied. This will be the outcome of the query " RRT1 || RRT2 " & "~IPT1 ^ IPT2 ^ IPT3 ".

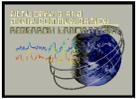
The first output of this query is the decision which will be used later to disclose protected health information based on generated special instruction and deliverables. Since we got a decision to disclose protected health information by rule number 2 from Figure 6, we will go through next step to see the list of special instructions and deliverable for each record in Table 12. Table 13 consists of the query output generated from Table 12 where each row represents the outcome of one record from Table 12.

In Table 13, preferences for patient and covered entity in the first record R.1.1 deny disclosing protected health information for research purposes. As a result, there weren't any special instructions or data to be delivered, which results in a deny decision for releasing this record. R1.3, R1.4 and R1.6 from Table 13 have special instruction for disclosing information for research. List of deliverables generated based on the special instruction in the release process. Final output can be seen in Table 14.

ID	Name	PRI	Date	Patient IPT = research	Covered Entity IPT = research	Conditional PRI Status
1	Abaad	PRI1	05/02/2011	Deny	Deny	
2	David	PRI3	12/5/2011	Disclose	Disclose	
3	Maria	PRI1	04/10/2011	Disclose	Deny	
4	Kamron	PRI1	07/11/2011	Disclose	Disclose	
5	Raja	PRI5	13/10/2011		Disclose	Authorized
6	Tena	PRI1	03/06/2011	Disclose	Disclose	
7	Nelo	PRI5	15/08/2011		Disclose	Authorized
8	Lala	PRI2	02/12/2011	Disclose	Deny	
9	Mao	PRI1	01/09/2011	Deny	Disclose	
10	Tina	PRI5	22/02/2011		Deny	Unauthoriz ed

Table 12. Patient Table

Query ID	Patient ID	Deliverables	Special Instructions	Decision
R1.1	1	DENIAL TEXT	NOT APPLICABLE	DENY
R1.2	2	DENIAL TEXT	NOT APPLICABLE	DENY
R1.3	3	REC#3,PRI1, 04/10/2011	DE-IDENTIFY RECORDS, TIMELIMIT=N/A , FEES=N/A	RELEASE
R1.4	4	REC#4,PRI1, 07/11/2011	DE-IDENTIFY RECORDS TIMELIMIT=N/A , FEES=N/A	RELEASE
R1.5	5	DENIAL TEXT	NOT APPLICABLE	DENY
R1.6	6	REC#4,PRI1,	DE-IDENTIFY RECORDS,	RELEASE



		03/06/2011	TIMELIMIT=N/A , FEES=N/A	
R1.7	7	DENIAL TEXT	NOT APPLICABLE	DENY
R1.8	8	DENIAL TEXT	NOT APPLICABLE	DENY
R1.9	9	DENIAL TEXT	NOT APPLICABLE	DENY
R1.10	10	DENIAL TEXT	NOT APPLICABLE	DENY

Table 13. Query processed for each record.

ID	PRI	Date
3	PRI1	04/10/2011
4	PRI1	07/11/2011
6	PRI1	03/06/2011

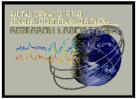
Table 14. Final Output to query

4.2. Researcher 1st Example Explanation

To show what have been triggered to release protected health information, we have created two tables that describe how record 1 and 3 from Table 12 processed. Table 15 and 16 shows detailed processing for the first and third records from Table 12 and they consist from 6 steps for processing data. Once the output is generated, snapshots can be taken from all previous steps to disclose protected health information and store this information in a log to be used later for auditing purposes.

Step 0	Request	Requester	Purpose	Record Items	Process Request
Step 1	Query	ReqT1	PPT1	PRI1	Process Query
Step 2	Pre-Conditions	PCT1 → PCT2 → PCT3 → PCT4 →		PCT for ReqT1	Y
				PCT for PPT1	Y
				PCT for PRI1	N
Step 3	Conditional Purpose	CPT1 → CPT3 → CPT8 → CPT10 →		Conditional Purpose & Waiver Rule	
Step 4	Action	AT4		Disclose	
Step 5	Special Instruction	IPT1	According to waiver		Disclose
		IPT2	Individual preferences		
		IPT3	Individual restriction		Deny
	Time Fee			N/A	
Step 6	Delivered	RR1	Limited Data to disclose		De-identified Data
		RR2	Data delivery method		Not specified
Released Record		None			
NOTE		Record is ready to release but restricted by individual preferences			

Table 15. Detail Explanation for record R1.1



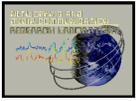
Step 0	Request	Requester	Purpose	Record Items	Process Request	
Step 1	Query	ReqT1	PPT1	PRI1	Process Query	
Step 2	Pre-Conditions	PCT1 → PCT2 → PCT3 → PCT4 →			PCT for ReqT1	Y
					PCT for PPT1	Y
					PCT for PRI1	N
Step 3	Conditional Purpose	CPT1 → CPT3 → CPT8 → CPT10 →			Conditional Purpose & Waiver Rule	
Step 4	Action	AT4			Disclose	
Step 5	Special Instruction	IPT1	According to waiver		Disclose	
		IPT2	Individual preferences		Disclose	
		IPT3	Individual restriction			
	Time				N/A	
	Fee				N/A	
Step6	Delivered	RR1	Limited Data to disclose		De-identified Data	
		RR2	Data delivery method		Not specified	
Released Record		3,PRI1, 04/10/2011				
NOTE		Anonymous Record released.				

Table 16. Detail Explanation for record R1.3

4.3. Researcher 2nd Example

Researcher request in natural language “A researcher wants to disclose Psychotherapy notes for patients who were registered in year > 2010. The researcher has authorization to access protected health information for research purposes. He/She requested to receive the protected health information by email”.

Requester Role:	Researcher	ID:	3456
Authorized By:	Yes - Research	Waiver Rule:	No
Purpose 1:	Research	Single Record Items:	Psychotherapy
Purpose 2:	None	Multi Record Items:	None



Multi Purpose:	None		
Date From:	1/1/2011	Date To:	12/12/2011
Record Format:	Default		

Table 17. Researcher Query

Query in SQL Format

Select Patient_Data(PHI) from “Table 12 (Patients Records)” where PRI= “Psychotherapy” and Date between “2010” and “2011” and Requester= “Researcher” and Purpose = “Research” groupby Requester && Purpose having Waiverrule=“None” && Authorization = “Patients_Autorization”

In this example, Rule number 1 from Figure 4 will be triggered because the researcher has an authorization. In addition, we will get a table similar to Table 14 regarding list of instruction and deliverables for releasing information that will be released by email based on RR2. Table 18 is the actual output. Note that personal information is presented in this table because the researcher has authorization from patients to disclose protected health information.

ID	Name	PRI	Date
5	Raja	PRI5	13/10/2011
7	Nelo	PRI5	15/08/2011

Table 18. Output result of the query

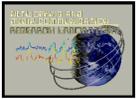
5. COMPARISON OF FORMALIZATION APPROACHES

In this section, we will use the same example explained in section 4 for a researcher requesting the use of PHI and doesn't have authorization. Based on our knowledge at the time of writing this our approach, we found three recent papers that attempted to provide a solution to formalize HIPAA privacy rules. We will show how each work handle the example that we explained earlier to disclose PHI.

The main concept behind formalizing privacy rules of HIPAA Act in [8] is the process of combining related clauses together in which different parts of legal text expressed by combining associated clauses in the form of permitted_by and forbidden_by which are called rules. By referring to section 4.1 "request example", a researcher without authorization “164.508.b.3.i” triggers a forbidden_by rule which has precedence over permitted_by rule generated by the waiver. In this case, if the covered entity provided a waiver “164.512.i.1.i” for researcher to disclose PHI for research purpose, researcher would be able to do that. This is due to the conflict between the forbidden_by and permitted_by to resolve overlapping problem between clauses. In such an approach, formalization process becomes quite unreliable in precision and error prone. On the other hand, implementation of this approach requires less analysis and faster deployment compared to our approach.

In [1] and [10], legal text is translated to prolog using several steps to produce production rules that are based on Hohfeldian Concepts. These concepts consist of 8 different categories and they are; right, obligation, privilege, no-right, power, liability, immunity and disability. The author implemented direct formalization of HIPAA legal text (clause by clause). HIPAA clauses contain references to internal and external clauses which have not been considered in this approach. This would create overlapping of rules. By using the same example discussed in section 4.1, a researcher will not be able to disclose PHI because there is no proposed mechanism in this study to resolve this overlap. However, using this approach to formalize legal text without external and internal references is reliable and similar to the previous study, requires less time for development.

In [16], authors proposed Least Fixed Point (LFP) Logic for assigning particular semantic modal and signature which specifies the privacy regulations. Privacy LFP is used to formalize legal text and information which is processed by predicate send (p1, p2, m), maysend(p1,p2, m). P1 principal sends message m to p2. So the result will be true when p1 sends m to p2 in the send predicate. On the other hand, predicate maysend(p1,p2, m) shows



that the transmission of message m is according to the law from p_1 to p_2 . Positive norm is explained as a transmission that might occur if at least one condition of positive norm is satisfied, for example, HIPAA clause 164.506.c.2 is considered as a positive norm because it states that if the purpose is for treatment then protected health information will be disclosed. So the permitting clause or rule is treated as a positive norm. Negative norms are the rules that are defined as information that will be released only if it satisfies all negative norms. By referring to the request example discussed in 4.1, a researcher without authorization is a negative norm and no PHI will be disclosed.

Other approaches have an advantage when it comes to the timeline required to formalize legal text due to the clause by clause approach. Also, with a non complex legal, they outperform our approach which requires further analysis and more detailed output. Regardless of precision issue with these studies, none of them considered patients and covered entities preferences which are equally important legal requirements as the decision. How information will be release and what will be released is also another important legal requirement. We have summarized the main differences between this study and others in Table-19 as shown below.

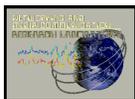
HIPAA- Comparison of our approach with other Approaches					
<i>A: World Rule Modal , B: 1st Approach [10] , C: 2nd Approach [1] ,D: 3rd Approach [16]</i>					
	Considered items in the study	A	B	C	D
1	Clauses are linked within the same section	Y	Y	Y	Y
2	Direct cross references is considered	Y	Y	Y	Y
3	Indicates applied rule as a result of a query	Y	Y	Y	Y
4	Provide type of action taken as a result of a query	Y	N	Y	Y
5	What information will be released is considered	Y	N	Y	Y
6	Study cover all the privacy sections of HIPAA	Y	N	N	Y
7	How information will be released is considered	Y	N	N	N
8	World model diagram for HIPAA	Y	N	N	N
9	Unreferenced Information between sections is covered	Y	N	N	N

Table-19. Comparison of All Approaches

6. CONCLUSION

Formalizing legal text is a complex process that consumes time and efforts but validating the outcome could consume more time and efforts [13]. We proposed a new methodology to formalize legal text which is required to create information system to facilitate the process of exchanging data electronically with lowest human intervention. Precision is important in this matter and this is what differentiates our work from others. The methodology that we proposed requires analyzing legal text to understand the flow of how information is processed. Then, splitting complex information into small manageable pieces (tags) to ease the integration process based on information of requesters, information owner, holder and the law that govern the exchange of the information. Entity relation diagram that explain the course of information is essential to understand how different information are connected together.

Creating a high level diagram is also required to understand the entire process of splitting complex information and generating the output. Pre-processing of the legal text would generate a searchable table that assists in selecting a more precise decision in disclosing or denying release of information. We found that missing important factors that might produce less precise decisions is caused by direct formalization of legal text without

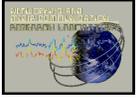


understanding the big picture. Following our model in formalizing legal text will prevent such approach and assure producing a more precise discussion. This methodology can be used with any legal text and HIPAA privacy rules is just an example. We have showed how input and output is analyzed and getting the right decision is only half the way to perfection. Dealing with patients and medical providers' preferences is one important subject but considering how information is released and what will be released is not less important than that.

We validated our solution by comparing it to three studies in this area. We have seen a less precise decision that led to deny disclosing protected health information where in fact this information should be disclosed.

7. REFERENCES

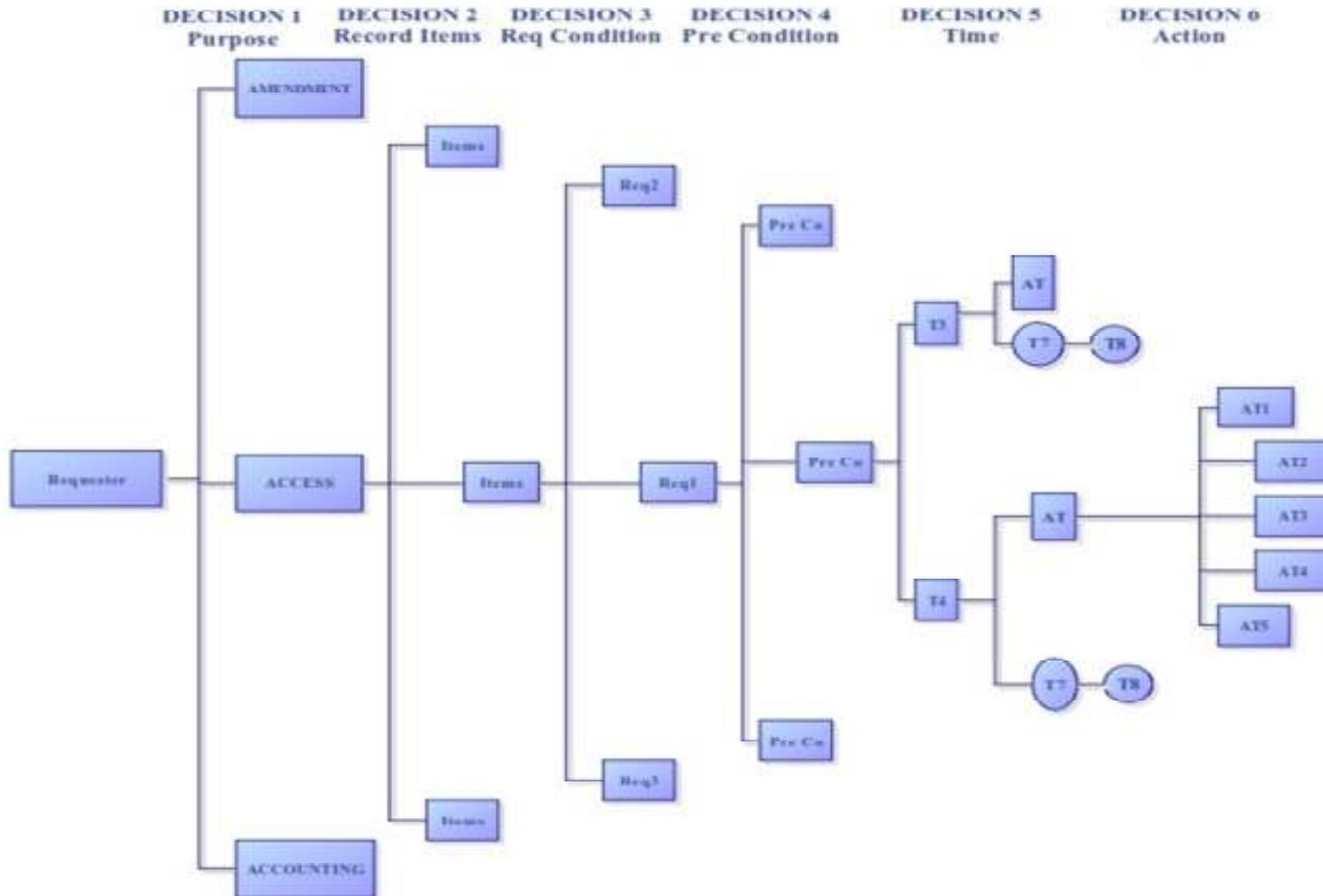
- [1] Maxwell, J.C., Annie I. Anton, "Developing Production Rule Models to Aid in Acquiring Requirements from Legal Texts", Proc. of the 17th Intl. IEEE Requirements Engineering Conf., Atlanta, 2009, pp. 101-110
- [2] Maxwell, J.C., Annie I. Anton, "Validating Existing Requirements for Compliance with Law Using a Production Rule Model" Proc. of the 2nd Intl. IEEE Workshop on Requirements Engineering and the Law, Atlanta, 2009, pp. 1-6.
- [3] T.D. Breaux, "Legal Requirements Acquisition for the Specification of Legally Compliant Information Systems", Ph.D. Dissertation, North Carolina State University, 2009.
- [4] As Office for Civil Rights (2003) Summary of the HIPAA privacy rules.<http://www.hhs.gov/ocr/privacy/hipaa/understanding/summary/privacysummary.pdf>
- [5] <https://www.privacyrights.org/fs/fs8a-hipaa.htm>
- [6] <http://www.hhs.gov/ocr/privacy/hipaa/understanding/summary/index.html>
- [7] <http://www.hipaa.com/2009/02/president-obama-to-sign-arras-hitech-provisions-tuesday-february-17-2009-in-denver-co/>
- [8] "HIPAA Administrative Simplification 45 CFR Parts 160, 162, and 164", U.S. Department of Health and Human Services, Office for Civil Rights, 2006 <http://www.hhs.gov/ocr/privacy/hipaa/administrative/privacyrule/adminsimplpregtext.pdf>
- [9] Roberta B. Ness. A year is a terrible thing to waste: early experience with HIPAA. *Annals of Epidemiology*, 15(2):85-86, 2005.
- [10] Peifung E. Lam, John C. Mitchell & Sharada Sundaram, "A Formalization of HIPAA for a Medical Messaging System". Stanford University, Stanford, CA. Lecture Notes in Computer Science, 2009, Volume 5695/2009, 73-85, <http://www.springerlink.com/content/e6281457716k0128/>
- [11] J.C. Maxwell, Annie I. Anton, "A Refined Production Rule Model for Aiding in Regulatory Compliance", (in submission) *IEEE Trans. on Software Engineering*, 2010 North Carolina State University Technical Report, TR-2010-3, 2010.
- [12] T.D. Breaux, Annie.I. Antón, "Analyzing Regulatory Rules for Privacy and Security Requirements", *IEEE Trans. on Software Engineering*, 34(1), Jan.-Feb. 2008, pp. 5-20.
- [13] P.N. Otto, Annie I. Antón, "Addressing Legal Requirements in Requirements Engineering", Proc. of the 15th IEEE International Requirements Engineering Conference, New Delhi, 2007, pp. 5-14.
- [14] R.J. Brachman, and Levesque, H.J., *Knowledge Representation and Reasoning*, Elsevier, 2004.
- [15] H. DeYoung, D. Garg, D. Kaynar, and A. Datta. Logical specification of the GLBA and HIPAA privacy laws. Technical Report CMU-CyLab-10-007, Carnegie Mellon University, 2010.
- [16] H.DeYoung, D. Garg, L. Jia, D. K. Kaynar, and A. Datta. Experiences in the logical specification of the HIPAA and GLBA privacy laws. In *WPES*, pages 73–82, 2010.
- [17] [17] Wilson J (2006). "Health Insurance Portability and Accountability Act Privacy rule causes ongoing concerns among clinicians and researchers". *Ann Intern Med* 145 (4): 313–6. PMID 16908928.
- [18] [18] D. M. Sherman. A prolog model of the income tax act of Canada. In *ICAIL '87: Proceedings of the 1st international conference on Artificial intelligence and law*, pages 127{136, 1987.

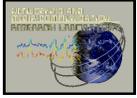


- [19] [19] Marc A. Borrelli. Prolog and the law: using expert systems to perform legal analysis in the United Kingdom. *Softw. Law J.*, 3(4):687, 715, 1990.
- [20] Helen Nissenbaum. Privacy as contextual integrity. *Washington Law Review*, 79(1):119{158, 2004.
- [21] Kimble, Joseph, "The Great Myth That Plain Language Is Not Precise," *BUS L TODAY*, July-August 2000, at 48. Reprinted in *7 SCRIBES J LEGAL WRITING* 109 (1998-2000).



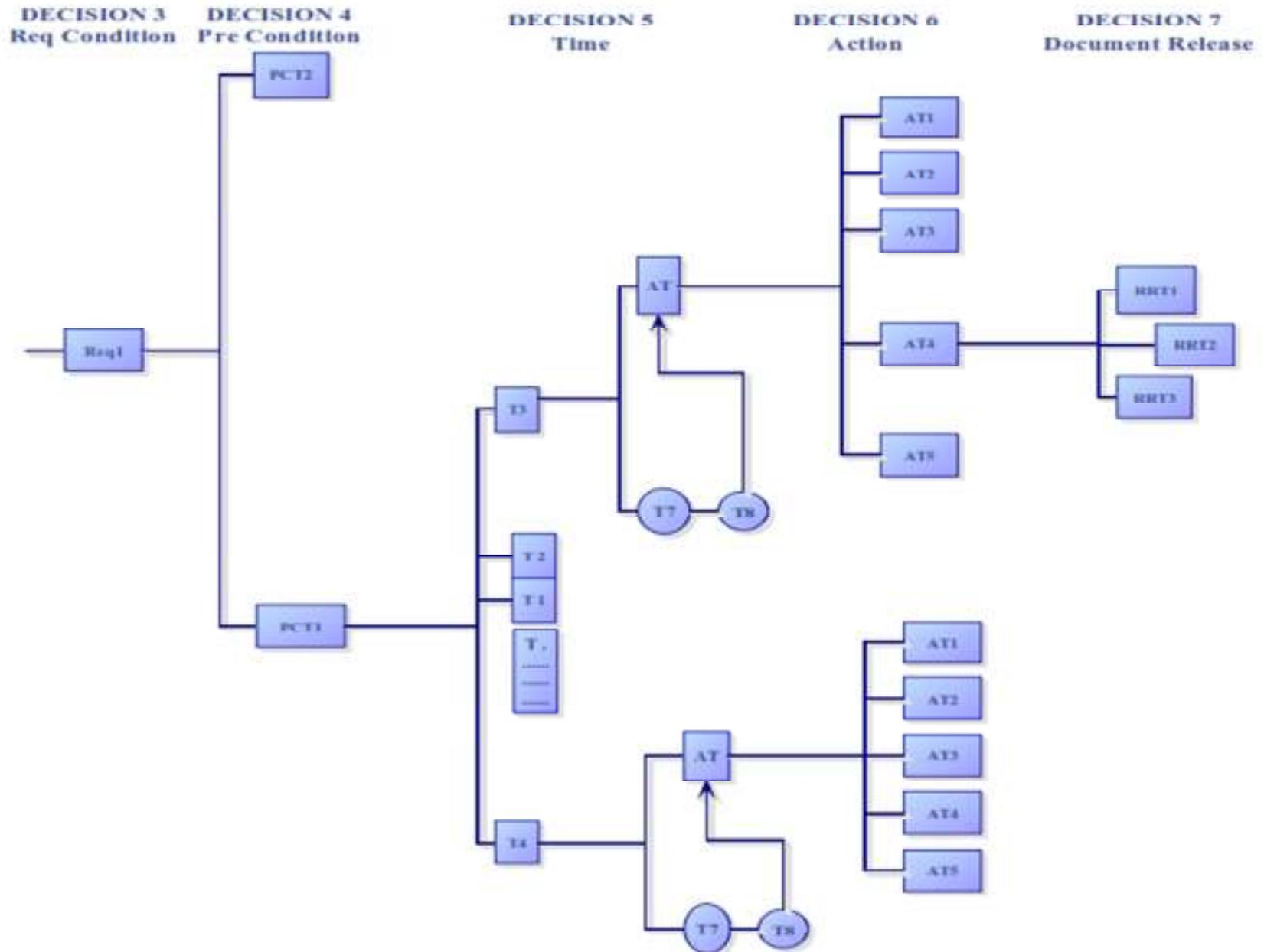
Appendix A: A Complete Decision Tree for Rules Generation

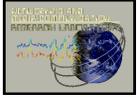




Technical Report 2012-06-02
Internetworking and Media Communications Research Laboratories
Department of Computer Science, Kent State University
<http://medianet.kent.edu/technicalreports.html>

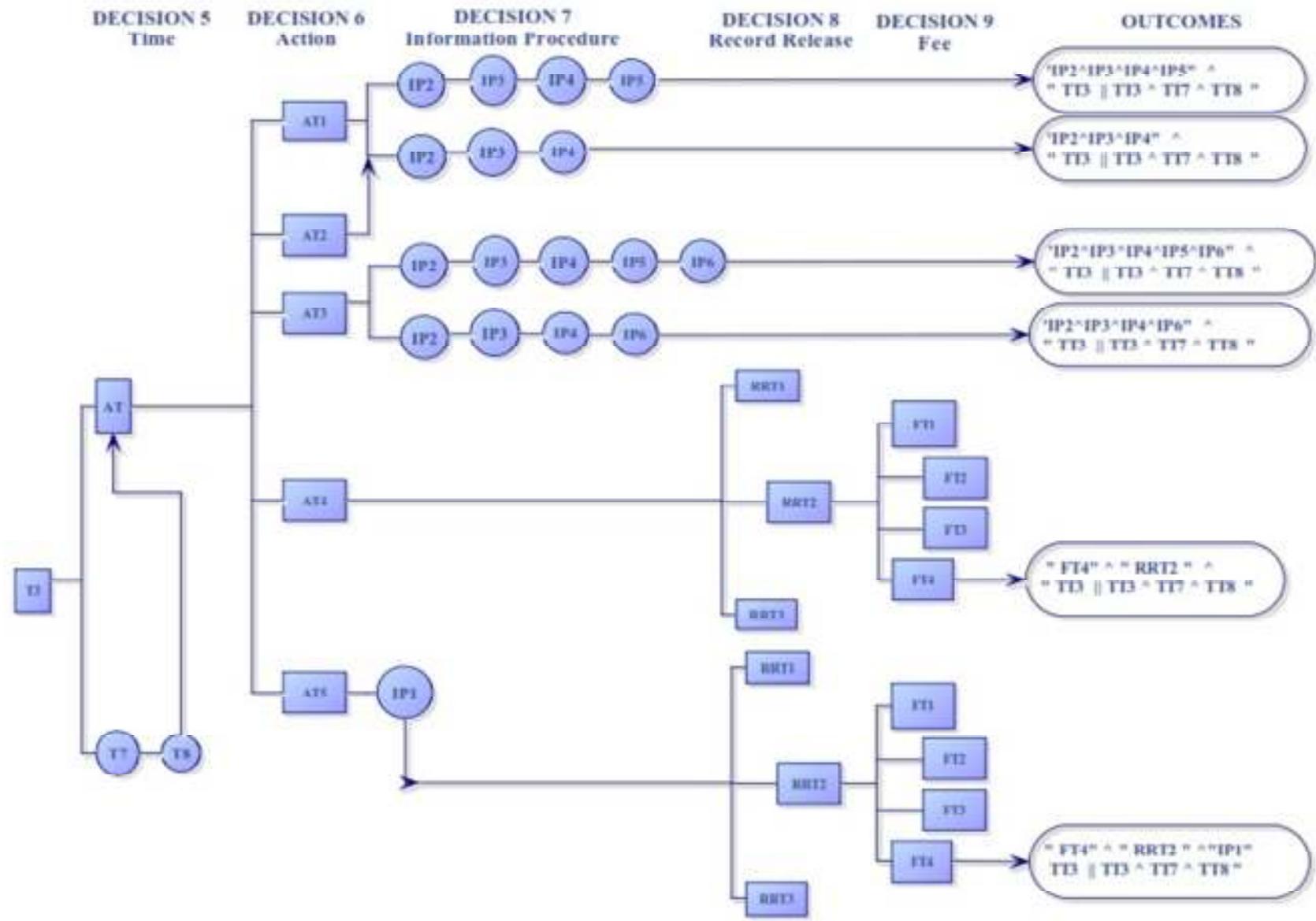
Figure A: Decision Tree for Rule Generation

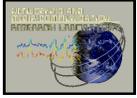




Technical Report 2012-06-02
Internetworking and Media Communications Research Laboratories
Department of Computer Science, Kent State University
<http://medianet.kent.edu/technicalreports.html>

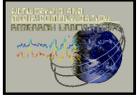
Figure B: Decision Tree for Rule Generation





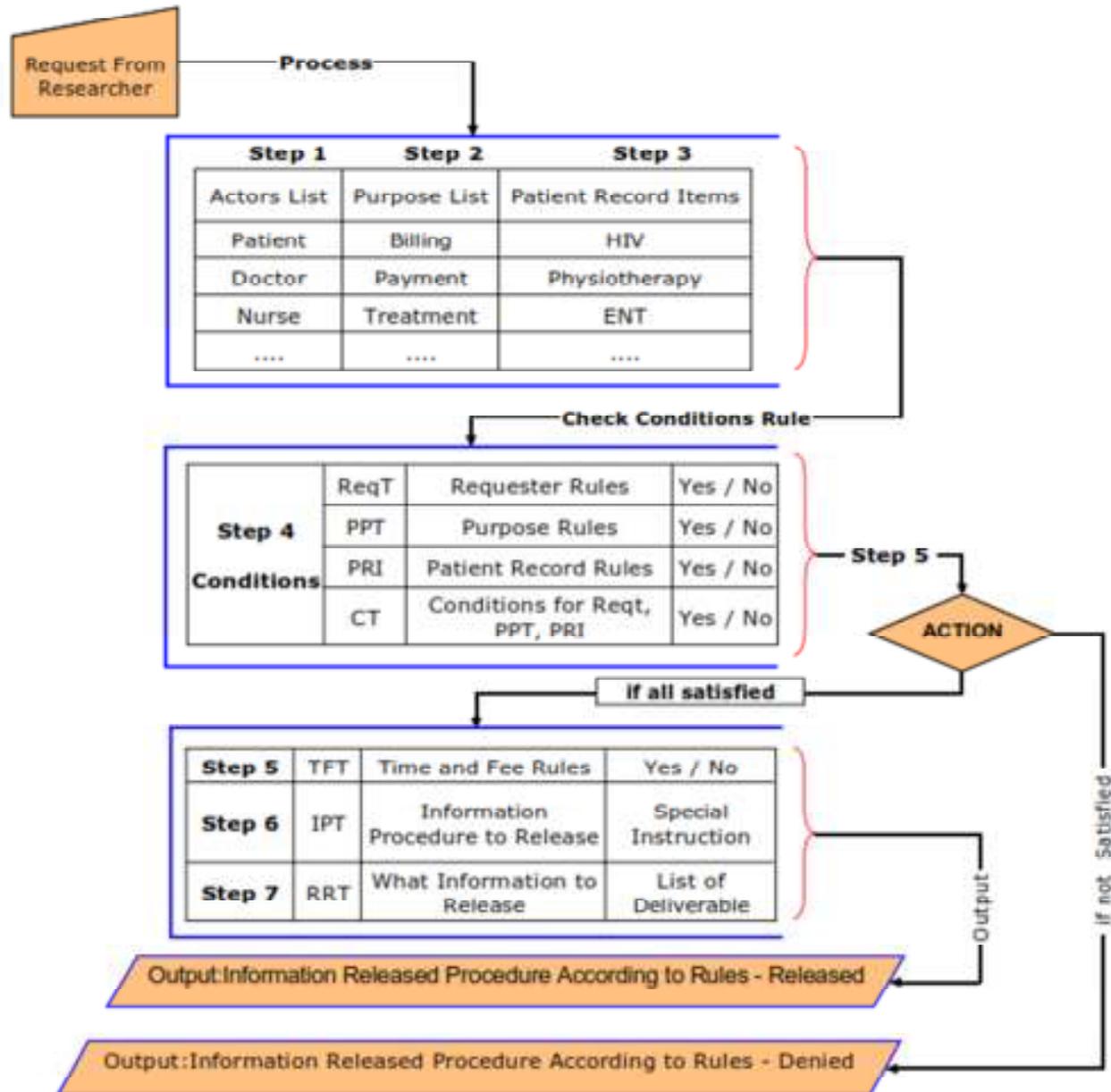
Technical Report 2012-06-02
Internet Networking and Media Communications Research Laboratories
Department of Computer Science, Kent State University
<http://medianet.kent.edu/technicalreports.html>

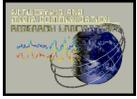
Figure C: Decision Tree for Rule Generation



Technical Report 2012-06-02
Internetworking and Media Communications Research Laboratories
Department of Computer Science, Kent State University
<http://medianet.kent.edu/technicalreports.html>

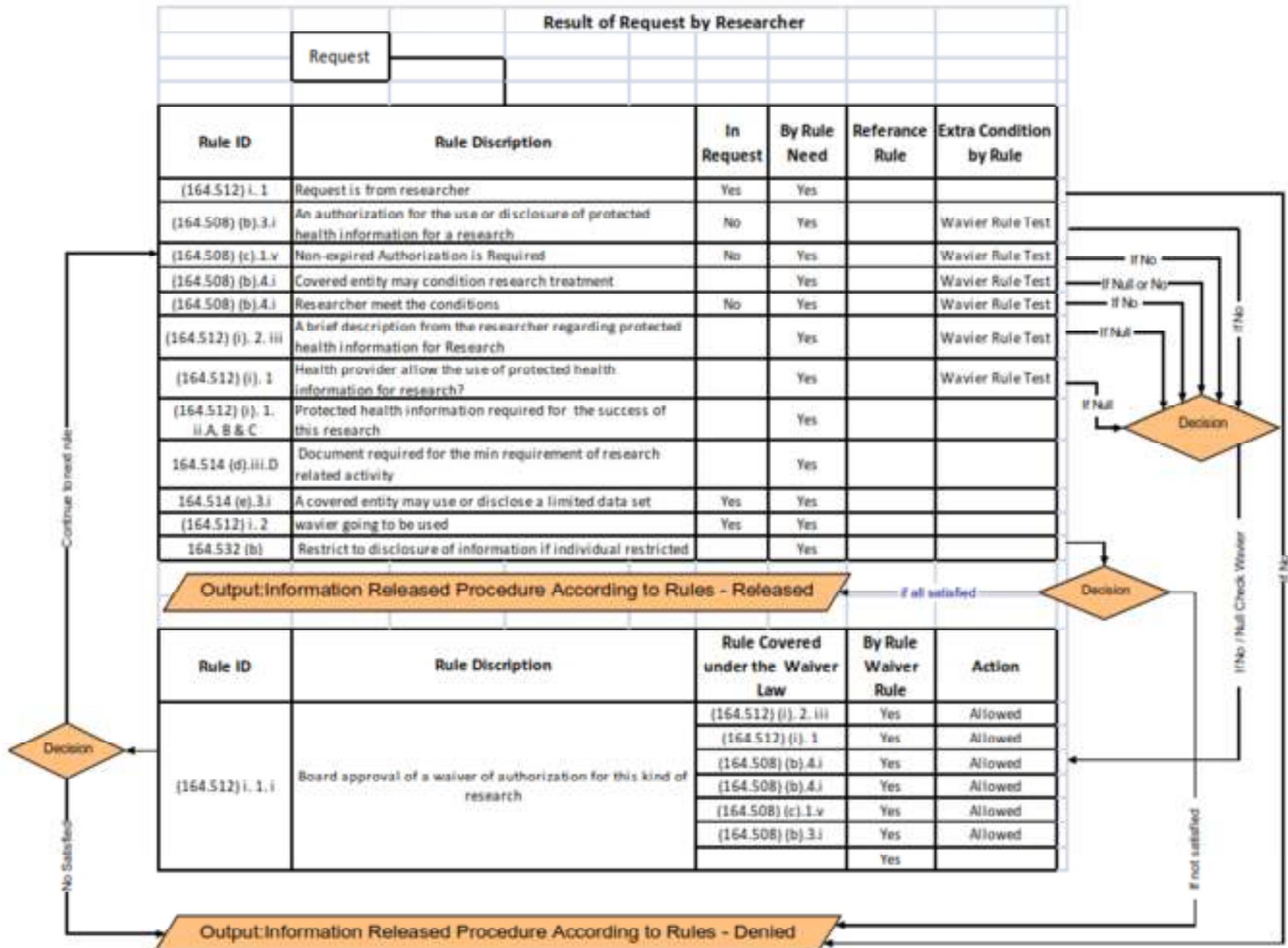
Appendix B: Request Flow Procedure

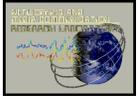




Technical Report 2012-06-02
Internetworking and Media Communications Research Laboratories
Department of Computer Science, Kent State University
<http://medianet.kent.edu/technicalreports.html>

Example A: Researcher Request Flow Examples after Rules Generation from HIPAA.





Technical Report 2012-06-02
Internetworking and Media Communications Research Laboratories
Department of Computer Science, Kent State University
<http://medianet.kent.edu/technicalreports.html>

Example B: Insurance OR Plan Sponsor Related Rules in HIPAA.

