A Domain Analysis Model for eIRB Systems: Addressing the Weak Link in Clinical Research Informatics

Shan He[a], Scott P. Narus,[a,b] Julio C. Facelli[a], Lee Min Lau[a,c], Jefferey R. Botkin[a], John F. Hurdle[a]

[a]Department of Biomedical Informatics, University of Utah, Salt Lake City, UT, USA

[b]Intermountain Medical Center, Intermountain Healthcare, Murray, UT, USA

[c]3M Health Information Systems, Murray, UT, USA

Corresponding Author

Name: Shan He

Email: shan.he@utah.edu

Address: 26 S 2000 E HSEB RM 5700, Salt Lake City, UT 84112, United States

Phone: 1-801-477-4260

Fax: 1-801- 581-4297

Abstract

Institutional Review Boards (IRBs) are a critical component of clinical research and can become a significant bottleneck due to the dramatic increase, in both volume and complexity of clinical research. Despite the interest in developing clinical research informatics (CRI) systems and supporting data standards to increase clinical research efficiency and interoperability, informatics research in the IRB domain has not attracted much attention in the scientific community. The lack of standardized and structured application forms across different IRBs causes inefficient and inconsistent proposal reviews and cumbersome workflows. These issues are even more prominent in multi-institutional clinical research that is rapidly becoming the norm. This paper proposes and evaluates a domain analysis model for electronic IRB (eIRB) systems, paving the way for streamlined clinical research workflow via integration with other CRI systems and improved IRB application throughput via computer-assisted decision support.

1. Introduction

Remarkable growth has occurred in biomedical research in recent years [1]. The sheer amount of research now being conducted has resulted in an unprecedented increase in workload and placed a severe burden on Institutional Review Boards (IRBs[1]) [2]. In addition, in recent years there has been a growing trend towards multisite clinical research, complicating the effectiveness of an institutionally-based oversight system [3,4]. Moreover, biomedical research enabled by new technologies in healthcare such as retrospective cohort research involving secondary use of patient data from electronic health record (EHR) systems and genetic research involving specimens from biobanks have raised new ethical questions and concerns [5,6] that require additional review resources.

IRBs are charged with reviewing of all research projects that involve humans to ensure that they comply with local, state, and federal laws, as well as with the ethical standards set forth by local policy. An IRB serves its own research community by applying high standards of intellectual integrity and careful attention to federal research regulations. They strive to provide investigators and study teams the support and resources they need to conduct high quality research and foster research practices that protect participants. The nature this mandate is complex, we provide several examples throughout this paper of specific IRB activities.

As the volume of research grows, the number of IRBs also increased substantially. Catania et al. showed that the number of U.S. IRBs increased by 41% from 2004 to 2008 [2]. According to the March 2013 data from the Office for Human Research Protections (OHRP), there are 2,937 actively registered U.S. IRB organizations (IORGs) and 1,983 non-U.S. IORGs, which include 3,589 individual U.S. IRBs and 2,252 non-U.S. IRBs, respectively [7].

---

[1] "IRB" is a generic term used by U.S. Department of Health and Human Services to refer to a group whose function is to review research to assure the protection of the rights and welfare of the human subjects. Each institution may use different names such as Research Ethics Committee, Committee on Human Studies, the Committee on Clinical Investigations, etc. For the sake of simplicity, this paper will use the term IRB to encompass all of these names.

There is a substantial literature discussing the problems faced by IRBs, including redundancy of duplicative review for multisite studies, inconsistency in decision-making across different IRBs, and inefficiency in communication among different stakeholders [8–11]. Studies show that investigators complain that the IRB application process is burdensome and, in some instances, waiting to obtain IRB approval has delayed project initiation substantially [8,12]. Despite the increasing interest by the informatics community in developing clinical research informatics (CRI) systems [13,14] and its long-standing interest in data standards for system interoperability and data sharing [15], informatics research in the IRB domain has not attracted much attention.

Consequentially the IRB review process is largely manual, using electronic systems primarily for communication. These systems are often built on simple databases that facilitate the transfer of *non-computabl*e electronic documents. This point was confirmed by our analysis of IRB application systems at all Clinical and Translational Science Award (CTSA) centers [16]. We found that 72% of CTSA institutions used some form of online IRB application systems during the study year (2012). However, the capability of these online systems varies greatly across organizations. Some systems simply allow investigators to upload application-related documents in Word or PDF format, while others have a "smart form" feature that can dynamically guide investigators through relevant online forms. Even for those systems that support forms entry, the preponderance of fields are free-text. The unstructured information in those fields is difficult to process for automated analysis or for data sharing between CRI applications.

In this paper, we propose and evaluate a domain analysis model to standardize the information elements within the IRB oversight domain. There have been many efforts to model various aspects of biomedical research in general. For example, the Protocol Representation Model (PRM) from the Clinical Data Interchange Standards Consortium (CDISC) [17] focuses on the characteristics of a study and the definition of activities within clinical trial protocols. The

Biomedical Research Integrated Domain Group (BRIDG) project [18] developed a comprehensive domain analysis model for protocol driven research and its associated regulatory artifacts. The regulatory subdomain of BRIDG was designed for the Food and Drug Administration's (FDA) regulated product submission process instead of IRB oversight. As a result it is too coarse-grained for computers to process by modeling the IRB submissions at the document level. The Ontology of Clinical Research (OCRe) is a formal ontology for annotating existing human studies and supporting federated query on data and meta-data across studies from different sources [19]. It focuses on the design and analysis phase of studies. To the best of our knowledge, however, there is no existing research on modeling the IRB oversight domain.

The goals of the work presented here are: 1) to develop a platform-independent domain analysis model that captures the structured data elements and high-level business processes for the IRB oversight domain; and 2) to evaluate the model's ability to represent the informational elements found in five different types of real-world IRB systems. The paper presents a detailed discussion of the domain analysis process, the resulting IRB model, and the results of the evaluation.

2. Model Development

Our design strategy was to capture all essential information that IRBs require to provide human subjects research oversight, as well as capture information that could be meaningfully shared with other CRI systems. Such information is represented in a structured way when possible, serving as the foundation for future decision support based on predefined rules or through binding to an ontology. For example, if a planned study activities involve high-risk procedures such as ionizing radiation or informational risk such as secondary use of existing data or specimens, commensurate review procedures for specific risks could be suggested automatically. We try to avoid modeling verbatim concepts defined in regulations unless they have an extension definition (e.g., "vulnerable population" is defined by federal regulations [20] and OHRP

guidelines [21] by naming a list of vulnerable subject categories). We do not model entities that are primarily for human understanding (as opposed to computer interpretation), such as the standard language used in an informed consent. It may be possible in the future to facilitate automatic informed consent generation as a more advanced application of our model, however. Our intention is not to replace human review with a computer system, but to make the reviewer's job easier by utilizing information collected in a machine-understandable way and presenting the information to the reviewers in a consistent manner. In this way, we are trying to maximize the advantages of both human (IRB reviewers) and automated processes, while adding the capability to integrate IRB systems into other CRI applications.

The IRB domain analysis model (DAM) is an implementation-independent model built on a high level of abstraction of the IRB domain. A host of documented domain analysis methods are available from the software engineering field, but no one standard domain analysis process is considered "best"[22]. We adapted the Feature-Oriented Domain Analysis (FODA) method [23] and the conceptual modeling method presented by Embley et al. [24] for our work. Figure 1 illustrates the IRB domain analysis process and the corresponding model artifacts generated by each step (i.e., context analysis, domain modeling, and business process modeling), which we describe in Sections 2.1-2.3.

2.1 Context Analysis

The first phase, context analysis, defines the scope of the modeling domain. The relationships between the candidate domain and its parent domain, subdomains, and peer domains are shown in Figure 2. The context analysis was based on careful review of federal regulations that define the IRB oversight scope and responsibility [25]. We found significant overlap within the IRB and the study protocol domain. The study protocol is the blueprint of every research project. It gives the IRB a comprehensive view of the study, including subject eligibility criteria and

recruitment, planned study procedures and interactions, data management, and analysis plans. The study protocol is essential in order for IRBs to evaluate the potential benefits and risks of a study. However, some details in the study protocol that may be important for study management purpose are not relevant for IRB review (e.g., the coding systems used for recording study condition and adverse events, the technique used for reporting study subject accrual data to the study sponsor, study acronyms, etc.). There are also certain aspects of a study that might not be covered in the study protocol but that are important for IRBs to know for review such as the informed consent process, compensation to subjects, vulnerable population participants, status report, unanticipated problem report, etc. Context analysis is the basis for planning the domain modeling phase during which all relevant entities are identified.
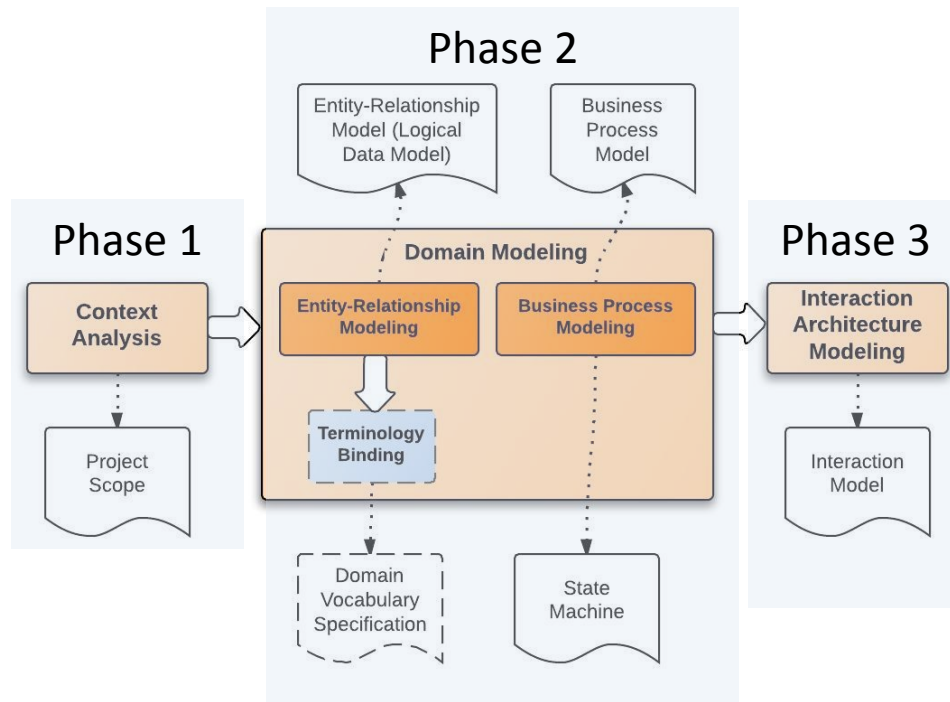

Figure 1. **Overview of the IRB Domain Analysis Process**. The IRB domain analysis process consists of three phases: context analysis, domain modeling and interaction architecture modeling. The white document-shaped boxes are the deliverables associated with each phase. The dashed boxes indicate that only preliminary work has been done.
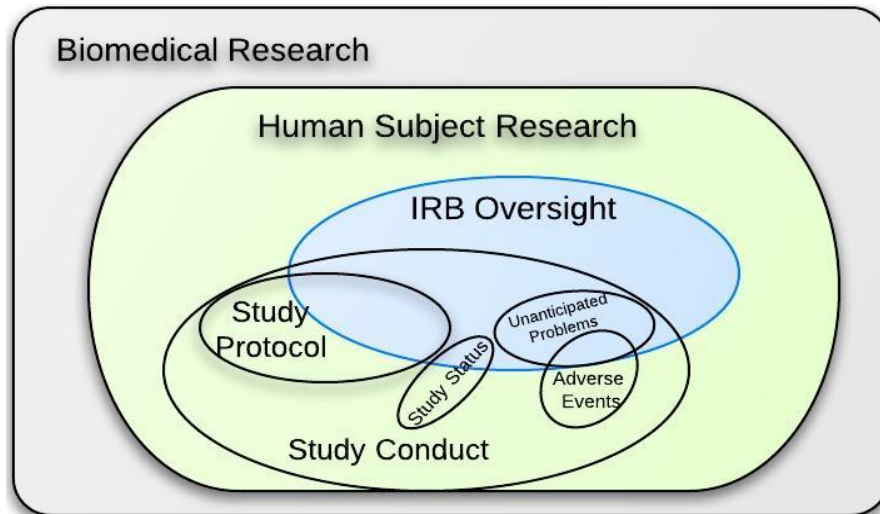
Figure 2. **Context Analysis for the IRB Oversight Domain**. This defines the scope of the modeling domain. Analysis of the relationship between the candidate domain and its parent domain, subdomains, and peer domains is useful to identify which data elements can be consolidated from existing standards and which should be defined.

2.2 Domain Modeling

The second phase, domain modeling, consists of entity-relationship modeling, which details the static (structural) semantics of the domain, as well as business process modeling, which illustrates the dynamic (behavioral) semantics of the domain.

2.2.1 Domain Modeling Process

The static artifact produced from entity-relationship modeling is called the entity-relationship model, which is independent of the underlying database design. Some literature refers to this as an information model or logical data model [26]. In this paper, these terms will be used interchangeably. Terminology binding connects the entity-relationship model with domain vocabulary specifications to standardize the terminology that describes the domain. In this paper, only a preliminary terminology binding effort was conducted – a comprehensive domain vocabulary development is considered as future work. The dynamic artifacts produced in the business process model were based on expert interview and literature review in this domain.

These include the high-level, generic IRB application and review processes, augmented by a state machine that describes the status transition of an IRB application. This paper only discusses the development process for the IRB entity-relationship model (Figure 3) in detail since it is the basis for supporting any internal or external automated tasks and since it comprises the most important artifact in the IRB DAM.
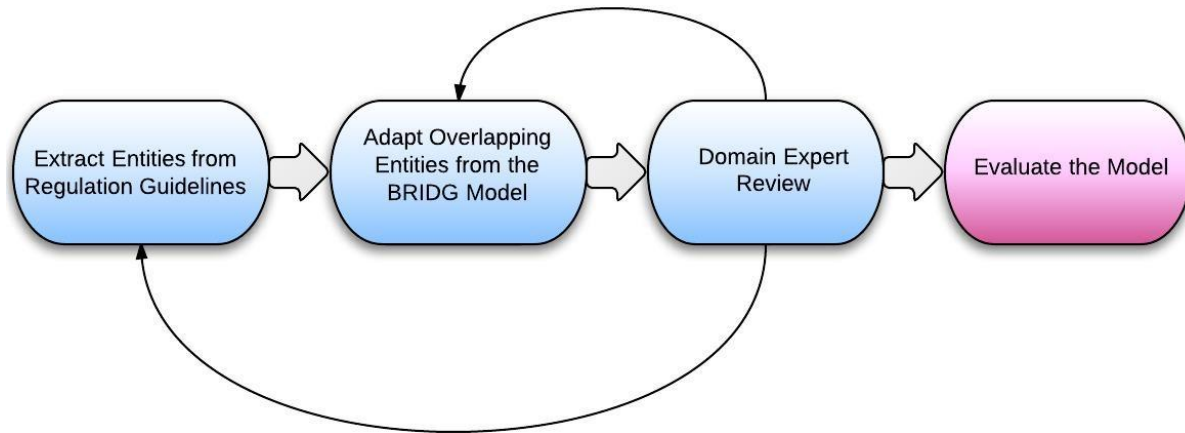

Figure 3 Overview of the Entity-relationship Modeling Process. This is an iterative process. Feedback from domain experts led us to revisit the knowledge sources and revise the draft model.

First, all the key entities in the IRB oversight domain were extracted from regulation-based guidelines, primarily the AAHRPP (Association for the Accreditation of Human Research Protection Programs, Inc.) accreditation standards, with OHRP, FDA regulatory guidance documents and HIPAA educational materials as complementary knowledge sources. AAHRPP is an independent, nonprofit organization that accredits high-quality human research protection programs. It has published a series of accreditation standards that are compliant with U.S. federal regulations and international ethical principles. Specifically, the *Evaluation Instrument for Accreditation* (Version January, 2012)[27], and *IRB Evaluation Checklist* (Version December, 2010) [28] were used as guidelines in the IRB domain modeling process. Each entity in the model is described by a set of attributes, each of which is bound to a data type defined in the

HL7 Version 3 Data Type Abstract Specification (Release 2) [29]. There are too many entities to list here, but the complete model is available online as described below. In the online version of the model, entities from different knowledge sources are distinguished by different colors as explained by the diagram legend.

Second, relevant entities and their attributes from the BRIDG model were used wherever possible, especially when the regulation guidelines did not provide sufficient detail. For example, AAHRPP does not define what a "full protocol" includes. In such cases, the study protocol-related classes from BRIDG were used. However, BRIDG focuses primarily on modeling clinical trials (either interventional or observational) and it is missing the category of retrospective studies, which use existing health data for secondary analysis. We reorganized and modified some entities and attributes in the BRIDG model to accommodate retrospective studies or social/behavioral science studies.

Third, a draft version of the IRB entity-relationship model developed from the first two steps was reviewed by two domain experts (a former IRB Chair and the Associate Director of the University of Utah IRB) to determine model accuracy. Not all entities or attributes about study protocol in the BRIDG model proved interesting to the IRB. Such information elements were eliminated from the IRB model during the expert review. To make the model more understandable to these domain experts, a concept map was derived from the IRB class diagram as a review aid. The concept map includes only the entities in the domain, and their relationships, without further details about as attributes or data types. It shows the big picture of the IRB domain and is visually simpler to comprehend than the UML class diagram. Both domain experts independently marked each class and associated attributes as "yes" meaning "to include" or "no" meaning "not to include." After the independent reviews from each expert were completed, the results were consolidated and the conflicts were identified and resolved in a joint session.

The final step was to evaluate the comprehensiveness of the model by comparison with real-world IRB application systems. We mapped the online or Word/PDF forms from five large IRBs to the model. The evaluation results comprised of complicated mapping categories shows overall agreement between a real system and our model but also provided insights into future improvements to both model development and IRB application form design. We incorporated entities that were identified as missing from the evaluation process to the model. The details of this model evaluation will be discussed in Section 3.

2.2.2 Domain Modeling Results

2.2.2.1 The IRB Entity-relationship Model

The IRB DAM is represented in Unified Modeling Language (UML) and was developed using Enterprise Architect (Version 9.2). The IRB entity-relationship model comprises 97 entities and 132 relationships in total. It covers eight core areas of the IRB domain, which are illustrated over the concept map for display purposes only (Figure 4). The complete model in UML class diagram can be accessed online at http://irb-dam.bmi.utah.edu.

The Study Protocol core represents the informational entities that pertain to the plan of a human subject research project. A classification of study protocols is needed because different study types need different information elements to describe the study. For example, a retrospective study uses only existing health data and does not require recruiting or interventional procedure plans in the study protocol. In contrast, a prospective study protocol should describe in detail the recruitment process and all observational or interventional procedures that will be applied to study participants. A well-designed study protocol typology, with relevant information elements defined for each study type, can facilitate "smart" form design in eIRB systems so that investigators do not need to answer non-applicable questions. We developed a study protocol classification schema specifically for the IRB domain shown in Figure 5.
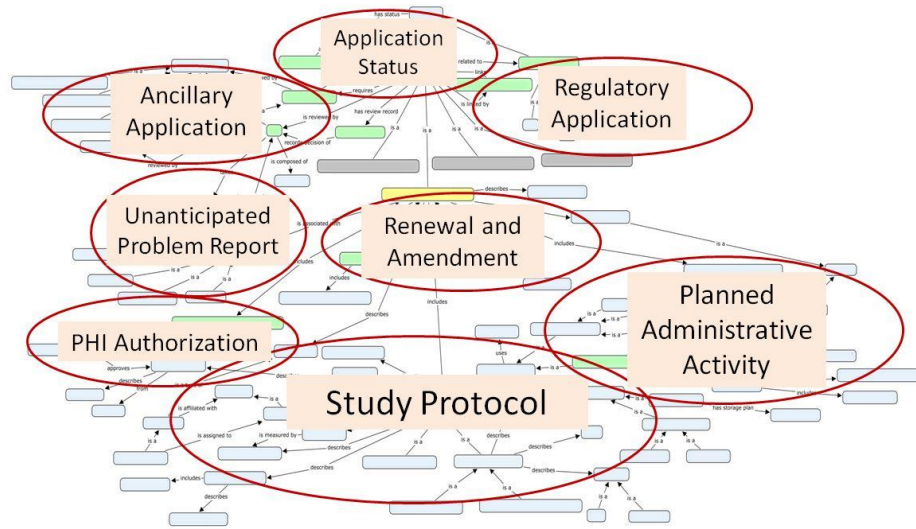
Figure 4. An Overview of the Concept Map for the IRB Oversight Domain. This diagram illustrates the eight core areas of the IRB domain over the concept map for display purposes only. It is not intended to show the individual entities and relationships in detail. To view that detail, navigate to: http://irb-dam.bmi.utah.edu/
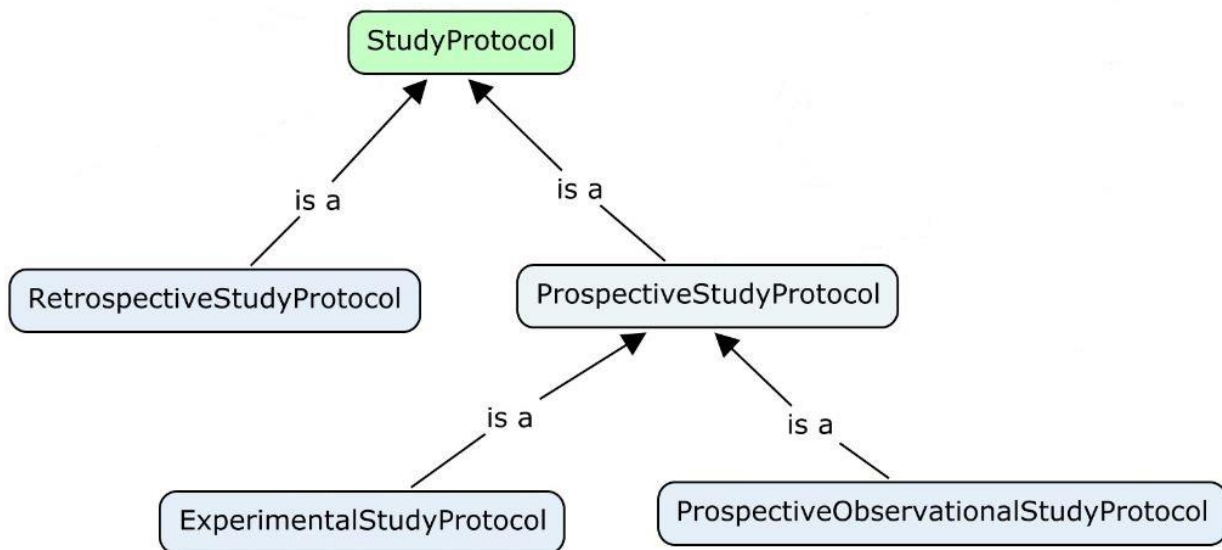

Figure 5. The Study Protocol Typology for the IRB Oversight Domain. This classification is designed especially for IRB review purpose. Each child study protocol type has its own characteristics that differentiate it from its parent. More specific study design types will be captured in the "designConfiguration" attribute defined in the StudyProtocol class.

<u>The Planned Administrative Study Activity core</u> represents those activities that are not directly related to the analysis of study outcomes such as participant recruitment procedures, compensation to study participants, and informed consent processes, etc. Since these administrative activities can raise significant ethical concerns, IRBs always require investigators to specify the details of these activities.

<u>The PHI Authorization core</u> covers the informational entities that are related to the IRB (or Privacy Board) as regulated by the HIPAA Privacy Rule. For example, the request for waiver or alteration of authorization describes the PHI data elements to access, as well as the justification of such access. Documentation of the IRB's review and approval of the request is also required by the HIPAA Privacy Rule. In fact, what kind of PHI access has been approved by IRB is essential to achieve *automated* access control in secondary-use datasets, an important long-term goal of our work.

<u>The Unanticipated Problem Report core</u> defines the informational elements that should be reported to the IRB as well as the elements documenting the corresponding actions taken by the IRB. For this core, we adopted some concepts from the Adverse Event subdomain of the BRIDG model. Since adverse events are often required to be reported to other bodies such as a local Data and Safety Monitoring Committee, the sponsor, or the FDA, a standardized format can facilitate automated reporting and avoid duplicate report preparation efforts.

<u>The Application Amendment and Renewal core</u> defines information entities such as amendment items and status report that are related to the IRB's continuing review. A standardized status report can be automatically generated from existing information stored in a clinical trial management system (CTMS) by defining certain report parameters.

The Application Status core represents information related to the review status of an IRB application and can be shared with other CRI applications to automate a streamlined workflow. This can eliminate the needs for manually delivering paper-based IRB approval letters to different stakeholders in the research domain.

The Ancillary Applications core covers areas such as radiation safety review committees, scientific review committees, conflict of interest committees, data and safety monitoring committees, etc., are also submitted to oversight committees in addition to the IRB, and many IRBs require ancillary application approvals before providing the final IRB approval. These ancillary oversight committees may or may not require extra information besides the standard IRB application. The IRB model presented is not intended to include the details of ancillary applications since many of them are dictated by local policies. However, the model is designed to be able to support any extension by local IRBs.

Finally, the Regulatory applications core covers those application types mandated by regulatory authorities such as the FDA. For studies involving investigational drugs or significant risk devices, regulatory applications such as investigational new drug (IND) or investigational device exemption (IDE) are required to be reviewed and approved by the FDA, in addition to being approved by the IRB. The model is designed to support such requirement by defining constraining relationships between an IRB application and regulatory applications.

2.2.2.2 Terminology binding

One of our major goals in developing the IRB DAM is to achieve interoperability between eIRB systems and other CRI systems. Implementation of the IRB entity-relationship model enables syntactic interoperability by specifying the structure of information being exchanged between different systems. However, to make the meaning of the exchanged content understandable to the receiving system, semantic interoperability should be achieved. Therefore every attribute with

the HL7 data type CD (*ConceptDescriptor*) in the IRB entity-relationship model should be bound to value sets where each value set consists of one or more coded concepts. Such value sets can be defined from scratch or by adopting existing terminologies, code systems, or ontologies if available. For example, many "study characteristic" related value sets defined in OCRe such as "phase," "blinding type," and "sampling method" can be bound to the corresponding attributes in our IRB model. Furthermore, certain consent related attributes can be bound to terms defined in existing research permission or informed consent ontologies [30,31] . We defined the value sets for a few attributes in the model as a preliminary IRB domain vocabulary specification effort and provide examples in Table 1.

We are specifically focused on defining the values sets for attributes concerning data-access requests such as *dataCategory* and *dataElementName* (as shown in Figure 6). The value sets are represented by the Research-Oriented Health Data Representation (ROHDR) model to define standardized representations of health data categories and specific data elements requested byresearchers. ROHDR is adapted from Common Data Model (CDM) Version 4 from the Observational Medical Outcomes Partnership (OMOP) [32] and can be accessed online at http://irb-dam.bmi.utah.edu/rohdr.A formal and comprehensive IRB domain vocabulary specification is necessary and should be considered as future work.

Table 1. Example value sets for concept descriptors described by the Entity-relationship model.

| Example IRB *Application Status* Value Set	Example IRB *Vulnerable Populations* Value Set
Approved	Children
ApprovedWithCondition	CognitivelyImpaired
Closed	EconomicallyDisadvantagedPerson
Created	EducationallyDisadvantagedPerson
Disapproved	ElderlyPersons
Modification Required	EmancipatedMinor
Renewed	Employees
Submitted	Fetuses
Suspended	HumanInVitroFertilization
Under Review	MatureMinor
	NormalVolunteers
	PregnantWomen
	Prisoners
	Students
	TerminallyIllPatients
	TraumaticOrComatosePatients
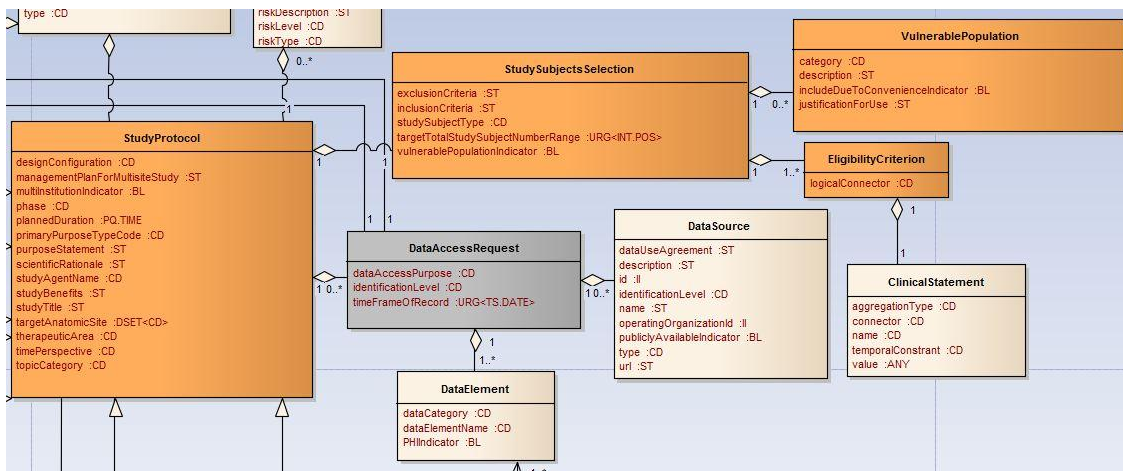
Figure 6. Data-Access Request Related Classes in the IRB Model: the *dataCategory* and *dataElementName* attribute from the *DataElement* class will be bound to the ROHDR model to achieve semantic interoperability.

2.2.2.3 The Business Process Model

As illustrated in Figure 1, the behavioral business process model specifies the high-level features of the eIRB system. It focuses on the end-user's perspective of the functionality of the application. Considering the potential variability in review workflows among different IRBs, the current business process model detailed on the Website below intentionally avoids detailed workflow design such as application review assignment, internal review processes, or meeting scheduling. The system interactions with investigators and IRB reviewers that we show in our business process model are generic enough, in our opinion, to be applicable across institutions. We expect that the model would be refined once others gain real-world experience with it. These models can be accessed online at http://irb-dam.bmi.utah.edu.

2.3 Interaction Architecture Modeling

The final phase of modeling, the interaction architecture model design, specifies interactions and information exchange between an eIRB system and other CRI systems to realize new features that are not supported by isolated systems. The interaction model is represented using the Business Process Model and Notation (BPMN) Collaboration Diagram. BPMN is a standard

Business Process Modeling Language (BPML) that is intended to provide a notation readily understandable by all stakeholders [33]. The BPMN Collaboration Diagram is one of the three sub-models supported by BPMN and is the most suitable for describing the interactions between different systems (participants) using Pools and message exchange between the participants using Message Flows. Example interactions can be found online. For example, the *BPMN 2.0 Models/IRB-Based Access Control for SUHD[2] systems* example illustrates an interaction model for automatically connecting an eIRB to a local external source of clinical data (protected health information).

3. Model Evaluation

The structural IRB entity-relationship model was evaluated to validate its support for representing informational elements found in diverse types of IRB applications at different institutions. The evaluation included comparisons with real-world IRB application systems from five representative institutions. These IRB application systems were chosen because they come from institutions across the nation and each of them is representative of a typical submission method or review setting. Table 2 summarizes the five IRB application systems used in the evaluation.

3.1 Evaluation Methods

The evaluation was performed from January to March 2013. The most up-to-date Web forms or Word/PDF application templates during that period of time at each institution were used for the evaluation. Each field defined in the Web form or Word/PDF templates was extracted as an information item and mapped to the IRB entity-relationship model (referred to as "Model") with one of the eight mapping types as the mapping results described in Table 3.

---

[2] SUHD stands for Secondary Use of Health Data

Table 2. Summary of the Five IRB Application Systems

IRB System	System Type	Review Setting
A	Locally customized commercial eIRB	Academic Medical Center
B	In-house developed eIRB	Academic Medical Center
C	Word templates submitted via e-mail or hard copy	Academic Medical Center
D	Ad hoc eIRB	Independent Commercial
E	Mixed submission method combining e-mail and Software as a Service (Saas) eIRB	Federal Centralized IRB

The number of application forms and the content design of each form vary across the five organizations. Some of the variations are related to the differences in review model and review scope of each IRB. Example form types include initial application form and continuing review application form for health sciences or for social and behavioral sciences. Most organizations also require additional application forms for requesting IRB exemption and for studies involving medical devices, specimens, data repositories, or vulnerable populations. Crossover fields in different forms were included only once for the purpose of mapping.

Table 3. Summary of the Mapping Results

IRB System	Total # of fields	Exact Mapping[1]	Equivalent Mapping[2]	Partial Mapping[3]	Supportable Mapping[4]	Derivable Mapping[5]	Out of Scope[6]	Not Defined[7]	Unclear[8]
A	280	23.9%	14.3%	17.1%	21.1%	5.4%	7.5%	*9.6%*	1.1%
B	241	14.9%	20.7%	35.7%	7.5%	6.6%	4.9%	*8.7%*	0.8%
C	263	20.5%	23.6%	16.3%	10.6%	9.1%	2.7%	*13.7%*	3.4%
D	302	13.6%	8.3%	32.5%	18.2%	2.6%	14.9%	*9.6%*	0.3%
E	141	17.7%	6.4%	21.3%	22.7%	2.8%	16.3%	*12.1%*	0.7%

[1] Exact mapping: the form field can be exactly mapped to an attribute of a class in the Model.

[2] Equivalent mapping: the form field can be semantically equal-mapped to the Model by combining more than one attributes from one or more classes.

[3] Partial mapping: the Model has a general attribute covering more than one related form fields but lacks the specificity defined in the form fields.

[4] Supportable mapping: the form field is supported by defining value set(s) for a certain attribute in the Model.

[5] Derivable mapping: the form field cannot be directly mapped to a class or attribute in the Model but it can be derived from other attribute(s).

[6] Out of scope: the form field is defined according to local regulations or policies and it is intentionally excluded from the core model. However, it is possible to extend the Model to support such local policies.

[7] Not defined: the Model does not have a corresponding class or attribute defined for the form field.

[8] Unclear: the definition of the form field is not clear.

3.2 Evaluation Results

The evaluation results produced granular and complicated mapping categories but showed overall agreement between these real systems and our model. As shown in Table 3, about 10-15% of items defined by the IRB organizations are not covered in the Model. These undefined items reflect elements that are not covered by the current knowledge sources used for developing the Model but are worth considering due to the important role they play in human subject protection. During the mapping analysis, several areas such as humanitarian use device and use of radioactive drugs were identified that are regulated in federal laws but that are not included in AAHRPP guidelines. Therefore, corresponding classes and attributes in these areas were added to the Model after the mapping analysis. There are also undefined form fields in the Model that are based on best practices. For example, some IRBs ask for extra details for placebo-controlled studies. Federal regulations do not address this specific type of study but some IRBs require more information due to potential risks posed by this type of study. A collection of such specializations is valuable because it can inform best practices in the IRB domain.

The model evaluation process revealed limitations of the IRB DAM in representing certain aspects of real-world IRB applications, especially in defining information elements about subjective evaluations and justifications from investigators for a certain study activity, or foreseen events that can be fully expressed only with free text. This limitation is caused by the nature of information models whose strength lies in representing discrete and machine-understandable data elements but not free-text information.

4. Discussion

Although our prototype implementation of an IRB DAM showed decent coverage across several IRB types, a formal evaluation of performance improvement and user satisfaction when the model is used as part of a eIRB-CRI integrated workflow is needed to make the proposed solution more convincing. The development of the IRB DAM was initially motivated by addressing the need to integrate a local CRI data query system and an eIRB system to realize automated access control of PHI based on IRB approval.[34] However, the value of the IRB DAM extends beyond this use case. With a standard IRB model, an eIRB system can in theory be integrated with any CRI system (e.g., clinical trial management systems, electronic data capture systems, clinical trial registries, other eIRBs, etc.) to streamline the clinical research workflow, which will be part of our future work to demonstrate the utility of the IRB model.

The author who developed the model performed the evaluation of the model, which may cause some mapping biases. However, the actual value of the evaluation does not lie in the specific numbers listed in the mapping result (Table 3). As demonstrated by the complicated mapping results, the purpose of the evaluation is not to categorize the model simply as "good" or "bad." There is no gold standard regarding to the design of IRB application forms. The five IRB application systems chosen in the evaluation phase are representative, but this does not mean their application forms are perfect. The comparison between the IRB model and the real-world IRB application forms identified several important insights about the model:

- The model has covered the core information elements required by representative IRBs, as shown by the small percentage of the *Not Defined* mapping category;

- For fields not covered in the model, some belong to local context and should not be included in the core model, and some may be potentially considered in the future version of the model and national guidelines according to best practices;

- Items that are currently free text can be defined in a structured format according to the model as indicated by the *Equivalent mapping* category;

- There are ambiguous fields in the current IRB application forms that should be clarified, as indicated by the *Unclear* mapping category;

- The evaluation process identified valuable sources for future development of the domain vocabulary for the Model since many *Supportable Mapping* fields suggest the possible values for a certain attribute in the Model.

In short, the evaluation provided insight with regard to future improvement to both the IRB model development and real-world IRB application form design.

Like any modeling effort, developing an IRB DAM that meets real-world application requirements needs many rounds of iteration and revision. As future work, we envisage continuing iterative development of the model by collaborating with more IRB domain experts and clinical researchers. A formal evaluation of the expressiveness of the data request related classes needs to be performed, possibly by annotating previously submitted IRB applications. We plan to promote the adoption of the IRB model by collaborating with other CTSA centers that have eIRBs. At the same time, we plan to integrate the IRB DAM with the BRIDG model through a harmonization process developed by BRIDG.

Structured and computable IRB application information could provide the foundation for automated review decision support with predefined rules, thus enhancing review quality and efficiency. We plan to develop more advanced applications of the IRB model such as facilitated computer decision support for IRB reviewers in our institutional eIRB system. For example, applications could be automatically assigned to IRB members with corresponding experience or expertise based on the area of research proposed. Also, depending on whether the planned study

activities involve high-risk procedures such as ionizing radiation, or informational risks such as secondary use of existing data or specimens, commensurate review procedures for risk can be suggested by the system. This could help IRBs allocate their limited resources to ensure human subject protection while enabling responsible research to proceed.

5. Conclusion

This paper focuses on the technical aspect of the issues in the IRB oversight domain. We described the development and evaluation process of the IRB DAM in detail. We demonstrated that the IRB model is broadly representative across a variety of IRB types. This paper fills a gap in standardization and modeling efforts in the clinical research informatics domain of IRB oversight and support.

Acknowledgements

References
1. Iii HM, Dorsey ER, Matheson DHM, Thier SO. Financial Anatomy of Biomedical Research. JAMA. 2005;294(11):1333–42.
2. Catania JA, Lo B, Wolf LE, Dolcini MM, Pollack LM, Barker JC. Survey of U.S. Human Research Protection Organizations: Workload and Membership. J Empir Res Hum Res Ethics. 2008;I(4):57–69.
3. Burman WJ, Reves RR, Cohn DL, Schooley RT. Breaking the camel's back: multicenter clinical trials and local institutional review boards. Ann Intern Med [Internet]. 2001 Jan 16

[cited 2013 Jun 9];134(2):152–7. Available from:
http://www.ncbi.nlm.nih.gov/pubmed/11177319

4.   Cook AF, Hoas H. Protecting Research Subjects: IRBs in a Changing Research Landscape. IRB. 2011;33(2):14–20.

5.   Nass SJ, Levit LA, Gostin LO. Beyond the HIPAA Privacy Rule : Enhancing Privacy , Improving Health Through Research. Heal. (San Fr. 2009.

6.   Lowrance WW, Collins FS. Ethics. Identifiability in genomic research. Science [Internet]. 2007 Aug 3;317(5838):600–2. Available from:
http://www.ncbi.nlm.nih.gov/pubmed/17673640

7.   Office for Human Research Protections (OHRP) Database for Registered IORGs & IRBs, Approved FWAs, and Documents Received in Last 60 Days [Internet]. [cited 2013 Jan 3]. Available from: http://ohrp.cit.nih.gov/search/search.aspx?styp=bsc

8.   Silberman G, Kahn KL. Burdens on research imposed by institutional review boards: the state of the evidence and its implications for regulatory reform. Milbank Q [Internet]. 2011 Dec [cited 2013 Mar 28];89(4):599–627. Available from:
http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3250635&tool=pmcentrez&rendertype=abstract

9.   Gold JL, Dewa CS. Institutional review boards and multisite studies in health services research: is there a better way? Health Serv Res [Internet]. 2005 Feb;40(1):291–307. Available from:
http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1361138&tool=pmcentrez&rendertype=abstract

10.   Greene SM, Geiger AM. A review finds that multicenter studies face substantial challenges but strategies exist to achieve Institutional Review Board approval. J Clin Epidemiol [Internet]. 2006 Aug [cited 2013 Jun 9];59(8):784–90. Available from:
http://www.ncbi.nlm.nih.gov/pubmed/16828670

11.   Burman WJ, Reves RR, Cohn DL, Schooley RT. Breaking the camel's back: multicenter clinical trials and local institutional review boards. Ann Intern Med [Internet]. 2001 Jan 16 [cited 2013 Jun 9];134(2):152–7. Available from:
http://www.ncbi.nlm.nih.gov/pubmed/11177319

12.   Whitney SN, Alcser K, Schneider C, McCullough LB, McGuire AL, Volk RJ. Principal investigator views of the IRB system. Int J Med Sci [Internet]. 2008 Jan [cited 2013 Jun 21];5(2):68–72. Available from:
http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2288790&tool=pmcentrez&rendertype=abstract

13.   Embi PJ, Payne PRO. Clinical research informatics: challenges, opportunities and definition for an emerging domain. JAMIA [Internet]. [cited 2013 Feb 12];16(3):316–27. Available from:
http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2732242&tool=pmcentrez&rendertype=abstract

14.   Kahn MG, Weng C. Clinical research informatics: a conceptual perspective. JAMIA [Internet]. 2012 Jun [cited 2013 Feb 12];19(1e):e36–42. Available from:
http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3392857&tool=pmcentrez&rendertype=abstract

15.   Richesson RL, Krischer J. Data Standards in Clinical Research: Gaps, Overlaps, Challenges and Future Directions. 2007;14(6):687–96.

16.   CTSA Institutions [Internet]. [cited 2012 Sep 2]. Available from:
https://www.ctsacentral.org/institutions

17.   The Protocol Representation Model Version 1.0. 2010.

18. Fridsma DB, Evans J, Hastak S, Mead CN. The BRIDG project: a technical report. JAMIA. 2008;15(2):130–7.

19. Sim I, Carini S, Tu S, Wynden R, Pollock BH, Mollah SA, et al. The human studies database project: federating human studies design data using the ontology of clinical research. AMIA Summits Transl Sci Proc. 2010. p. 51–5.

20. Code of Federal Regulations Title 45 Part 46. 2009.

21. IRB Guidebook - U.S. Department of Health and Human Services [Internet]. [cited 2013 May 4]. Available from: http://www.hhs.gov/ohrp/archive/irb/irb_chapter6.htm

22. Ferré X. An Evaluation of Domain Analysis Methods. Int Work Eval Model Methods Syst Anal Des. 1999.

23. Cohen SG, Hess JA, Novak WE, Peterson AS. Feasibility Study Feature-Oriented Domain Analysis (FODA) Kyo C . Kang. 1990;(November).

24. Embley DW, Thalheim B. Handbook of Conceptual Modeling: Theory, Practice, and Research Challenges. 2011.

25. Code of Federal Regulations TITLE 45 PUBLIC WELFARE DEPARTMENT OF HEALTH AND HUMAN SERVICES PART 46 PROTECTION OF HUMAN SUBJECTS [Internet]. Available from: http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html#46.102

26. Ambler S. Data Modeling 101 [Internet]. [cited 2013 Apr 3]. Available from: http://www.agiledata.org/essays/dataModeling101.html

27. Evaluation Instrument for Accreditation. 2012.

28. IRB Evaluation Checklist. 2010.

29. HL7 Version 3 Standard: Data Types - Abstract Specification, Release 2 [Internet]. [cited 2013 May 6]. Available from: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=264

30. Obeid J, Gabriel D, Sanderson I, Sc M. A Biomedical Research Permissions Ontology : Cognitive and Knowledge Representation Considerations. Assoc Comput Mach Annu Comput Secur Appl Conf. 2010. p. 9–13.

31. Grando MA, Boxwala A, Schwab R, Alipanah N. Ontological Approach for the Management of Informed Consent Permissions. IEEE Second Int Conf Healthc Informatics, Imaging Syst Biol. IEEE; 2012. p. 51–60.

32. Observational Medical Outcomes Partnership Common Data Model Specifications. 2012.

33. OMG. Business Process Model and Notation (BPMN) Version 2.0 [Internet]. 2011. Available from: http://www.omg.org/spec/BPMN/2.0/

34. He S, Hurdle JF, Botkin JR, Narus SP. Integrating a Federated Healthcare Data Query Platform With Electronic IRB Information Systems. AMIA Annu Symp Proc. 2010. p. 291–5.