Epoch: an Ontological Framework to Support Clinical Trials Management

Ravi D. Shankar1  Susana B. Martins1  Martin J. O’Connor1  David B. Parrish2  Amar K. Das1

1Stanford Medical Informatics, Stanford University School of Medicine, Stanford, CA, USA
2The Immune Tolerance Network, Pittsburgh, PA, USA

{rshankar, smartins, moconnor, akd}@stanford.edu, dparrish@immunetolerance.org

ABSTRACT
The increasing complexity of clinical trials has generated an enormous requirement for knowledge and information specification at all stages of the trials, including planning, documentation, implementation, and analysis. We are building a knowledge-based framework (Epoch) to support the management of clinical trials. We are tailoring this approach to the Immune Tolerance Network (ITN), an international research consortium developing new therapeutics in immune-mediated disorders. In the broad spectrum of trial management activities, we currently target two areas that are vital to the successful implementation of a trial: (1) tracking study participants as they advance through the trials, and (2) tracking biological specimens as they are processed at the trial laboratories. The core of our software architecture is a suite of ontologies that conceptualizes relevant clinical trial domain. Our approach can provide ITN and other research organizations a stable and consistent knowledge source for clinical-trial software applications.

Categories and Subject Descriptors
I.2.4 [Artificial Intelligence]: Knowledge Representation Formalisms and Methods – frames and scripts, representations (procedural and rule-based), semantic networks. I.3 [Computer Applications]: Life and Medical Sciences – medical information systems.

General Terms
Management, Design, Standardization

Keywords
Clinical Trial, Knowledge Base, Ontology, OWL, SWRL

1. INTRODUCTION
Clinical researchers undertake a clinical trial to test the safety and effectiveness of a new drug or procedure in human subjects generally after promising laboratory studies. The increasing complexity of clinical trials has generated an enormous requirement for knowledge and information management at all stages of the trials — planning, specification, implementation, and analysis. In the past few years, there has been a spurt of modeling projects to support applications that target different requirements in the broad spectrum of clinical trial activities, such as trial registry [17], trial authoring [9,11], and trial execution [12,19]. There are ongoing efforts supported by HL7 [7] and CDISC [2] in partnership with entities such as NCI’s cabIG project to develop the BRIDG model [1] that defines functions and behaviors throughout standard clinical trials. This work is aimed at bridging standards within the clinical research and healthcare domains, but does not address the knowledge specification needed for day-to-day activities in clinical trials management.

In this paper, we describe Epoch, an ontological framework that supports the management of clinical trials at the Immune Tolerance Network, or ITN [8,16], which develops new therapeutics for immune-mediated disorders. The Epoch framework is borne out of the collaborative research efforts of Stanford Medical Informatics (SMI) and the ITN in addressing the informatics needs of collecting, managing, integrating and analyzing clinical trial and immunoassay data. Our focus is currently on two application areas: (1) tracking participants of the trial as they advance through the studies, and (2) tracking clinical specimens as they are processed at the trial laboratories. The core of our system is a suite of ontologies that encodes knowledge about the clinical trial domain that is relevant to trial management applications. Our models have been influenced by past and ongoing modeling work, including the CDISC and BRIDG projects. We discuss in this paper, the requirements for knowledge specification and management in ITN clinical trials and how the Epoch framework supports these efforts.

2. CLINICAL TRIAL SPECIFICATION AND IMPLEMENTATION IN THE IMMUNE TOLERANCE NETWORK
The Immune Tolerance Network (ITN) is an international consortium that aims to accelerate the development of immune tolerance therapies through clinical trials and integrated mechanistic (biological) studies. The ITN is involved in planning, developing and conducting clinical trials in autoimmune diseases, islet, kidney and liver transplantation, allergy and asthma, and operates a dozen core facilities that conduct bioassay services. At the ITN, the successful conduct of a clinical trial depends upon the interaction of professionals working for various entities,
including the ITN, contract research organizations, clinical study sites, and core laboratories. Several groups within these entities collaborate in facilitating the specification and implementation of the trials and related biological assay studies (Figure 1). The Protocol Group is primarily responsible for translating a researcher’s clinical study goals into a clinical trial design. The group generates a document that includes the study design, the implementation protocol, the measured outcomes, and other relevant clinical knowledge. The document also includes a Schedule of Events table that describes clinical activities involving study participants at different time points during the study period. ITN clinical trials are augmented by a series of mechanistic studies designed to uncover the basic biological features of clinical tolerance. ITN’s Assay Group chooses a set of special clinical procedures and tests (assays) that will increase understanding of tolerance. The group generates a Specimen Table that has operational information such as clinical specimens that need to be drawn from participants, the specimen collection time points, the specimen processing instructions, the core laboratories where the assays will be conducted, assay instructions, etc. Based on the specimen table, the Operations Group creates a Tube Table that lists all the different types of specimen containers required per participant at different time points during the study. A Tubes Manufacturer consults the tube table and creates groupings (kits) of physical specimen containers and ships them to the clinical sites. ITN has contracted with Cimarron Software, Inc to build a specimen workflow application called ImmunoTrak based on Cimarron’s Laboratory Workflow Systems product. ImmunoTrak will inform clinical trial personnel at the sites on the flow of activities to be performed at a participant’s visit, and will allow the personnel to log the participant’s visit, specimen collection, and shipping and receiving of bar-coded specimen containers. The configuration of ImmunoTrak will be based on the schedule of events, specimen table, the tube table and the kits definition. The Core Labs receive the collected specimens and perform assays as defined by the specimen table. A Contract Research Organization (CRO) is an external group to ITN that creates Case Report Forms (CRFs) used at clinical visits to collect clinical trial data. The structure of the CRFs are based on the schedule of events table generated by the Protocol Group. Throughout the clinical trial period, data collected by ImmunoTrak on participants’ visits, collected specimens, clinical data collected using the CRFs, and results of the assays performed at the Core Labs are sent to ITN’s Data Center for storage and analysis. Several applications use the data to monitor the progress of the clinical trial and to perform trial-related clinical analyses.

3. CLINICAL TRIALS MANAGEMENT ACTIVITIES
Clinical trials management encompasses several activities in conducting different stages of a trial, from the trial’s inception to its completion. Some example activities are coordinating the authoring and approval process of the protocol, recruiting participants, managing clinical and laboratory sites, facilitating and monitoring participant care, collecting and processing clinical specimens, collecting, storing and analyzing trial data, and publishing the trial results. Our work focuses currently on two operational activities: participant tracking and specimen tracking.

3.1 Participant tracking
Participants are recruited into the protocol, and then, they are advanced through the different phases of the protocol plan. The participants are tracked to determine the recruitment status at each clinical site, to monitor the progression of the participants in the trial, to ensure appropriate inventory of clinical supplies, such as specimen containers, at the sites, to gauge participation levels at all sites and across all trials, and, more importantly, to monitor for serious adverse events. A set of participant states can be derived from the protocol plan and the associated schedule of events. Some example states are potential, eligible, not eligible, enrolled,
active, withdrawn, suspended, and completed. Participants need to be tracked at different levels of granularity, such as participant states, phases, study visits, and clinical activities.

3.2 Specimen tracking
Mechanistic specimens are collected from participants at different visits based on clinical assessments and clinical studies (biological assays) planned in the protocol. These specimens are then stored in pre-determined containers and shipped to bio-repositories. The specimens (or portions of them) are shipped to the core laboratories that can perform specific assays on the specimens. The assay results are then sent to a data warehouse for storage and subsequent analysis. The bio-repositories may also archive portions of the specimens for future interrogation. The trials managed by ITN generate enormous amount of specimen traffic across different sites. Tracking the specimen from the point of collection to the point of processing and archival becomes paramount to maintain the integrity of the operation. Appropriate type and number of specimen containers should be stocked at the clinical sites in preparation for the anticipated participant visits. At the time of a participant’s visit, appropriate specimens should be collected and stored in matching containers. The containers are shipped to the bio-repositories, and then to the core laboratories based on the shipping instructions in the specimen table and the specimen flow of the protocol. Specimens have to be accounted for at all times using shipping and receiving logs.

A general approach to clinical trials management is to employ disparate applications such as laboratory management systems and hospital information systems that monitor specific sections of the participant and specimen workflows. The information generated by these applications along with data out of loosely controlled sources such as spreadsheets, documents and email messages are then assembled to determine the operational state of the clinical trial. The lack of common nomenclature among the different sources of the tracking information and the unreliable nature of the data generation can lead to significant operational and maintenance challenges. Studies need to be tracked for the purposes of general planning, gauging progression, monitoring patient safety, and managing personnel and clinical resources. The tracking effort is compounded by the fact that an ITN trial often is carried out at multiple sites, geographically distributed, sometimes across the world.

4. EPOCH FRAMEWORK FOR CLINICAL TRIALS
We are building a knowledge-based framework called Epoch that can support the undertaking of multi-site clinical trials. We are specifically developing three types of methods:

1. Knowledge acquisition methods that use a standardized knowledge representation (ontology) to annotate protocol and assay specific elements with metadata and that permit specification of knowledge on immune disorders
2. Ontology-database mapping methods to integrate the knowledge base of study metadata and biomedical knowledge with primary data stored in the ITN data repository
3. Concept-driven querying methods for the data repository to support integrated data management plans and create high-level, mechanistic-oriented abstractions for data analysis.

A clinical trial protocol (the plan for a trial) lays out specification,
implementation and data analysis details. For example, it includes the reason for undertaking the study, the number of participants that will be in the study and the recruitment process, the sites (clinical and laboratory) where the study will be conducted, the study drug that the participants will take, the medical tests that the participants will undergo, the data that will be collected, and the statistical analyses that will be performed on the data. Our methods use a structured and standardized knowledge representation that conceptualizes the protocol entities relevant to our management applications. We highlight four pieces of protocol concepts that are required to support these activities:

1. The **protocol schema** divides the temporal span of the study into phases such as the treatment phase and follow-up phase, and specifies the temporal sequence of the phases. It also includes information on the arms of the protocol.

2. The **schedule of events** enumerates a sequence of protocol visits that are planned at each phase, and, for each visit, specifies the timing and a list of protocol events (assessments, procedures and tests) that are planned at that visit.

3. The **specimen table** lays out the clinical specimens that will be collected from the participant, when they will be collected, and the assays (special analyses) that will be performed on them.

4. The **specimen flow** describes the workflow associated with the processing of the specimens. The specimens are shipped from the collection site to bio-repository sites and core laboratory sites where they are assayed.

At the core of our knowledge management architecture (Figure 2) is a suite of ontologies:

- The **clinical ontology** includes terms that specify clinical and biological knowledge on immune tolerance disorders and other concepts relevant to ITN clinical trials.

- The **protocol model** is a knowledge model of the clinical trial protocol. It simplifies the complexity inherent in the full structure of the protocol by focusing only on concepts required to support participant and specimen tracking applications. Other concepts are either ignored or partially represented. The main concepts represented in the protocol ontology are the protocol schema and the schedule of events. The temporal constraints associated with participant visits and protocol events are specified using a temporal model developed in our laboratory [14].

- The **specimen ontology** models the workflow of specimens – collection, shipping and processing workflow of specimens at the clinical laboratory and bio-repository sites.

- The **assay ontology** models characteristics of mechanistic studies relevant to immune disorders. An assay specification includes the clinical specimen that can be analyzed using that assay, and the workflow of the specimen processing at the core laboratories.

- The **site model** provides a structure to store site-related data such as protocols implemented at the site, participants on each protocol, relevant clinical resources and study coordinators.

- The **specimen container ontology** is a catalog of different specimen containers. Each container is characterized with attributes such as associated specimen type, container size, manufacturer, additives, material, closure type, processing instructions and shipping instructions.

- The **virtual trial data ontology** that encapsulates the study data that is being collected, such as participant clinical record, specimen workflow logs, and site related data.

Figure 3 shows a subset of concepts represented in the ontologies and the relationships among these concepts. These ontologies are used to annotate protocol and assay specific elements with metadata. The annotations and the relationships between them can then be used by different tools that participate in tracking participants and specimens. The Epoch ontologies thus provide a common nomenclature and semantics required to support an integrated and consistent clinical trial management. A suite of applications allows different ITN groups to specify the protocol, the specimen table and the assays for a clinical trial using the Epoch ontologies. The resulting knowledge base is then used to configure another set of applications that collect trial data at the clinical site. The clinical trial monitoring applications such as the participant and specimen tracking applications then employ query methods driven by the Epoch knowledge base to analyze the collected trial data.
5. EPOCH MODELING METHODOLOGY
We have developed the Epoch ontologies in OWL Web Ontology Language [15]. OWL is a W3C standard language for use in Semantic Web where machines can provide enhanced services by reasoning with facts and definitions expressed in OWL. An OWL ontology consists of classes, properties and individuals. Classes are interpreted as sets of objects that represent individuals in the domain of discourse. Properties are binary relationships that link individuals. We have built hierarchies of classes representing the concepts in the Epoch ontologies. We then create individuals of the protocol ontology to encode specific clinical trials. The clinical ontology, the assay ontology and the specimen container ontology are created as repositories that can be used to uniformly specify all clinical trials. We also represent collected trial data as individuals of OWL classes in the data model, and thus facilitate mechanisms for reasoning with the data using the Epoch ontologies. SWRL, the Semantic Web Rule Language [18], is a W3C recommendation for a rule language that can be used to express rules in terms of OWL concepts and that can reason about OWL individuals. We have used SWRL to specify constraints in our ontologies, and also investigating ways to adapt SWRL for our query methods. For example, we use SWRL to specify temporal constraints found in the protocol ontology and the assay ontology. Figure 4 illustrates a SWRL rule to specify the constraint: On days that both immunotherapy and omalizumab are administered, omalizumab will be injected 60 minutes after the immunotherapy.

Protégé [10] is a software tool that supports the specification and maintenance of terminologies, ontologies and knowledge-bases. Protégé has several software plug-ins including an OWL editor and a SWRL editor (shown in Figure 4). We used Protégé to create the ontologies in OWL. We, then, entered specific protocols and assays using Protégé’s knowledge-acquisition facilities.

6. IMPLEMENTING SPECIMEN TRACKING – AN ILLUSTRATION OF THE EPOCH FRAMEWORK
A high-level functional architecture of the Epoch framework consists of several components interacting with each other to provide knowledge based reasoning and query methods (Figure 5).

The Epoch Knowledge Base contains the ontologies enumerated in Section 2. It also stores specific instantiations of the ontologies for different clinical trials. The repository uses a file backend to store the ontologies.

The Knowledge Base Server provides a programmatic interface (API) that other components can use to access the contents of the ontology repository. The domain specific API is built on top of the generic Protégé-OWL API.

The Clinical Trial Database is a relational database system that stores data related to the implementation and execution of clinical trials. The types of data include participant enrollment data, specimen shipping and receiving logs, participant visits and activities, and clinical results.

The Model-Database Mapper facilitates runtime access to relational data in the Clinical Trial Database as instances of the Epoch data model. It uses a mapping ontology to connect data model concepts to database entities i.e. properties of an OWL class are mapped to columns of a relational table. The BioSTORM disease-surveillance framework [4] employs techniques to map disparate data sources to a data model. We are translating these methodologies to use OWL, SWRL, and rule-based inference engines such as Jena. We are also investigating the possibility of using D2RQ [5], a language to describe mappings between relational database schema and OWL/RDFS ontologies.

The Inference / Rule Engine executes temporal and non-temporal constraints – that have been expressed as SWRL rules - in Epoch ontologies. We have developed a SWRL built-in deployment module that provides a general mechanism to define Java implementations of SWRL built-ins, dynamically load them, and invoke them from a rule engine. It interfaces with the Model-Data
Figure 5 A high-level functional architecture of the Epoch framework

Mapper to allow SWRL rules to be executed on data stored in the Clinical Trial Database. Here is an example of a SWRL rule to check if a participant’s visit time fell within that visit’s time window:

\[
\text{VisitRecord}(\text{?vrecord}) \land \\
\text{hasVisitId}(\text{?vrecord}, \text{?vid1}) \land \\
\text{hasParticipantId}(\text{?vrecord}, \text{?pid}) \land \\
\text{temporal:hasValidTime}(\text{?vrecord}, \text{?vtO}) \land \\
\text{Visit}(\text{?v}) \land \\
\text{hasVisitId}(\text{?v}, \text{?vid2}) \land \\
\text{swrlb:equal}(\text{?vid1}, \text{?vid2}) \land \\
\text{hasStartCondition}(\text{?v}, \text{?vsc}) \land \\
\text{temporal:inside}(\text{?vtO}, \text{?vsc})
\]

→

The empty consequent of the rule indicates that this rule is formulated as a query. This rule uses a built-in temporal:inside that takes in as arguments a time and a relative variable interval, and returns true if the time point is within the interval, and returns false otherwise. We are using JESS, a production rule-engine, to selectively execute the rules based on the context. For example, the rule that specifies the constraint on a visit time window will alone need to be executed when checking if a specific participant’s visit satisfied the constraint.

The Utility Functions consist of a set of generic modules that facilitate import and export of the knowledge base in different formats such as XML, concept-driven query methods, constraint satisfaction methods, etc.

The Clinical Trials Management Applications interoperate with the Epoch functional components at syntactic, structural and semantic levels to support the management of clinical trials.

Figure 6 shows the interaction of the components of the Epoch architecture to implement specimen tracking. The first step is to specify the specimen workflow in the Epoch knowledge base using the Protégé-OWL editor. Next, an export utility function generates a configuration file in XML format. This configuration file consists parts of the knowledge base required to configure ImmunoTrak, the specimen workflow system. At the execution time of the clinical trial, study coordinators at different clinical sites access ImmunoTrak to enter specimen collection, processing and shipping data. This data is stored in the Clinical Trial Database. The Specimen Tracking Application employs the Model-Database Mapper to access the tracking data through the Epoch data models. Using the Query / Rule Engine, the application can satisfy user queries for specific specimen collection and processing status, and execute any validation rules or temporal constraints, as specified in the ontologies, on the tracking data.

7. CONCLUSION

Participant tracking and specimen tracking are clinical trials management activities vital to the successful implementation of a clinical trial protocol, but that are difficult to specify, coordinate and enact in a consistent and integrated manner. In our work, we have developed a novel ontology-based framework that can inform professionals how to undertake these activities. To disseminate our knowledge into software applications for clinical
trials management, we must extend our framework into the areas of knowledge acquisition and data integration.

Developers who are familiar with the Epoch ontologies and domain experts who have relevant medical knowledge work as a team to encode clinical trial protocols. We have modularized the ontologies so that domain experts are exposed only to parts of the knowledge base relevant to them. We are building rich graphical user interfaces so that domain experts can enter and review protocol knowledge with minimal exposure to the complexities of the ontologies or Protégé.

We are also actively working on methodologies to integrate clinical trial data such as the specimen workflow data, participants’ clinical data, and assay results. We also employ the techniques we have developed in our research group to analyze and integrate temporal databases. Most protocol data will continue to reside in relational databases. To support analysis at the knowledge-level, we must create mechanisms to map relational data into knowledge-level data. We use a mapping tool called Synchronus [6] to perform run-time transformations of data from relational tables to the extended propositions described by the OWL temporal model. However, transforming all information to OWL individuals would impose an excessive run-time overhead and the system would scale poorly. Thus, instead of generating OWL individuals for each datum, it creates customized in-memory objects. A further optimization is direct mapping of temporal SWRL queries to Chronus II, a temporal database mediator that we developed and tested; it extends the standard relational model and the SQL query language to support valid-time temporal queries [14]. Instead of mapping relational tuples into memory for processing, we operate on those rows directly using Chronus II queries. Query results are then read into memory. This process is described in more details in a previous publication [13]. As mentioned earlier, we are also experimenting D2RQ, a language for mapping relational schema to OWL.

The Epoch ontological framework supports the consistent specification and implementation of clinical trials undertaken by large, distributed research consortiums, such as the ITN. We have been able to design ontologies and specific protocol knowledge bases that can drive participant tracking and specimen tracking applications. Our framework can also scale to implement other clinical trials management services, such as site management, core management, study tracking, and analyses across multiple trials.

8. ACKNOWLEDGMENTS
This work is supported in part by the Immune Tolerance Network, which is funded by the National Institute of Health under the grant NO1-AI-15416).

9. REFERENCES


[15] OWL Overview: http://www.w3.org/TR/OWL-features/


[18] SWRL Specification: http://www.w3.org/Submission/SWRL/