A Knowledge-Based System for Managing Clinical Trials
Ravi D. Shankar, MS¹, Susana B. Martins, MD, MSc¹, Martin O’Connor, MSc¹,
David B. Parrish, MS², Amar K. Das, MD, PhD³
¹Stanford Medical Informatics, Stanford University, CA, USA;
²The Immune Tolerance Network, Pittsburgh, PA, USA

Abstract
Clinical trials encompass a vast array of studies in treatment, prevention or diagnosis of medical conditions. There is an enormous requirement for knowledge and information management at all stages of the trials – planning, specification, implementation, and analysis. We are building Epoch, a knowledge-based system to manage clinical trials in the Immune Tolerance Network (ITN) in developing new therapeutics in immune-mediated disorders. In the broad spectrum of trial management activities, we are currently targeting two areas that are vital to the successful implementation of a trial – tracking study participants as they advance through the studies, and tracking clinical specimens as they are processed at the trial laboratories. The core of our system is a suite of medical ontologies that conceptualizes the clinical trial domain relevant to our participant and specimen tracking applications. We expect that our ontological solution can provide a stable and consistent platform to support our current and future requirements of clinical trial management.

1. Introduction
Clinical trials encompass a vast array of studies in treatment, prevention or diagnosis of medical conditions. The increasing complexity of the 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 efforts to support solutions that target different requirements in the broad spectrum of the trial management activities – trial registry [1], trial authoring [2][3], and trial execution [4][5]. Significant progress has been made to automate clinical guidelines [6][7] and the underlying frameworks can be applied to automate patient therapy during clinical trials. There is a major ongoing effort by HL7 (http://www.hl7.org/) and CDISC (http://www.cdisc.org/standards/) in partnership with entities such as NCI’s caBIG project (http://cabig.nci.nih.gov/) to develop the BRIDG model (http://www.bridgproject.org/) that defines functions and behaviors throughout standard clinical trials.

We describe a knowledge-based system called Epoch to manage clinical trials in the Immune Tolerance Network (ITN) in developing new therapeutics in immune-mediated disorders. We are currently building applications to track study participants as they progress through the studies, and to track clinical specimens as they are processed at the trial laboratories. The core of our system is a suite of medical ontologies that conceptualizes the clinical trial domain relevant to our participant and specimen tracking applications.

2. Participant and specimen tracking in clinical trials
Clinical researchers undertake a clinical trial generally after laboratory studies indicate the promise of a new drug or procedure. A clinical trial protocol (the plan for a trial) that is created lays out specification, implementation and data analysis details. For example, it includes the reason for doing 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.

2.1 Protocol concepts relevant to managing clinical trials
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. We have focused our work on two operational activities – participant tracking and specimen tracking. We highlight four pieces of protocol concepts (Figure 1) that are required to support these activities:

- 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.
- 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.
- 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.
- 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.

### 2.1 Participant tracking

Participants are recruited into the protocol and then they are advanced through the different phases of the protocol schema. A set of participant states can be derived from the protocol schema and the schedule of
events. Some example states are – potential, eligible, not eligible, enrolled, active, withdrawn, suspended, and completed. Participants can then be tracked at different levels of granularity – participant states, phases, visits, and events. Some of the reasons for tracking participants are to determine the recruitment status at each clinical site, to monitor the progression of the participants in the protocol, to ensure appropriate inventory of clinical supplies, such as specimen containers, at the sites, and to monitor participation levels at all sites and across all protocols.

2.2 Specimen tracking

Clinical specimens are collected from participants at different visits according to the clinical studies (assays) planned in the protocol. These clinical specimens are then stored in pre-determined containers and shipped to bio-repositories. The specimens (or portions of them) are sent 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.

General approach to participant tracking and specimen tracking 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 and emails are then assembled to determine the operational state of the protocol. The lack of common nomenclature among the different sources of the tracking information and the unreliable nature of the data generation leads to significant operational and maintenance issues.
3. Epoch, a knowledge-based system to track participants and specimens

We are developing software methods that support the participant and specimen tracking of the clinical trials. Our methods that we collectively call as the Epoch framework use a structured and standardized knowledge-representation that conceptualizes the protocol entities relevant to our target applications. 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 ontology 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 (see Section 2.1). The temporal constraints associated with participant visits and protocol events are specified using a temporal model developed in our laboratory [8].

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 (see Section 2.1).

The site ontology 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.

These ontologies are used to annotate protocol and assay specific elements with metadata. The annotations and the relationships between them (Figure 3) can then be consumed by different tools that participate in tracking participants and specimens. Thus, the ontologies will provide a common nomenclature and semantics required to support an integrated and consistent clinical trial management.

Figure 3 Relationships among a subset of protocol concepts

Figure 4 Protege - the OWL and SWRL editors
3.1 Knowledge-acquisition methodology

We developed these ontologies using OWL (the Web Ontology Language proposed by W3C) by building hierarchies of classes describing concepts in the ontologies and relating the classes to each other using properties. OWL can also represent data as instances of OWL classes—referred to as individuals—and also provides mechanisms for reasoning with the data and manipulating it. OWL also provides a powerful constraint language for precisely defining how concepts in an ontology should be interpreted. We use SWRL (the Semantic Web Rule Language) to specify temporal constraints found in the protocol ontology and the assay ontology. For example, 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 [9] 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 (Figure 4). We used Protégé to create the ontologies in OWL. We, then, entered specific protocols and assays using Protégé’s knowledge-acquisition facilities.

4. Specimen Tracking – an illustration of the Epoch system

We illustrate the Epoch system as it applies to specimen tracking (Figure 5). To implement a specific protocol, we create a knowledge base by entering the protocol details using the ontologies in the Epoch framework. This knowledge base will include the protocol schema, the schedule of events, the specimen table and the specimen flow. We then configure the specimen workflow system with relevant protocol knowledge on collection and processing of specimens. The configuration parameters are directly obtained from the protocol knowledge base previously created in Epoch.

The protocol study coordinator queries the knowledge base for details on the specimen containers required at each site. An order for the containers is sent to the manufacturer. The manufacturer sends the containers directly to the sites, and the container inventory (site knowledge base) is updated. At a participant visit, the study coordinator is presented with the assay events for that visit, and shuttles the
appropriate specimen containers to the lab coordinator. The lab coordinator draws the specimens from the participant, stores them in specific containers. The containers are then transported to the bio-repository sites and core laboratory sites in accordance with the specimen workflow. All the containers are bar coded and shipping and receiving logs are maintained at the sites.

The specimen tracking application uses the Epoch knowledge base to collate data from the specimen workflow system. The application can then generate dynamic reports on all aspects of the specimen management such as the status of specific specimens, and specimen container inventory at the sites.

5. Conclusion

Participant tracking and specimen tracking are vital to a successful implementation of a clinical trial protocol. We described an ontology-based system to provide a consistent and integrated solution. We continue to develop and improve on our framework especially in the areas of knowledge acquisition and data integration. We plan to build rich graphical user interfaces so that domain experts can enter protocol knowledge with minimal exposure to the complexities of the Epoch ontologies. We are actively working on methodologies to integrate clinical trial data such as the specimen work flow data, participants’ clinical data, and assay results. We will employ the techniques we have developed in our laboratory to analyze and integrate temporal databases. We believe that our knowledge-based system can scale to implement other clinical trials management activities such as site management, core management and study tracking.

6. Acknowledgement

The authors would like to thank Mark Musen, Daniel Rubin, and Samson Tu for their comments on our work. 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.

7. References


