Epoch Clinical Trial Ontologies to Support Clinical Trials Management

Ravi D. Shankar[1] Susana B. Martins[1] Martin J. O'Connor[1] David B. Parrish[2] Amar K. Das[1]

[1]Stanford Medical Informatics, Stanford University School of Medicine, Stanford, CA, USA

[2]The 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 called 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 three areas that are vital to the successful implementation of a trial: (1) tracking study participants as they advance through the trials, (2) tracking biological specimens as they are processed at the trial laboratories, and (3) visualization of clinical trial data. The core of our software architecture is a suite of ontologies that conceptualizes relevant clinical trial domain. In our presentation, we will discuss the Epoch suite of clinical trial ontologies, and the specification of clinical trials using our ontologies. We will demonstrate how we use our ontologies to configure clinical trial data collection and visualization applications. We will discuss how our current work supports semantic interoperability among clinical trial applications using semantic web technologies.

Categories and Subject Descriptors

I.2.4 [**Artificial Intelligence**]: Knowledge Representation Formalisms and Methods – *frames and scripts, representations (procedural and rule-based), semantic networks.* J.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. The lifecycle management of a complex clinical trial typically involves multiple applications facilitating activities such as trial design specification, clinical sites management, laboratory management, and participants tracking. These disparate applications form a suite of subsystems that are banded together as a clinical trial management system. 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. The subsystems support different but related aspects of a clinical trial, and require clinical trial data flow and knowledge exchange between the subsystems. Thus, there is a strong impetus to integrate these diverse subsystems at syntactic, structural and semantic levels so as to improve clarity, consistency and correctness in specifying clinical trials, and in acquiring and analyzing clinical data.. The situation becomes especially critical with the need to manage complex clinical trials at various sites, and to facilitate meta-analyses on trials.

2. EPOCH FRAMEWORK FOR CLINICAL TRIALS

We are building Epoch, a knowledge-based approach to support a suite of clinical trial management subsystems. Our initiative uses semantic technologies to provide a consistent basis for the subsystems to interoperate. We are adapting this approach to the Immune Tolerance Network (ITN), 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. Many groups, internal and external to ITN, collaborate in facilitating the specification and implementation of the trials and related biological assay studies. Therefore, 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. Studies need to be tracked for the purposes of general planning, gauging progression, monitoring patient safety, and managing personnel and clinical resources. The management effort is especially compounded by the fact that an ITN trial often is carried out at multiple sites, geographically distributed, sometimes across the world.

Our focus is currently on three application areas: (1) tracking participants of the trial as they advance through the studies, (2) tracking clinical specimens as they are processed at the trial laboratories, and (3) visualization of clinical trial data that includes tracking data, clinical data and results of mechanistic studies. 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. 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.

We have developed the Epoch ontologies in OWL Web Ontology Language. 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. 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. 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, 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 to support mapping of our ontologies to other knowledge formats, and also investigating ways to adapt SWRL for our query methods. We have described our approach extensively elsewhere [1].

3. EPOCH DEMONSTRATION

We will present the Epoch suite of ontologies and discuss the use of these ontologies in encoding clinical trials. Protégé 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. We used Protégé to create the ontologies in OWL. We, then, entered specific protocols using Protégé's knowledge-acquisition facilities. Typically, specifying a protocol is a collaborative effort by personnel with different specialties. For example, the protocol group specifies the clinical trial protocol, the assay group designs mechanistic studies associated with the trial, and the operations group works on the implementation details of the protocol. We have built additional software modules that we call KWIZ to create views of the protocol knowledge base to provide focused working areas for domain users to encode parts of the protocol they are responsible for. We have also created a protocol browser that displays encoded information using domain-specific graphical user interfaces. We will show the Epoch ontologies and example encodings of clinical trials in Protégé and in KWIZ. We will also show the domain-specific view of the OWL ontologies using our protocol browser.

We have used the Epoch ontologies and instantiations of these ontologies to drive ITN's clinical trial management applications that support participant-tracking, specimen-tracking, and clinical trial data visualization. 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. The participant-tracking application tracks participants at different levels of granularity, such as participant states, phases, study visits, and clinical activities. 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. The specimen-tracking application uses a workflow specification to track the movement of specimens across different clinical trial facilities. The clinical trial data includes data generated from the participant-tracking application, specimen-tracking application, participants' clinical data, enrollment data, mechanistic study results, etc. A visualization application presents the data to clinical trial personnel using query-driven user interfaces. All these three applications are initially configured with protocol specifications such as participant's visit flow and the specimen work flow. We have created techniques using the SWRL language to automatically configure these applications with protocol encodings created using the Epoch ontologies. We will demonstrate our methodology that supports the sharing of semantics among clinical trial management applications. We will also discuss our continuing work in mapping the clinical trial data to the Epoch ontologies, and hence, support knowledge-driven queries on data. Thus, we will show how our approach supports semantic interoperability among clinical trial management applications using semantic web technologies.

1. R. D. Shankar, S. B. Martins, M. J. O'Connor, D. B. Parrish, A. K. Das. Towards Semantic Interoperability in a Clinical Trials Management System. The 5th International Semantic Web Conference (ISWC 2006), Athens,GA,2006