

AN ONTOLOGICAL APPROACH TO REPRESENTING AND REASONING WITH TEMPORAL CONSTRAINTS IN CLINICAL TRIAL PROTOCOLS

Ravi D. Shankar, Susana B. Martins, Martin J. O'Connor and Amar K. Das
Stanford Medical Informatics, Stanford University
251 Campus Drive, MSOB X215, Stanford, California, 94305, USA

Keywords: Ontology, temporal reasoning, clinical trials, biomedical informatics, Semantic Web, OWL.

Abstract: Temporal constraints play an important role in the specification and implementation of clinical trial protocols, and subsequently, in the querying of the generated trial data. Protocols specify a temporal schedule of clinical trial activities such as tests, procedures, and medications. The schedule includes temporal constraints on the sequence of these activities, on their duration, and on potential cycles. In this paper, we present our approach to formally represent temporal constraints found in clinical trials. We have identified a representative set of temporal constraints found in protocols to study immune tolerance. Our research group has developed a temporal constraint ontology that allows us to formulate the temporal constraints to the extent required to support clinical trials management. We use this ontology to provide temporal annotation of clinical activities in an encoded clinical trial protocol. We have developed a temporal model that represents time-stamped data and facilitates interval-based temporal operations on the data. Using semantic web technologies, we are building a knowledge-based framework that integrates the temporal constraint ontology with the temporal model to support queries on clinical trial data. Using our approach, we can formally specify temporal constraints, and reason with the temporal knowledge to support management of clinical trials.

1 INTRODUCTION

Clinical trials are formal studies on participants to systematically evaluate the safety and efficacy of new or unproven approaches in the prevention and treatment of medical conditions in humans. A clinical trial protocol is a document that includes study objectives, study design, participant eligibility criteria, enrollment schedule, and study plan. It specifies a temporal schedule of clinical trial activities such as tests, procedures, and medications. The schedule includes temporal constraints on the sequence of these activities, on their duration, and on potential cycles. A temporal constraint is defined as an interval-based temporal annotation on a domain entity in relationship with other entities. Temporal constraints are fundamental to the descriptions of protocol entities, such as the following specifications: *Participants will be enrolled at least two days apart; Participant is ineligible if he/she had vaccination with a live virus within the last 6 weeks before enrollment; The first*

dose will be infused over a minimum of 12 hours; Visit 10 for the participant occurs 3 weeks \pm 2 days from the day of transplant. There is an enormous requirement on the execution of a clinical trial to conform to the temporal constraints found in the protocol. 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 a trial often is carried out at multiple sites, geographically distributed, sometimes across the world. The validity of the findings of the clinical trial depends on the clinical trial personnel and the participants performing clinical trial activities as planned in the protocol. More importantly, the treatment and assessment schedules should be strictly followed to ensure the safety of participants.

We have developed an ontological framework that we call Epoch (Shankar et al., 2006), to support the management of clinical trials at the Immune Tolerance Network, or ITN (Rotrosen et al., 2002)

(<http://www.immunetolerance.org/>). As part of this effort, we have developed a suite of ontologies that, along with semantic inferences and rules, provide a formal protocol definition for clinical trial applications. We use the OWL Web Ontology language (<http://www.w3.org/2004/OWL/>), which 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. Central to our ontological effort is the modeling of temporal constraints that we identified in clinical trial protocols. We have created a temporal constraint ontology to formally represent temporal constraints. The ontological representation can then be used to construct rules that can be used in turn, for reasoning with temporal constraints. Thus, at protocol specification phase, a domain expert can capture the essence of temporal constraints using higher-level ontological constructs. At a later time, a software developer can fully encode the constraints by creating rules in terms of temporal patterns and other protocol entities in the ontologies. We are using SWRL, the Semantic Web Rule Language (<http://www.w3.org/Submission/SWRL/>) to write the rules. At execution time of the protocol, the rule elements use the protocol knowledge specified in the Epoch ontologies, and the clinical trial data collected in the clinical trial databases to reason with the temporal constraints. In this paper, we discuss our work in identifying temporal constraints found in ITN's clinical trial protocols. We then discuss our temporal constraint ontology using some patterns that we found in the temporal constraints. We then show how we use the temporal constraint ontology along with other Epoch ontologies to create rules that are then executed at runtime to support clinical trial management.

2 TRIAL CONSTRAINTS

A clinical trial protocol defines a protocol schema that 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 a schedule of activities that enumerates a sequence of protocol visits that are planned at each phase, and, for each visit, specifies the time window when the visit should happen and a list of protocol activities (assessments, procedures and tests) that are planned at that visit. Activities such as medication need not be confined to visits and can be planned to occur in a time window within a protocol phase. An activity can have sub activities

that impose additional temporal constraints. For example, an assessment activity can include collection and processing of biological specimens with associated temporal constraints.

Here is a representative set of temporal constraints that we found in the ITN protocols that we are encoding:

1. Visit 17 must occur at least 1 week but no later than 4 weeks after the end of 2003 ragweed season.
2. Administer Rapamune 1 week from Visit 0 daily for 84 days.
3. Visit 1 should occur 2 weeks \pm 3 days after transplant.
4. Screening visit evaluations must occur between 30 days prior to Visit -1 and 45 days prior to Visit 0.
5. The vital signs of the participant should be obtained at routine time points starting at 10 minutes post infusion, then at 20-minute intervals until the participant is discharged.
6. Administer study medication at weekly intervals for 3 months.
7. Clinical assessments are required twice a week until Day 28 or discharge from hospital.
8. The first and second blood draws are 10 days apart, and the third draw is 11-14 days after the second.
9. On days that both IT and omalizumab are administered, omalizumab will be injected 60 minutes after the IT.
10. Monitor cyclosporine levels 3 times per week while in-patient, then weekly as out-patient.

As evident in the constraints, clinical activities—we are using the terms *activity* and *event* interchangeably—are temporally dependent on each other. The temporal annotations in the constraints are specified in relative terms typically with reference to one or more clinical events. At the protocol execution time, the actual times of these events found in the clinical data will be used to reason with the constraints. There can also be fuzziness in the relative start and end times as well as in the duration of the activity. An activity can be repeated at a periodic interval for a specific number of times or until a condition is satisfied. The periodic interval can be a single offset or a set of offsets. The temporal annotation of an activity or the temporal ordering of activities can be conditional on other events.

3 TEMPORAL REPRESENTATION

We have developed a temporal constraint ontology that can be used to formally specify the temporal constraints found in the clinical trial protocols. We briefly describe the core entities of the underlying temporal representation below:

Anchor defines an unbound time point that can be used to specify temporal relations among activities. It can be used as a reference point to define the start of another event before or after the anchor. In example 1 (of the constraints listed earlier), *end of 2003 ragweed season* is an anchor used to define the start of *Visit 17*. During the execution of the protocol, an anchor is bound to the absolute time of the anchor as recorded in the clinical trial data.

Duration is the difference between two time points. It is used typically to specify how long an activity lasts. In example 2, *84 days* is a duration.

Anchored Duration relates two activities with a temporal offset. In example 2, the activity *administer Rapamune* is offset from the anchor *Visit 0* by *1 week*.

Varying Duration is defined as duration with a high variance and a low variance. In example 3, *2 weeks ± 3 days* specifies a varying offset between *transplant* and *Visit 1*.

Start and End Expression constrains the start and the end of an activity and is expressed as offsets before or after one or more reference events. In example 4, the start of the activity *Screening visit evaluations* is 30 days before the anchor *Visit -1* and the end is 45 days before another anchor *Visit 0*.

Cyclical Plan Expression formulates events that are repeated at periodic intervals. The repetition ends typically when a specific number of cycles is reached or until a specific condition is satisfied. There are two types of cyclical plans with subtle differences. The first type has a single anchor point with potentially multiple intervals. In example 5, the *vital signs* assessments are planned at 10, 30, 60, 90, 120, and 180 minutes after *infusion*. If the participant gets off schedule because the assessment is made at minute 35 instead of minute 30, then the participant gets back on schedule with the next assessment at minute 60. This type of cyclical plan is used generally with assessments and tests where evaluations need to be made at specific intervals after a clinical intervention. The second type of cyclical plan can potentially have multiple anchors with a single offset. In example 6, the plan is to administer medication at weekly intervals for 3 months. The initial anchor is the event of administering the first dose. According to the

schedule, the second dose will be 1 week later, and the third 1 week later from the second dose. If the participant gets off schedule because the drug was administered 5 days after first dose and not 7 days, then the participant gets back on schedule with the next dose at 7 days from the last dose. This type of cyclical plan is used typically with drug administration where fixed intervals between dosages need to be maintained for safety and efficacy purposes.

Conditional Expression allows associating different temporal annotations with a single activity based on a condition. There are three patterns of conditional expressions – *if-then*, *if-then-else* and *until-then* patterns. Example 9 illustrates the *if-then* pattern – the temporal constraint between the administrations of two drugs is dependent on the condition that the two drugs are administered on the same day. Example 10 illustrates the *until-then* pattern – the monitoring activity is performed 3 times a week until the participant is in in-patient status, and when the status changes to out-patient then the activity is performed weekly.

4 EPOCH ONTOLOGIES

In order to support clinical trial management activities, the Epoch knowledge-based approach provides three methods: 1. knowledge acquisition methods that allow users to encode protocols, 2. ontology-database mapping methods that integrate the protocol and biomedical knowledge with clinical trial data including clinical results and operational data stored in the ITN data repository, and 3. concept-driven querying methods that support integrated data management, and that can be used to create high-level abstractions of clinical data during analysis of clinical results. At the center of all these methods is the suite of Epoch ontologies that provide a common nomenclature and semantics of clinical trial protocol elements. We shall describe each of the core ontologies.

4.1 Protocol Ontology

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 clinical trial management. Other concepts are either ignored or partially represented. The main concepts represented in the protocol ontology are the protocol schema and the schedule of activities.

4.2 Temporal Constraints Ontology

The *temporal constraints ontology* models the class of temporal constraints found in clinical trial protocols (see Section 3).

4.3 Virtual Trial Data Ontology

The *virtual trial data ontology* encapsulates the study data that is being collected, such as participant clinical record, specimen workflow logs, and site related data. A mapping component can then map clinical trial data (found in a relational database) to these virtual data records using a mapping ontology. The data model concept is similar to the Virtual Medical Record (Johnson et al., 2001) specification promoted in the clinical guideline modeling efforts.

4.4 Temporal Model

The *temporal model* provides a valid-time model of the temporal component of clinical trial data. In this model, all facts have temporal extent and are associated with instants or intervals denoting the times that they are held to be true. The core concept in the model is the *extended proposition* class that represents information that extends over time. There are two types of extended propositions in the model: 1. *extended primitive propositions* that represent data derived directly from secondary storage, and 2. *extended abstract propositions* that are abstracted from other propositions. These extended propositions can be used to consistently represent temporal information in ontologies. For example, a set of participant visits in a clinical trial data can be represented by defining a class called *VisitRecord* that inherits the *valid time* property from *extended proposition* class. The *valid time* property will then hold a visit's actual occurrence time. Similarly, an extended primitive proposition can be used to represent a drug regimen, with a value of type string to hold the drug name and a set of periods in the valid time property to hold drug delivery times. A more detailed discussion of the temporal model can be found elsewhere in the literature (O'Connor et al., 2006).

5 OWL IMPLEMENTATION

We have developed these ontologies in OWL 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 ontology should be interpreted. The Semantic Web Rule Language (SWRL) allows users to write Horn-like rules that can be expressed in terms of OWL concepts and that can reason about OWL individuals. SWRL provides deductive reasoning capabilities that can infer new knowledge from an existing OWL knowledge base. We use SWRL to specify temporal constraints. Once all temporal information is represented consistently using the temporal model, then SWRL rules can be written in terms of this model and the temporal constraint ontology. However, the core SWRL language has limited temporal reasoning capabilities. A few temporal predicates called *built-ins* are included in the set of standard predicates, but they have limited expressive power. SWRL provides an extension mechanism to add user-defined predicates. We used this mechanism to define a set of temporal predicates to operate on temporal values. These predicates support the standard Allen temporal operators (Allen, 1993). Using these built-in operators in conjunction with the temporal model, we can express complex temporal rules. Here is an example SWRL rule to check if participants conform to the visit schedule specified in the protocol:

```
Participant(?p) ^
hasVisitRecord(?p, ?vr) ^
hasVisitId(?vr, ?vid1) ^
hasValidTime(?vr, ?vt) ^
Visit(?v) ^
hasVisitId(?vr, ?vid2) ^
hasStartExpression(?vr, ?se) ^
swrlb:equal(?vid1, ?vid2) ^
temporal:inside(?vt, ?se) ^
-> ConformingParticipant(?p)
```

This rule uses concepts such as *Participant* and *Visit* from the *protocol ontology* and the concept of *Start Expression* from the *temporal constraint ontology*. The class of actual visits undertaken by a participant is the *VisitRecord* in the *virtual trial data ontology*, and is modeled as an extended proposition. The rule uses two built-ins – *equal*, that checks if two strings are equal, and *inside*, which is a built-in that we developed to check if an absolute time is within an anchored varying duration (see Section 3). Protégé (Knublauch, 2004) (<http://protege.stanford.edu/>) is a

software tool that supports the specification and maintenance of terminologies, ontologies and knowledge-bases in OWL. It has a plug-in called SWRL Tab (O'Connor et al., 2005), an editor for SWRL rules. We used Protégé to create the ontologies in OWL and SWRL. We then encoded specific protocols using Protégé's knowledge-acquisition facilities. The data generated from the implementation and execution of clinical trials is stored in a relational database. The types of data include participant enrollment data, specimen shipping and receiving logs, participant visits and activities, and clinical results. We have implemented a dynamic OWL-to-relational mapping method and have used SWRL to provide a high-level query language that uses this mapping methodology. A *schema ontology* describes the schema of an arbitrary relational database. A *mapping ontology* describes the mapping of data stored in tables in a relational database to entities in an OWL ontology. A *mapping software* uses the data source and mapping ontologies to dynamically map data to entities in the *clinical trial data ontology*. A detailed description of the mapping techniques can be found elsewhere in the literature (O'Connor et al., 2007). We are currently using JESS (<http://www.jessrules.com/>), a production rule-engine, to selectively execute the SWRL 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. Thus, a temporal constraint is defined first using the temporal constraint ontology, then is formulated as a rule, finally, is reasoned with real clinical data using dynamic mappings.

6 RELATED WORK

Over the years, many expressive models have been developed to represent temporal constraints (Bettini et al., 2002), (Combi, 2004), (Terenziani, 2002), (Duftschmid, 2002). Shahar's approach (Shahar, 1996) identifies temporal abstractions of data and properties using interpolation-based techniques and knowledge-based reasoning. In recent years, there have been a number of initiatives to create clinical trial protocol models that encapsulate clinical trial activities and associated temporal constraints found in a protocol. These ontologies are then used to automate different clinical trial management activities such as eligibility determination, participant tracking, and site management. The

ontologies can also be used when subsequently analyzing the clinical trial data.

In the past few years, we have seen considerable interest in building knowledge-based systems that automate clinical trial protocols and clinical practice guidelines. Our Epoch framework employs a task-based paradigm that combines an explicit representation of the clinical trial domain with rules that capture the logical conditions and temporal constraints found in the trial management process. There have been a number of proposals on task-based clinical guideline representation formats – EON (Musen et al., 1996), PROforma (Fox et al., 1996), GLIF (Boxwala et al., 2004), etc. that deal with temporal constraints on patient data and on activities found in clinical guidelines.

In the area of clinical trials, several modelling efforts have addressed different requirements of trial management activities. An ontology to represent temporal information and cyclical event patterns in clinical trial protocols has been proposed by (Weng et al., 2002). The Trial Bank Project (Sim et al., 2003) is a trial registry that uses a protocol ontology to capture information on randomized clinical trials such as intervention, outcomes, and eligibility criteria. The underlying knowledge base can support systematic reviewing and evidence-based practice.

There is an ongoing effort by CDISC (<http://www.cdisc.org/>), an industry-lead, multidisciplinary organization, to develop and support the electronic acquisition, exchange, submission and archiving of clinical trials data. As part of this effort, CDISC is developing the Structured Protocol Representation that identifies standard elements of a clinical trial protocol that can be codified to facilitate the data interchange among systems and stakeholders including regulatory authorities, biopharmaceutical industry, statisticians, project managers, etc. A parallel effort is the BRIDG (Weng et al., 2007) project, a partnership of several organizations including CDISC, the HL7 (<http://www.hl7.org/>) standards body, the National Cancer Institute and the Federal Drug Administration, that consumes the Trial Design Model work to build a comprehensive domain analysis model representing protocol-driven biomedical/clinical research. The BRIDG model is a work in progress to elaborately define functions and behaviors throughout clinical trials, and uses the Unified Modeling Language (UML) for representation. The model, in its current state, lacks formalization of and reasoning with temporal constraints, and thus, cannot fully support the requirements of ITN's clinical trial management.

7 DISCUSSION

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. 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 the Epoch framework is a suite of ontologies that encodes knowledge about the clinical trial domain that is relevant to trial management activities. This focus on just supporting trial management activities is also reflected in our approach to temporal constraint reasoning. Thus, in the temporal constraint ontology and in our reasoning approach with rules, we have limited ourselves to the types of temporal constraints, to the complexity of formalism and to the levels of reasoning to just support the clinical trial management activities. For example, we do not support checking temporal constraints for consistency. We continue to work on the *temporal constraints ontology* to support newer and more complex constraints. With any complex constraint, one concern is the power, or lack thereof, of our reasoning approach with SWRL rules,

Since we use OWL ontologies and SWRL rules, native RDF Store (storing data as RDF triples) would have been a natural solution for storing clinical trial data, and then seamlessly operate on the data using our ontologies and rules. ITN uses a legacy relational database system to store clinical trial data, and therefore, prevents us from using native RDF Stores as our backend. We have built techniques to map the database tables to our *virtual trial data ontology* OWL classes. With these solutions, our data model remains flexible and independent of the structure of the data sources. We are yet to undertake a thorough evaluation of our dynamic mapping methodology especially in the area of scalability

An often over-looked aspect of knowledge-based reasoning approaches is the task of knowledge-acquisition. Currently, we use the Protégé-OWL editor to build the Epoch models. Based on the class and property definitions, Protégé automatically generates graphical user interface (GUI) forms that can be used to create instances of these classes (OWL *individuals*). Thus, domain specialists can use to enter a specification of a protocol, say for a transplant clinical trial, using these Protégé-generated forms. Unfortunately, domain specialists

find it cumbersome and non-intuitive to use the generic user interfaces as they are exposed to the complexities of the Epoch ontologies, the OWL expressions and the SWRL rules. We are building custom graphical user interfaces that hide the complexities of the knowledge models, and that facilitate guided knowledge-acquisition. Providing a friendly user interface to enter SWRL rules can be challenging.

The knowledge requirements borne out of the need for managing clinical trials align well with the touted strengths of semantic web technologies – uniform domain-specific semantics, flexible information models, and inference technology. Using these technologies, we have built a knowledge-based framework for temporal constraints reasoning that is, above all, practical.

ACKNOWLEDGEMENTS

This work was supported in part by the Immune Tolerance Network, which is funded by the National Institutes of Health under Grant NO1-AI-15416.

REFERENCES

- Allen, J.F., 1993. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11): 832-843.
- Bettini, C., Jajodia, S., Wang, X., 2002. Solving multi-granularity constraint networks. *Artificial Intelligence*, 140(1-2):107-152.
- Boxwala, A.A., Peleg, M., Tu, S. W., Ogunyemi, O., Zeng, Q. T., Wang, D., Patel, V. L., Greenes, R. A., Shortliffe, E. H. 2004. GLIF3: A Representation Format for Sharable Computer-Interpretable Clinical Practice. *Journal of Biomedical Informatics*, 37(3):147-161.
- Combi, C., Franceschet, M., and Peron, A, 2004. Representing and Reasoning about Temporal Granularities. *Journal of Logic and Computation*, 14(1):51-77.
- Duftschnid, G., Miksch, S., Gall, W. 2002. Verification of temporal scheduling constraints in clinical practice guidelines. *Artificial Intelligence in Medicine* 25(2): 93-121.
- Fox, J., Johns, N., Rahmzadeh, A., Thomson, R. 1996. PROfarma: A method and language for specifying clinical guidelines and protocols. *Proceedings of Medical Informatics Europe*.
- Johnson, P.D., Tu, S. W., Musen, M. A., Purves, I., 2001. A Virtual Medical Record for Guideline-Based Decision Support. *Proceedings of the 2001 AMIA Annual Symposium*, 294-298.

- Knublauch, H. Fergerson, R.W., Noy, N.F. and Musen, M.A., 2004. The Protégé OWL Plugin: An Open Development Environment for Semantic Web applications. *Proceedings of the Third International Semantic Web Conference*. 229-243.
- Musen, M.A., Tu, S.W., Das, A.K., Shahar, Y. 1996. EON: A component-based approach to automation of protocol-directed therapy, *Journal of the American Medical Informatics Association*, 3(6): 367–388.
- O'Connor, M.J., Knublauch, H., Tu, S.W., Grossof, B., Dean, M., Grosso, W.E., Musen, M.A., 2005, Supporting Rule System Interoperability on the Semantic Web with SWRL. *Proceedings of the Fourth International Semantic Web Conference*. 974-986.
- O'Connor, M.J., Shankar, R.D. Das, A.K., 2006. An Ontology-Driven Mediator for Querying Time-Oriented Biomedical Data. *Proceedings of the 19th IEEE International Symposium on Computer-Based Medical Systems*. 264-269.
- O'Connor, M.J., Shankar, R.D. Tu, S.W., Nyulas, C., Musen, M.A., Das, A.K., 2007. Using Semantic Web Technologies for Knowledge-Driven Queries in Clinical Trials. *Proceedings of the 11th Conference on Artificial Intelligence in Medicine*.
- Rotrosen, D., Matthews, J.B., Bluestone, J.A., 2002. The Immune Tolerance Network: a New Paradigm for Developing Tolerance-Inducing Therapies. *Journal of Allergy and Clinical Immunology*, 110(1):17-23.
- Shahar Y., Musen, M.A.. 1996. Knowledge-Based Temporal Abstraction in Clinical Domains. *Artificial Intelligence in Medicine*, 8:267-298.
- Shankar, R.D., Martins, S.B., O'Connor, M.J., Parrish, D.B., Das, A.K. 2006. Epoch: an ontological framework to support clinical trials management. *Proceedings of the International Workshop on Health Information and Knowledge Management*, 25–32.
- Shankar, R.D., Martins, S.B., O'Connor, M.J., Parrish, D.B., Das, A.K., 2006. Towards Semantic Interoperability in a Clinical Trials Management System. *Proceedings of the Fifth International Semantic Web Conference*. 901-912.
- Sim, I., Olasov, B., Carini, S. 2003. The Trial Bank system: capturing randomized trials for evidence-based medicine. *Proceedings of the American Medical Informatics Association Fall Symposium*. 1076.
- Terenziani, P., 2002. Toward a Unifying Ontology Dealing with Both User-Defined Periodicity and Temporal Constraints About Repeated Events, *Computational Intelligence* 18(3):336-385.
- Weng, C., Kahn, M., Gennari, J.H. 2002. Temporal Knowledge Representation for Scheduling Tasks in Clinical Trial Protocols. *Proceedings of the American Medical Informatics Association Fall Symposium*. 879 – 883.
- Weng, C., Gennari, J.H., Fridsma, D.B. 2007. User-centered semantic harmonization; A case study. *Journal of Biomedical Informatics* 40: 353–364.