

CNTRO 2.0: A Harmonized Semantic Web Ontology for Temporal Relation Inferencing in Clinical Narratives

Cui Tao, PhD Harold R. Solbrig, BS Christopher G. Chute, MD, DrPH
{tao.cui, solbrig.harold, chute}@mayo.edu
Mayo Clinic College of Medicine, Rochester, MN

Abstract

The Clinical Narrative Temporal Relation Ontology (CNTRO) has been developed for the purpose of allowing temporal information of clinical data to be semantically annotated and queried, and using inference to expose new temporal features and relations based on the semantic assertions and definitions of the temporal aspects in the ontology. While CNTRO provides a formal semantic foundation to leverage the semantic-web techniques, it is still necessary to arrive at a shared set of semantics and operational rules with commonly used ontologies for the time domain. This paper introduces CNTRO 2.0, which tries to harmonize CNTRO 1.0 and a list of existing time ontologies or top-level ontologies into a unified model — an OWL based ontology of temporal relations for clinical research.

1 Introduction and Background

The rapid increase in the volume of electronic health records (EHR) available for research purposes provides new opportunities to create semantically interoperable healthcare applications and solutions for evidence-based medicine. An important aspect of EHR is the temporal ordering of clinical events. Time is essential in clinical research [1]. Exposing the temporal dimension in medical data analysis provides new research paths such as (1) uncovering temporal patterns at the disease and patient level to better understand the progression of a disease, (2) explaining past events such as the possible causes of a clinical situation, and (3) predicting future events such as possible complexities based on a patient’s current status.

One important objective for enable meaningful use of EHR is to develop software applications “to realize the true potential of EHR to improve the safety, quality, and efficiency of care” [2]. “EHRs must enable knowledge extraction and application” [3]. In order to facilitate clinical researchers to expose the temporal dimension in medical data analysis, software platforms that allow users to ask free-form queries and retrieve temporal information automatically from clinical records are highly desired. First, the temporal information interwoven in clinical narratives needs to be extracted and annotated to allow computer systems to be able to locate the information of interest. Second, temporal relations and assertions that are not explicitly expressed in the original documents need to be automatically inferred in order to enable the full capacity and true potential of secondary use of EHR for meaningful use. Third, temporal-oriented questions need to be captured in

computer queries to query the annotated and inferred information.

The Semantic Web and the Web Ontology Language (OWL) [4] provide a suitable environment for modeling the temporal dimension of the clinical data, reasoning and inferring new knowledge, and querying for the information desired. The Semantic Web provides a standard mechanism with explicit and formal semantic knowledge representation, and automated reasoning capabilities. OWL is built on formalisms that adhere to Description Logic (DL) and therefore allows reasoning and inference. The Semantic Web Rule Language (SWRL) [5] can be used to add rules to OWL and enable Horn-like rules that can be used to infer new knowledge from an OWL based ontology and reason about OWL individuals. Once we have an ontology that can represent temporal assertions in the clinical domain precisely, we can annotate temporal expressions and relations with respect to the ontology and store the instances as RDF triples [6]. The information then becomes “machine-understandable”. Tools and services such as reasoners, editors, querying systems, and storage mechanisms that have been developed by the Semantic Web community can be directly applied to the temporal data.

Toward this direction, we have developed an ontology in OWL for modeling temporal information in clinical narratives, and evaluated this ontology using real-world clinical notes [7]. We call this ontology the Clinical Narrative Temporal Relation Ontology (CNTRO)¹. CNTRO can model the temporal information found both in structured databases and in natural-language based clinical reports.

While CNTRO provides a formal semantic foundation for us to leverage the semantic-web techniques, it is still necessary to arrive at a shared set of semantics and operational rules with commonly-used ontologies for the time domain. During the development process of CNTRO, we have studied and compared a list of well-recognized OWL ontologies for time-related components and relations, such as the Time Ontology [8], the SWRL Temporal ontology [9], and the Basic Formal Ontology (BFO) [10]. These ontologies either focus on well-structured data in databases, model the temporal relations between temporal data instead of events, or target on top level meta definitions on the temporal aspects, therefore do not fit the requirements to model the temporal information in clinical narratives. We then decided to build CNTRO 1.0 as a stand along temporal ontology that fits the requirements of our

¹<http://www.cntro.org>

domain. CNTRO and these existing ontologies, however, do share some commonalities in different aspects. This paper builds on these previous threads along this line and attempts to harmonize them into a unified model – CNTRO 2.0 – an OWL based ontology of temporal relations for clinical research. The purpose of this ontology is to allow temporal information of clinical data be semantically annotated and queried, and to use inference to expose new temporal features and relations based on the semantic assertions and definitions of the temporal aspects in the ontology.

The rest of the paper is organized as follows. We introduce CNTRO 1.0 briefly in Section 2. Section 3 compares the time and event definitions in CNTRO and related ontologies, and discusses how we extended the event class using BFO. Section 4 describes how we harmonized temporal-relation definitions with the Time Ontology and the SWRL Temporal Ontology. Section 5 discusses the improvements on modeling granularity and units. And Section 6 summarizes the paper and provides conclusion remarks.

2 CNTRO 1.0

Figure 1 shows the graphical view of the CNTRO 1.0 introduced in [7]. OWL classes are represented by a rectangle with rounded corners and data types are represented by ovals. Subclass relationships are represented by hollow-headed arrows and object and data properties by solid-headed arrows with the arrow pointed to the ranges.

CNTRO centered by the class, *Event*, which is used to represent any occurrence, state, perception, procedure, symptom or situation that occurs on a time line in clinical narratives.

The *Time* class is the superclass of all the OWL temporal representation classes: *TimeInstant*, *TimeInterval*, *TimePhase*, and *TimePeriod*. The *TimeInstant* class is to represent a specific point of time on the time line. For each time instant, CNTRO can also represent its level of granularity, the original form, and the normalized form. The *TimeInterval* class represents a duration of time. It could have two relations (OWL object properties), *hasStartTime* and *hasEndTime*. Each of them links to *TimeInstant*. A *TimeInterval* could also have a *Duration*. An instance of the *Duration* class represents the time length of a *TimeInterval*. We use an OWL data type property *hasValue* and an OWL object property *hasUnit* to describe a *Duration*.

Many clinical events recur periodically. CNTRO has two OWL classes, *TimePhase* and *TimePeriod*, to represent intervals of time that recur periodically. A *TimePhase* represents each occurrence of the repeating interval and a *TimePeriod* specifies a reciprocal measure of the frequency at which the *TimePhase* repeats. The class *TimePhase* is a subclass of *TimeInterval*, therefore, we can also specify a *StartTime*, an *EndTime*, and a *Duration*. In addition, a relation (OWL ObjectProperty), *hasTimePeriod*, is defined to specify the relation between a *TimePhase* and a *TimePeriod*. For example, “every 8

hours for 10 days starting from today” is a *TimePhase*. Its *StartTime* is “today”. Its *Duration* is “10 days”. And its *TimePeriod* is “every 8 hours”.

We also define the certainty of a *Time* instance. For example, a physician can describe a time notation with ambiguities such as “early next week” and “in approximately two weeks”. In the CNTRO ontology, we defined a class called “Modality” which serves as a flag to indicate whether a time representation is approximated or not.

We can define the temporal relations between two events, or between an event and a time instance using the object property *hasTemporalRelation* and its subproperties. We use Allen’s temporal logic operators to define our temporal relation properties: equal, before, after, during, meet, start, finish, and during. We have also defined their logical characteristics. For example, *before* is a transitive property, and its inverse property is *after*. We can also use *TemporalRelationStatement* class to describe temporal relations between two events or between an event and a *Time* instance. The *TemporalRelationStatement* class is a subclass of *rdf:Statement*, we can define temporal subject, object, and predicate of a *TemporalRelationStatement*. Using *TemporalRelationStatement* to describe a temporal relation enables defining properties of the relation by reification. For example, we can add an offset time frame to the relation by using an OWL object property called *hasTemporalOffset*. The domain of *hasTemporalOffset* is *TemporalRelationStatement* and the range of it is *Duration*. This offset defines the relative timing of a pair of events. In order to model the sentence “patient’s bilirubin is elevated 2 weeks after the second cycle of chemotherapy”, for example, we can use a *TemporalRelationStatement* to represent “patient’s bilirubin is elevated” (object) *after* (predicate) “the second cycle of chemotherapy” (subject), and then add “2 week” as an instance of *TemporalOffset* to this *TemporalRelationStatement* instance.

3 Entity Definition

We first want to harmonize the time-related classes defined in CNTRO with the two well-acknowledged OWL ontologies for temporal information (the Time Ontology and the SWRL Temporal Ontology), as well as the Basic Formal Ontology. Table 1 shows the comparison of the time-related classes in the four ontologies. CNTRO defined a *Time* class which is the superclass of all the OWL temporal representation classes: *TimeInstant*, *TimeInterval*, *TimePhase*, and *TimePeriod*. The Time Ontology defined a class called *TemporalEntity* which has two subclasses, *Interval* and *Instant*. The SWRL Temporal Ontology defined a class called *ValidTime* which has two subclasses, *ValidInstant* and *ValidPeriod*. BFO has defined a class called *connected_temporal_region*, which has two subclasses: *temporal_instant*, and *temporal_interval*. We consider the classes for Instant among the four classes are equivalent classes. Same as the classes for Interval among the four classes. Since the *Time* class in CNTRO contains two additional subclasses, we did not map it with the cor-

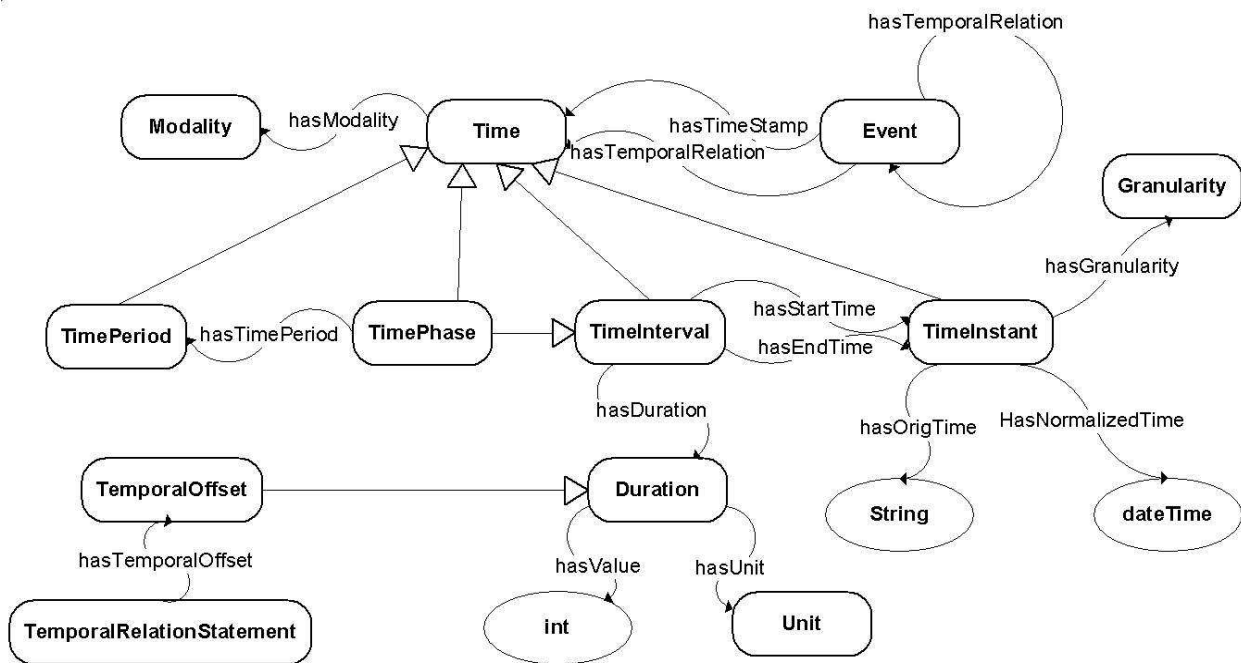


Figure 1: A Graphical View of CNTRO 1.0.

	CNTRO	TIME	SWRL Temporal	BFO
Overall Time Class	Time	TemporalEntity	ValidTime	connected_temporal_region
Instant	TimeInstant	Instant	ValidInstant	temporal_instant
Interval	TimeInterval	Interval	ValidPeriod	temporal_interval
Other	TimePhase, TimePeriod	-	-	-

Table 1: Comparison of the Time-related Classes

responding classes in Table 1 on the “Overall Time Class” level.

CNTRO 1.0 defined an *Event* class for describing any sort of occurrence, state, perception, procedure, symptom or situation that occurs on a time line in clinical narratives. CNTRO 2.0 expanded the definition of the *Event* class by introducing three new subclasses adopted from the BFO: *process*, *process_boundary*, and *fiat_process_part*. BFO defines a *process* as “a processual entity that is a maximally connected spatiotemporal whole and has bona fide beginnings and endings corresponding to real discontinuities”. We use *process* to annotate a clinical event that associates with a time interval. For example, in the sentence “patient’s last cycle of chemotherapy started from Jan. 10 and ended on Jan. 19”, the event “last cycle of chemotherapy” should be considered as a *process*. A *process_boundary* is defined as “a processual entity that is the fiat or bona fide instantaneous temporal process boundary”. We use this class to represent clinical events that associate with a time instant. For example, in the sentence “the patient checked in at 3pm”, the event “checked in” should be considered as a *process_boundary*. A *fiat_process_part* is defined in BFO as “a processual entity that is part of a process but that does not have begin-

nings and endings corresponding to real discontinuities.” We use this class to describe events with unknown or unspecified starting and ending time. For example, in the sentence “patient experienced headache during last week”, the event “patient experienced headache” should be considered as a *fiat_process_part*.

We also introduced two new subclasses of the *Event* class: *process_observable* and *process_repeat*. The *process_observable* class is to represent observable states, conditions, or qualities appeared in clinical narratives. BFO contains classes such as *quality* and *realizable_entity* that can be used to describe a continuant that could be changed over time, such as the blood pressure, the height, and the weight of a patient. What clinical narratives report, instead, are the results of the quantitative measurements for these continuant. We believe it is more reasonable to use *process_observable* to represent the observable results of the measuring or testing process. The *process_repeat* class is to represent the clinical events recur periodically. For example, in the sentence “take antibiotics every 8 hours for 10 days starting from today”, the event “take antibiotics” is a instance of *process_repeat*, with time stamp that can be represented by CNTRO *TimePhase* and *TimePeriod*.

4 Temporal Relation Harmonization

In this section, we describe how we harmonized the temporal relations defined in CNTRO with the Time Ontology and the SWRL Temporal Ontology. All the three ontologies adopted Allen’s temporal operators [11] to describe temporal relations. One important distinction between CNTRO temporal relations and Allen’s temporal operators, however, is that CNTRO temporal relations are defined for events, but Allen’s temporal operators are defined between time intervals. The SWRL Temporal Ontology extended Allen’s temporal operators so that they could be applied to both time intervals and time instants. The Time ontology in OWL, uses Allen’s temporal operators only for time intervals, and added a general *before* and a general *after* for relations between either instants or intervals.

CNTRO is an ontology describing the temporal relations of clinical events. Therefore the temporal relations defined in CNTRO are between either two *events*, or an *event* and a *time*. This definition also aligns with the Relation Ontology (RO) [12], which is a well-recognized ontology for fundamental relations [13]. In many cases in clinical narratives, physicians describe the relations between two events without indicating the time stamps of neither or one of the events. Here is an example: “patient’s bilirubin is elevated after the second cycle of chemotherapy”. In this case, it is not straightforward to use either the SWRL Temporal ontology or the Time ontology to capture the relation *after* since both the domain and the range of the *after* property in these two ontologies are defined as some sort of temporal expressions, but not events. Using CNTRO, however, we are able to represent this kind of relations as:

```
<event1> rdf:type Event;  
         rdfs:label "patient's bilirubin  
                 is elevated";  
<event2> rdf:type Event;  
         rdfs:label "the second cycle  
                 of chemotherapy";  
<event1> after <event2>
```

In CNTRO, the *Event* class connects to the time-associated classes by an object property called *hasTimeStamp*, therefore we can still capture or infer the temporal relations between two events based on their associated temporal information. Through the *hasTimeStamp* property, we can define the temporal relations in CNTRO using the temporal relations defined in either the SWRL Temporal ontology or the Time ontology, and therefore semantically harmonize them. For example, we can define the *after* property in CNTRO using the following SWRL rules where `<swrl:Builtin#after>` refers to the *after* operator defined in the SWRL Temporal Ontology.

```
Event(?e1), hasTimeStamp(?e1, ?t1),  
Event(?e2), hasTimeStamp(?e2, ?t2),  
<swrl:Builtin#after>( ?t1, ?t2)  
-> after(?e1, ?e2)
```

```
Event(?e1), hasTimeStamp(?e1, ?t1),  
Time(?t2),  
<swrl:Builtin#after>( ?t1, ?t2)  
-> after(?e1, ?e2)
```

Similarly, we can also define the temporal relations in CNTRO using the temporal relation properties defined in the Time Ontology too.

One advantage of defining the temporal relation properties in CNTRO using the SWRL Temporal Ontology is that we can leverage the SWRL Temporal Built-Ins library [14] to perform temporal operations. Augmented by the logical characteristics of the temporal relation properties defined by CNTRO, we can perform different kinds of reasoning to infer new time-related knowledge. Here we use a simple example to illustrate a case of temporal relation reasoning. We have already known that *event1* is after *event2*. If we further know that *event2* happened on 01/26/07, and *event3* happened on 01/24/07, we can therefore infer that *event2* happened after *event3* based on their time stamps using the temporal operator *after* defined in SWRL. Now we know that *event1* is after *event2*, and *event2* is after *event3*. Since we defined that *after* is a transitive property in OWL, we can further infer that *event1* is after *event3*. We have designed a framework for temporal relation inference using SWRL and Pellet-based reasoner [15]. Since the focus of this paper is the semantic Harmonization of CNTRO with related ontologies, we discuss the detailed information about temporal relation reasoning in a separate manuscript [16].

5 Granularity and Unit

In clinical narratives, physicians might describe time expressions in different levels of granularity. CNTRO has defined the *Granularity* class to represent the allowed levels of granularity in the time domain, i.e., year, month, day, hour, minute, second. The SWRL Temporal Ontology has also defined a class for granularity. We believe it is reasonable to map the CNTRO *Granularity* class to the SWRL *Granularity* class. The SWRL Temporal Built-Ins implemented standard Allen temporal operators, and allow temporal operations on different levels of granularity. For example, the operation “temporal:add(?t, “2009-11-01T00:00”, -10, temporal:Years)” returns “1999-11-01T00:00” to the variable ?t, since the specific granularity is “temporal:Years”. In our temporal relation inference framework [16], we adopted the temporal operations implemented by the SWRL Temporal Built-Ins. Therefore, it is important to formally define the relations between the CNTRO *Granularity* class and the SWRL *Granularity* class, plus the relationships between the allowed individuals of these two classes. After the mapping definition is done, we can apply the temporal operations defined in the SWRL Temporal Built-Ins to the information from clinical narratives represented with respect to CNTRO. For example, for the sentence “the patient was diagnosed with diabetes 10 years ago.” here is the RDF triple representation:


```

1. <event1> rdf:type bfo:fiat_process_part;
2.     rdfs:label "the patient
   was diagnosed with diabetes";
3.     hasTimeStamp <tInst1>;
4. <tInst1> rdf:type TimeInstant;
5.     hasOrigTime "10 years ago";
6.     hasGranularity "year";

```

We further know that the note date is “2009-11-01”, we can then using the operation “temporal:add(?t, “2009-11-01T00:00”, -10, temporal:Years)” to calculate the time of <event1>. Note that since the granularity here is “year”, the information on any finer granularity will be ignored. In this case, we can only infer that the <event1> happened in the year of 1999.

In clinical narratives, physicians could also describe the duration of an event in different time units, for example, 2 days, 2 hours, and etc. CNTRO has defined a *Unit* class and specified that a *Duration* could have a numeric value (*cntro:hasValue*) which is an integer, and a *Unit* associated with it (*cntro:hasUnit*). CNTRO 2.0 maps our definition for units to the Measurement Units Ontology (MUO) [17]. The MUO defined two foundational classes, *muo:UnitOfMeasurement* and *muo:QualityValue*. An OWL object property *muo:measuredIn* connects these two classes to describe the unit of measurement of a specific quality value. The numeric value of *muo:QualityValue* is represented using *rdf:value*. In CNTRO 2.0, we defined that *cntro:Duration* is a subclass of *muo:QualityValue* and *cntro:Unit* is a subclass of *muo:UnitOfMeasurement*. Therefore, we can use the *muo:measuredIn* and *rdf:value* directly to represent the value of a duration without defining the two properties: *cntro:hasUnit* and *cntro:hasValue*.

6 Conclusion Remark

In this paper, we introduce a new version of the Clinical Narrative Temporal Relation Ontology, CNTRO 2.0. We compared CNTRO 1.0 with a list of existing time ontologies such as the Time Ontology and the SWRL Temporal ontology, and top-level ontologies such as the Basic Formal Ontology and the Measurement Units Ontology. We mapped the corresponding overlapped components such as the classes for time instants and time intervals among these ontologies. We used temporal relations defined in the Time ontology and the SWRL Temporal ontology to define the temporal relations in CNTRO. We extended the CNTRO *Event* class based on BFO. We also linked MUO to CNTRO for representing time units. We believe CNTRO 2.0 can serve as a harmonized unified model for representing temporal relations for clinical research.

Acknowledgment This research is supported by the National Science Foundation under Grant #0937060 to the Computing Research Association for the CIFellows Project and the ONC Strategic Health IT Advanced Research (SHARP) award under Grant #90TR0002-01.

References

- [1] Shahar Y. Timing Is Everything: Temporal Reasoning and Temporal Data Maintenance in Medicine. In: Proceedings

- of Artificial Intelligence in Medicine. Joint European Conference on Artificial Intelligence in Medicine and Medical Decision Making (AIMDM'99). Aalborg Denmark; 1999. p. 30–46.
- [2] Blumenthal D, Tavenner M. The “Meaningful Use” Regulation for Electronic Health Records. *The New England Journal of Medicine (NEJM)*. 2010 July;363(6):501–504.
- [3] Ahmad A. Electronic Health Records Beyond Meaningful Use. *ACM Transactions on Database Systems*. 2010 August;1(3):265–267.
- [4] OWL Web Ontology Language Reference;. <http://www.w3.org/TR/owl-ref/>.
- [5] A Semantic Web Rule Language Combining OWL and RuleML;. <http://www.w3.org/Submission/SWRL/>.
- [6] Resource Description Framework (RDF);. <http://www.w3.org/RDF/>.
- [7] Tao C, Wei WQ, Savova G, Chute CG. A Semantic Web Ontology for Temporal Relation Inferencing in Clinical Narratives. In: Proceedings of the American Medical Informatics Association (AMIA) 2010 Annual Symposium. Washington DC; 2010. .
- [8] Time Ontology in OWL;. <http://www.w3.org/TR/owl-time/>.
- [9] The SWRLTab’s Valid-Time Temporal Ontology;. <http://swrl.stanford.edu/ontologies/built-ins/3.3/temporal.owl>.
- [10] Basic Formal Ontology (BFO);. <http://www.ifomis.org/bfo>.
- [11] Allen JF. Maintaining knowledge about temporal intervals. *Communications of the ACM*. 1983;26(11):832–843.
- [12] Relation Ontology;. <http://www.obofoundry.org/ro/>.
- [13] Bittner T. Logical properties of foundational mereogeometrical relations in bio-ontologies. *Applied Ontology*. 2009;4(2):109–138.
- [14] SWRL Temporal Built-In Library;. <http://protege.cim3.net/cgi-bin/wiki.pl?SWRLTemporalBuiltIns>.
- [15] Pellet: OWL 2 Reasoner for Java;. <http://clarkparsia.com/pellet/>.
- [16] Tao C, Sobrig HR, Sharma DK, Wei WQ, Savova G, Chute CG. Time-Oriented Question Answering from Clinical Narratives Using Semantic-Web Techniques. In: the 9th International Semantic Web Conference (ISWC 10). Shanghai, China; 2010. .
- [17] Measurement Units Ontology;. http://forge.morfeo-project.org/wiki_en/index.php/Units_of_measurement_ontology.