

LODE: Linking Open Descriptions of Events

Ryan Shaw¹, Raphaël Troncy^{2,3} and Lynda Hardman³

¹ University of California, Berkeley, USA, <ryanshaw@ischool.berkeley.edu>

² EURECOM, Sophia Antipolis, France, <raphael.troncy@eurecom.fr>

³ CWI, Amsterdam, The Netherlands, <lynda.hardman@cwi.nl>

Abstract. People conventionally refer to an action or occurrence taking place at a certain time at a specific location as an *event*. This notion is potentially useful for connecting individual facts recorded in the rapidly growing collection of linked data sets and for discovering more complex relationships between data. In this paper, we provide an overview and comparison of existing event models, looking at the different choices they make of how to represent events. We describe a model for publishing records of events as Linked Data. We present tools for populating this model and a prototype “event directory” web service, which can be used to locate stable URIs for events that have occurred, provide RDFS+OWL descriptions and link to related resources.

1 Introduction

Though their specific methods differ significantly, both historians and journalists work to produce narrative chains of events to explain phenomena in the past. The resulting historical records of events constitute valuable cultural heritage of interest to academics as well as the general public. The Linked Data⁴ effort seeks to publish and connect RDF data sets on the Web using dereferenceable URIs for identifying web documents, real-world objects, links between them and/or other pieces of information. Yet, while standard and widely used vocabularies have emerged for representing people, places, and other types of entities as Linked Data, none has yet emerged specifically for events.

The term “event” has several meanings. It is used to mean both phenomena that have happened (e.g. things reported in news articles or explained by historians) and phenomena that are scheduled to happen (e.g. things put in calendars and datebooks). Various standards and formats have been proposed for representing the latter as structured data, usually for personal information management purposes. In this paper, we focus on the former category: phenomena that have happened in the past.

This paper makes two contributions. First, we compare existing models for representing historical events (Section 2). These models serve different communities and have different strengths. Our goal is not to propose yet another ontology *per se*, but rather to build an *interlingua* model that solves an interoperability

⁴ <http://linkeddata.org/>

problem by providing a set of axioms expressing mappings between existing event ontologies (Section 3). Second, we present tools for populating this model with data coming from existing sources, such as Wikipedia timelines. We describe a prototype of an “event directory”⁵ web service which can be used to locate stable URIs for past events and to provide RDFS+OWL descriptions of those events and links to related resources (Section 4). Finally, we give our conclusions and outline future work in Section 5.

2 Comparison of Existing Event Models

A number of different RDFS+OWL ontologies providing classes and properties for modeling events and their relationships have been proposed (see Table 1). In

Event model	Ontology URL
CIDOC CRM	http://cidoc.ics.forth.gr/OWL/cidoc_v4.2.owl
ABC Ontology	http://metadata.net/harmony/ABC/ABC.owl
Event Ontology	http://purl.org/NET/c4dm/event.owl#
EventsML-G2	http://www.iptc.org/EventsML/
DOLCE+DnS Ultralite	http://www.loa-cnr.it/ontologies/DUL.owl
F	http://events.semantic-multimedia.org/ontology/2008/12/15/model.owl
OpenCYC Ontology	http://www.opencyc.org/

Table 1. Ontologies for representing events

this section, we present an analysis based on their main constituent properties: type (Section 2.2), time (Section 2.3), space (Section 2.4), participation (Section 2.5), causality (Section 2.6) and composition (Section 2.7). This builds upon previous work in which we examined a number of different non-RDFS+OWL models for representing information about events [9].

2.1 Event Models Overview

Though all of the ontologies presented in Table 1 provide classes and properties suitable for representing events, they were created to serve different purposes. The CIDOC CRM [2] and ABC [6] ontologies aim at enabling interoperability among metadata standards for describing complex multimedia objects held by museums and libraries. The events they intend to describe include both historical events in the broad sense (e.g. wars, or births) as well as events in the histories of the objects being described (e.g. changes of ownership, or restoration).

The Event Ontology (EO) [7] was developed by the Centre for Digital Music to be used in conjunction with music-related ontologies. Although intended to describe events such as performances or sound generation, there is nothing specific to the music domain. It is currently the most commonly used event ontology in the Linked Data community. EventsML-G2 has been developed by the International Press Telecommunications Council (IPTC) for exchanging structured

⁵ We provide an interface for searching and browsing linked descriptions of events at <http://www.linkedevents.org>

information about events among news providers and their partners. It describes both planned, past or breaking events as reported in the news.

DOLCE+DnS Ultralite (DUL) is a lightweight “upper” ontology for grounding domain-specific ontologies in a set of well-analyzed basic concepts. It is a combination and simplification of the DOLCE foundational ontology and the Constructive Descriptions and Situations pattern for representing aspects of social reality [3]. The F Event Model is a formal model of events built on top of DUL. It provides additional properties and classes for modeling participation in events, as well as parthood relations, causal relations, and correlations between events. F also provides the ability to assert that multiple models represent views upon or interpretations of the same event [8]. OpenCYC is also an “upper” ontology, but at the other end of the spectrum from DUL: rather than being a lightweight set of core concepts it provides hundreds of thousands of concepts intended to model “all of human consensus reality”.

2.2 Fundamental Types of Events: Aspect and Agentivity

Given their different intended applications, these ontologies define events in varying ways. Table 2 provides a comparison of the prose descriptions for the top-level event classes. Furthermore, all of these ontologies, with the exception of EO,

<code>cidoc:E2.Temporal-Entity</code>	“[E2.Temporal_Entity] comprises all phenomena, such as the instances of E4.Periods, E5.Events and states, which happen over a limited extent in time.”
<code>abc:Event</code>	“An Event marks a transition between Situations.”
<code>eo:Event</code>	“An arbitrary classification of a space/time region, by a cognitive agent.”
<code>eventsml:Event</code>	“...something that happens and is subject to news coverage.”
<code>dul:Event</code>	“Any physical, social, or mental process, event, or state.”
<code>f:Event</code>	“...perduring entities (or perdurants or occurants) that unfold over time, <i>i.e.</i> , they take up time..”
<code>cyc:Situation</code>	“...a state or event consisting of one or more objects having certain properties or bearing certain relations to each other.”

Table 2. Definitions of top-level event-related classes

make an attempt to distinguish among some fundamental types of events. The basis upon which these distinctions are made vary.

One way to distinguish types of events is their *aspect*, *i.e.* whether the event involved is an ongoing activity or process, or the completion of some activity or transition between states. For example, OpenCYC defines a concept called **Situation** and uses *aspect* to distinguish between two main specializations of this concept: **StaticSituation** and **Event**. The former denotes a situation in which some state of affairs has persisted throughout the situation’s interval of time, while the latter denotes a situation in which some change has occurred during the situation’s interval of time.

CIDOC makes a similar but conceptually less clear distinction between two types of **E2.Temporal_Entity**: **E3.Condition_State** and **E5.Event**. It is less clear because CIDOC also introduces the concept **E4.Period**, a type of temporal entity that is not static, but does not necessarily involve a change of state. **E3.Condition_State** is also defined narrowly to denote only descriptions of “the prevailing physical condition of any material object or feature” which would seem to exclude descriptions of, for example, the relative state of two things. **E3.Condition_State** is similar to the ABC ontology’s **Situation** concept, instances of which describe the states of tangible things at particular times. The ABC ontology then uses this **Situation** concept to narrowly define an **Event** concept as a transition between two different **Situation** instances. This makes it difficult to describe an event that is characterized by a change in the relationship between two things rather than a change in the state of a single object.

Another distinction is whether an *agent* is identified as having produced the event. Both OpenCyc and DUL distinguish an **Action** as a particular type of **Event**, and CIDOC distinguishes an **E7.Activity** as a particular type of **E5.Event**. The ABC ontology also distinguishes an **Action** concept as something performed by an agent, but rather than being a specialization of the **Event** concept, it is defined as disjoint with the **Event** concept, which can “contain” actions via a **hasAction** property. Thus the ABC ontology suggests that events are fully described as sets of actions taken by specific agents, which may be an issue for modeling events such as earthquakes.

One potential problem with building these types of classifications into an ontology for modeling things that happened is that they force a knowledge engineer to adopt a particular perspective on what happened. This is desirable for precise modeling in specific domains that share a descriptive paradigm, but it is undesirable if the goal is to enhance access to documents which may present different interpretations of the same events. Distinctions based on aspect or agentivity are not necessarily inherent to what happened, but instead are rooted in particular interpretations. Whether a historical event or a event reported in the news involves an identifiable change or not, or whether agency can be assigned, is often a matter of debate, and its resolution should not be a prerequisite for representing what happened using a concept from an ontology.

This desire to separate events from their interpretations is what drives the approach taken by DUL, which provides a **Situation** concept, instances of which may describe different views or interpretations of the same **Event** instance. Using the DUL ontology, the types of classifications discussed above would be applied to instances of **Situation** rather than to instances of **Event**⁶.

2.3 Events and Temporal Intervals

Temporality is a major distinguishing feature of events as entities, requiring modeling spans of time and relating events to these. The relationship between

⁶ DUL does specialize its **Event** concept on the basis of agentivity, providing the **Action** concept for events that have at least one participating agent and the **Process** concept for events that are not recognized having participating agents.

events and chronological spans of time is analogous to the relationship between places and spatial coordinate systems. In each case, instances of the former have persistent, socially attributed meanings, while the latter are arbitrary systems for subdividing an abstract space. One approach to linking events to ranges of time uses datatype properties, directly relating event instances with RDF literals representing calendar dates (and thus typed using one of the date-related XML Schema datatypes such as `xsd:date` or `xsd:dateTime`). Another approach introduces a class for representing temporal intervals, and uses object properties to link event instances with instances of this class. Temporal interval instances can then be linked to calendar values using datatype properties.

ABC, CIDOC, and EO all take the second approach, with ABC and CIDOC introducing classes for temporal intervals, and EO using the `TemporalEntity` class from OWL-Time [5]. DUL allows both approaches: dates for an event can be directly asserted using the `hasEventDate` datatype property, or the temporal interval involved can be made explicit by instantiating the `TimeInterval` class and linking an event instance to it using the `isObservableAt` object property.

The advantage of associating dates directly with events is simplicity: there are fewer abstractions to deal with, and it is simple to filter or sort events using standard date parsing and comparison routines. This also makes it simple to export lists of events for visualization on a timeline. But the tradeoff for this simplicity is an inability to express more complex relationships to time, such as temporal intervals that do not coincide with date units, or uncertainty about when precisely an event took place within some bounded temporal interval. This is a problem for representing historical events.

By introducing classes for representing temporal intervals, one can use a temporal calculus for reasoning about these more complex relationships. For example, if the precise date of a historical event is not known but some boundaries can be established within which it must have occurred, the time between these boundaries can be represented as a temporal interval, and a containment relationship can be asserted between that interval and the (unknown) interval during which the event occurred. The drawback to such an approach is that it can be off-puttingly complex as it introduces a number of abstract entities. The problem also arises of how to either mint URIs to identify these entities or deal with the problems introduced by using blank nodes.

2.4 Events, Spaces and Places

Events can be linked to abstract temporal regions (Section 2.3) and to abstract spatial regions or to semantically significant places. ABC, CIDOC and EO only support linking to spatial regions. CIDOC provides a class (`E53.Place`) for “extent in space” to which events can be related via the `P7.took_place_at` property. Instances of `E53.Place` may have names (`E44.Place_Appellation`), but there is no way to link an event to a place name except through a specific spatial extent. ABC’s `Place` class also emphasizes spatial location rather than meaningful place. EO’s place property has a range of `wgs84:SpatialThing`, which is also defined in terms of spatial extent.

Only DUL makes an explicit place/space distinction between **Place** and **SpaceRegion**. An event instance can be related to a **Place** via the **hasLocation** property, or related to a **SpaceRegion** via the **hasRegion** property. This is the most flexible approach, as it allows one to make assertions about events that occurred in places not easily resolvable to geospatial coordinate systems. For example, scholars of ancient history may work with documents that do not distinguish between real and mythical events. These scholars may wish to indicate that some event is recorded as having occurred at a mythical place. Similar problems are posed by contemporary events which may occur at virtual places such as those found within massive multi-player online environments. In both cases it is convenient to be able to associate events to such places without having to specify geospatial coordinates for them. Furthermore, making a clear distinction between named places and spatial regions enables one to deal properly with the phenomenon of places changing their absolute spatial location over time.

2.5 Participation in Events

The event ontologies also provide properties for linking agents, such as people and organizations, and the things involved in them.

Object Involvement in Events ABC defines two types of properties for relating an **Event** to a tangible thing (an **Actuality** in ABC parlance). The **involves** property does not imply anything beyond simple involvement. The **hasResult** property relates an **Event** to a tangible thing or attribute of a thing which exists as a result of that **Event**. ABC also defines various sub-properties of these two properties that further specialize these meanings. For example **destroys** is a specialization of **involves** implying that the involved **Actuality** ceased to exist as a result of its involvement in the **Event**.

CIDOC defines a property **P12.occurred_in_the_presence_of**, which like ABC's **involves** relates an **E5.Event** to a **E77.Persistent_Item** (endurant) without committing to any implied role for that item beyond simple involvement. **P12.occurred_in_the_presence_of** is the root of a hierarchy of properties expressing more specialized forms of involvement such as **P25.moved** and **P31.has_modified**. Unlike ABC's **Actuality**, CIDOC's **E77.Persistent_Item** encompasses not only tangible entities but also intangible concepts or ideas, making CIDOC's **P12.occurred_in_the_presence_of** a broader concept than ABC's **involves**. DUL defines a **hasParticipant** for relating an **Event** to an **Object**. Like CIDOC's **E77.Persistent_Item**, DUL's **Object** includes social and mental objects as well as physical ones. EO's **factor** property, having no range defined, is similarly broad. EO also defines a **product** property that, like ABC's **hasResult**, links an **Event** to some thing that exists as a result of that **Event**.

Agent Participation in Events ABC defines a **hasPresence** property for weakly asserting that an agent was present at an event without implying that

the agent took an active role. It is specialized by the `hasParticipant` property, which does imply an active or causal role for the agent. CIDOC's equivalent of ABC's `hasPresence` is `P11.had_participant`, and its equivalent of ABC's `hasParticipant` is `P14.carried_out_by`. DUL's `involvesAgent` property is a specialization of `hasParticipant` for relating an `Event` to an `Agent`. EO provides the `agent` property for the same purpose.

F stands apart from the other ontologies in what it offers for modeling participation. Using DUL, one can assert that a given object or agent participated in an event. F uses the *descriptions and situations* (DnS) pattern[3] to enable a further classification of this participation as an instance of some role-based class. For example, using DUL one might state that the agents Brian Boru and Máel Mórda mac Murchada participated in the Battle of Clontarf. Using F, one can further state that the Battle of Clontarf is classified as a *battle*, that *battles* have *commanders*, and that Brian and Máel Mórda are classified as *commanders*.

CIDOC's `P14.1.in.the.role.of` property provides some support for classifying an agent's participation in an event as an instantiation of a particular role. However, since it is defined as a property of the `P14.carried_out_by` property, it requires the use of OWL Full. Furthermore, there does not seem to be a way to associate roles with generic event schemas in the manner described above.

2.6 Events, Influence, Purpose and Causality

Event models vary in their approaches to modeling relations of causality, purpose, or influence. Both EO and CIDOC provide properties for making broad assertions linking events to any relevant thing (tangible or not). CIDOC defines `P15.was_influenced_by`, while EO defines `factor`. EO does not distinguish between a thing's participation in an event and a thing's influence upon an event, using the same property for both relations. Likewise, it seems that the only difference between CIDOC's `P12.occurred_in.the.presence_of` and `P15.was_influenced_by` is whether the relevant thing was physically present (and, by implication, a `E77.Persistent_Item`). The only support that ABC offers for making assertions about causality is the `hasResult` property.

In historical discourse there is often a lack of consensus about causality, purpose, or influence. Thus simple properties like these are unlikely to be adequate for modeling assertions about such relations. Here the F model's DnS pattern provides a more powerful and flexible modeling tool. Unlike the other models, F takes the position that only other events can stand in causal relation to an event. Rather than directly linking events via a property expressing causality, events are included in an `EventCausalitySituation`. The `EventCausalitySituation` includes not only the events being classified as the cause and the effect, but also the theory under which causality is being asserted. Using the F model's interpretation pattern, one can assert that a given `EventCausalitySituation` is part of a specific interpretation of an event. Thus multiple, potentially conflicting causality relations can be asserted for the same set of events by specifying the interpretive context in which the relations are made.

2.7 Events, Parts and Composition

Often, it is desirable to model an event A as being part of some other event B . While an event A 's being part of event B implies that event B 's timespan contains event A 's timespan, event parthood is more than temporal containment. One may get married during the Olympics, but that does not make one's marriage part of the Olympics. Thus, event ontologies must distinguish between mere temporal containment and mereological relationships between sub-events and some greater event. Ontologies that make a distinction between temporal spans and events can clearly distinguish between the two types of relationships.

CIDOC distinguishes between time-spans and periods/events, and provides the `P86.falls_within` property to express containment relations among time-spans, and the `P9.consists_of` property to express part-of relationships among events. EO defines a `sub_event` property, and ABC defines an `isSubEventOf` property for expressing mereological relationships among events. Since ABC conceptualizes events as sets of actions taken by specific agents, it also provides the `hasAction` property for linking events to the actions they contain.

DUL defines two properties for linking events to sub-events: `hasPart` and `hasConstituent`. `hasPart` can be used both for temporal containment relationships such as “the 20th century contains year 1923” and for semantic relationships such as “World War II included Pearl Harbour”. `dul:hasConstituent` attempts to capture the notion that we sometimes model aspects of the world as consisting of layers at different levels of abstraction, which are not strictly parts of one another. Thus society is constituted of individual people, even though you might not want to say that people are “parts” of society because people and societies exist at different levels of abstraction. This distinction is useful for events as well, as it allows us to describe a large and complex event like the French Revolution as being constituted of many smaller events, even though these smaller events may not be “parts” of the larger event in the same sense that a set is part of a tennis match.

In keeping with its use of the DnS pattern, F enables one to define a high-level description of how an event can be composed of smaller events. Specific situations (i.e. specific groups of events) can then satisfy this description. This allows one to simply describe the conditions under which an event is considered to be part of another event, and infer parthood based on this description, rather than requiring parthood to be explicitly asserted every time. For large events that may contain large numbers of sub-events, this could be quite useful. And, of course, F's interpretation pattern allows for multiple, potentially conflicting decompositions of the same event.

3 Towards a Linked Data Event Model

We propose a minimal model that encapsulates the most useful properties of the models reviewed. Our goal is to enable interoperable modeling of the “factual” aspects of events, where these can be characterized in terms of the *four Ws*:

What happened, Where did it happen, When did it happen, and Who was involved. “Factual” relations within and among events are intended to represent intersubjective “consensus reality” and thus are not necessarily associated with a particular perspective or interpretation. Our model thus allows us to express characteristics about which a stable consensus has been reached, whether these are considered to be empirically given or rhetorically produced will depend on one’s epistemological stance. We exclude properties for categorizing events or for relating them to other events through parthood or causal relations. We believe that these aspects belong to an interpretive dimension best handled through the DnS approach of the F event model.

Table 3 shows the main properties of our model, aligned with approximately equivalent properties from the models discussed above. For the actual equivalence relations, see the ontology itself at <http://linkedevents.org/model/>.

ABC	CIDOC	DUL	EO	LODE
atTime	P4.has_time-span P7.took_place_at	isObservableAt	time place	atTime inSpace atPlace
inPlace		hasLocation		
involves	P12.occurred_in_the_presence_of	hasParticipant	factor	involved
hasPresence	P11.had_participant	involvesAgent	agent	involvedAgent

Table 3. Excerpt of approximate mappings between properties from various event models

Agentivity. Our model is agnostic with regard to judgements of aspect or agentivity (see Section 2.2). Users are free to model historical or reported events without taking a position on what has changed or where agency lies. This agnosticism has consequences for mapping our `Event` class to those defined by other models. We consider our `Event` class to be directly equivalent to those defined by EO and DUL, as both of these are also agnostic with respect to aspect and agentivity. Our event class is not equivalent to the `E5.Event` class, since CIDOC defines `E5.Event` to exclude ongoing states, activities, or processes. Because we wish to support the modeling of such static entities as events, we define our `Event` class to be a subclass of CIDOC’s `E2.TemporalEntity`, which is the superclass of `E5.Event` (via `E4.Period`) and `E3.Condition.State`. Our `Event` class is a subclass of `E2.TemporalEntity` because the latter is defined as “anything that happens over a limited extent in time”, which is more general than the definition we wish to give. Specifically, we want to restrict our definition to only include those things happening over a limited extent in time that have been reported as events by some agent, e.g. a historian or journalist.

Time. We link events to ranges of time via instances of a temporal interval class. Like EO, we use `TemporalEntity` from OWL-Time as our temporal interval class, so our `atTime` property is directly equivalent to EO’s `time` property.

atTime is a subclass of DUL's **isObservableAt** property, as it restricts the domain of the latter to include only events. Likewise, **atTime** is a sub-property of CIDOC's **P4.has_time-span** because it restricts the domain of the latter to include only events (as we define them here) rather than any temporal entity (recall that our event class is a subclass of CIDOC's **E2.TemporalEntity**). We also define **atTime** to be an OWL **FunctionalProperty**, meaning that an event can be associated with at most one interval of time. Where there may be disagreement about the interval of time associated with an event, this disagreement should be modeled at an interpretive level beyond the scope of our model, and the value of **atTime** should either be specified as the shortest temporal interval that includes the conflicting interpretations, or left unspecified.

Space. We follow DUL in making an explicit distinction between abstract spatial regions and semantically significant places. Our **inSpace** property relates an event to some subjectively imposed spatial boundaries, i.e. a region of space. Like **atTime**, **inSpace** is a **FunctionalProperty**, so an event can be related to at most one such region of space. **inSpace** is a sub-property of DUL's **hasRegion** because it restricts its domain to include only events, not all entities, and because it restricts its range to include only spatial regions, not any dimensional space. In keeping with EO, we use **SpatialThing** from the Basic Geo (WGS84 lat/long) Vocabulary as our spatial region class, so our **inSpace** property is directly equivalent to EO's **place** property. Because our concept of an event is broader than the one defined by the CIDOC CRM, **inSpace** is a super-property of CIDOC's **P7.took_place_at**. While the range of **inSpace** is an abstract spatial extent, it is often desirable to express relationships to socially defined places. We define an **atPlace** property to associate an event with some meaningful place(s), whether or not it is possible to define spatial boundaries for those places. Unlike **inSpace**, **atPlace** is not a **FunctionalProperty**, so an event can be related to any number of places. **atPlace** is a sub-property of DUL's **hasLocation** property, because it restricts the latter such that the domain includes only events and the range includes only places (not any entity).

Participation. Like DUL, we define a property for linking events to arbitrary things (**involved**) and a single specialization of this property for linking events to agents (**involvedAgent**). These two properties are directly equivalent to DUL's **hasParticipant** and **involvesAgent**, respectively. They are roughly equivalent to CIDOC's **P12.occurred_in_the_presence_of** and **P11.had_participant** (though not directly equivalent given our broader event concept). The mapping to EO is more complicated. **involved** is more specific than EO's **factor** property because it restricts the range of the latter to include only objects and not, for example, "abstract causes." But it is also more general, because it does not imply (as **factor** does) a "passive" role for the involved object. Thus there is no formal equivalence relationship stated between the two. **involvedAgent** is a super-property of EO's **agent** property because it generalizes the latter to include all relations to agents, whether or not their role is "active" or "passive."

Judgments of activity or passivity are higher-level interpretations that go beyond our goal of modeling only “factual” aspects.

Causality. Finally, as discussed above, our model contains no properties for expressing relations of influence, purpose, or causality. Therefore, there are no properties equivalent to CIDOC’s `P15.was_influenced_by` or EO’s `factor`. Similarly, we provide no properties for expressing parthood relations among events. We believe these higher-level interpretations are best handled via a layer of descriptions and situations over the basic statements expressible using our model. The F event model provides an exemplary blueprint.

4 Applications

For demonstrating the usefulness of our proposed model, we set up two experiments. First, we extract events from Wikipedia timelines in order to test whether we can represent these events accurately in the Web of Data (Section 4.1). Second, we load existing instances of events represented according to the various event models reviewed in this paper in order to test the interoperability we claim our model brings (Section 4.2). We provide an interface for searching, browsing and visualizing all these events at <http://www.linkedevents.org>.

4.1 Extracting Events from Wikipedia Timelines

The events found in Wikipedia timelines vary widely in scope and domain, making them a good challenge for modeling. We also demonstrate that Wikipedia timelines provide a source of structured data not yet tapped by projects such as DBpedia⁷ and Freebase⁸. Since timelines on related topics are spread throughout Wikipedia, extracting their events and modeling them as linked data is useful for enabling aggregated views of these events and for exploring related topics.

Timelines appear in Wikipedia in two major forms. *Dedicated topic-specific* timeline articles, such as “Timeline of historic inventions”, take the form of a list or table of events. As of October 2008, there were approximately 1000 such articles in Wikipedia. The list or table of events is usually divided into temporal groups (e.g. *September 1939* or *12th century*) by subheadings. Each event consists of (at a minimum) a date and a short description. The description generally contains words or phrases linked to other articles in the typical Wikipedia manner. The second form of timeline found in Wikipedia is *date-specific* timeline articles, such as “1996 in Ireland”. In addition to short lists of events in the form described above, these articles usually also include some type-specific lists of events such as births, deaths, and sporting events that took place in that year. The most general form of this type of article is the “Year” article (e.g. “1979”). Uses of a given year in any Wikipedia article are usually linked to the

⁷ <http://dbpedia.org/>

⁸ <http://freebase.com/>

corresponding “Year” article. Similarly, uses of a given day of the month (e.g. “May 24”) are usually linked to the corresponding “Month Day” article. These two types of article are highly mutually interlinked.

Date-specific timeline articles have a more standard format, making them more amenable to the extraction of structured data. But the events in date-specific timelines rarely have anything in common other than the year or day of the month with which they are associated. Since we were interested in linking events to one another via places, people, and other topics, we decided to focus on topic-specific timeline articles. Unfortunately, the formats for topic-specific timeline articles vary widely, making it difficult to create a generic parser and scraper. Many topic-specific timelines add additional fields for each event. For example, the “Timeline of Chinese history” includes a field for ruler or Emperor as well as the standard date and description. Other timelines group events in idiosyncratic ways, such as the “Timeline of punk rock” which categorizes the events of each year into “Bands formed”, “Disbandments”, “Albums [released]”, and “Singles [released]”. Furthermore, the timelines vary in the temporal granularity of their events: while some timelines specify specific days for their events, others only specify months or years. These variations illustrate how the structure of events can vary according to the topical context and the need for a flexible data model to accommodate them.

To populate instances of our event model, we wrote article-specific parsers for a number of the most active timeline articles. The parsers identify individual event entries within articles and from each entry extract the date and textual description. The parsers also extract the article subheading under which each entry appears for two reasons. First of all, the date specified in an entry is often given relative to the subheading. For example, events listed under the subheading *September 1939* may only specify a day of the month, with the month and year left implicit. Second, the subheadings provide a convenient means of linking back to the specific article section from which the event was extracted.

After the article-specific extraction, we use the extracted dates and descriptions to model our events. Dates are modeled using OWL-Time and linked to the event using the `atTime` property. Links to other Wikipedia articles found within the descriptions are used to identify other entities related to the event. We use type ontologies from DBpedia to determine what type of relation to create between an event and another entity. For example, if an event has the description “Canada declares war on Germany” and the word “Canada” is linked to the Wikipedia article of the same name, we then look up the corresponding resource in DBpedia (<http://dbpedia.org/resource/Canada>) and see what types have been assigned to it. <http://dbpedia.org/resource/Canada> has the type <http://dbpedia.org/ontology/Place> assigned to it, so we relate it to our event with the `atPlace` property. If DBpedia does not assign any usable types to the entity, we default to creating an `involves` relation.

Our initial set of events were extracted from four Wikipedia timelines:

- “Timeline of World War II” provides seven year-specific timelines of global events involving people at the granularity of single days.

- “Timeline of Irish History” provides events from a single geographic location spread over a wide temporal range, from the Stone Age to present day.
- “Timeline for the day of the September 11 attacks” provides a set of 147 very fine-grained events from a single day.
- “Timeline of evolution” tested our ability to model very coarse-grained events associated with times far in the past.

4.2 Interoperability with Legacy Event Collections

To evaluate the mappings between our model and other vocabularies, we combined our Wikipedia events with two collections of events modeled using other event vocabularies: the C4DM Event Ontology and the BIO⁹ vocabulary for biographical information. The goal was to be able to browse and view event descriptions using Cliopatria, a generic semantic search web-server [12]. We defined views and facets only in terms of our event model but rely on our mappings to translate the legacy event collections to these views.

Congressional Biographies. The Biographical Directory of the U.S. Congress provides short biographical articles, as a series of statements describing life events, on every member of the United States legislature from 1774 to the present. The consistent structure allows simple extraction and modeling of events. In earlier work 69,228 events were modeled using the BIO vocabulary.

The Emma Goldman Chronology. The Emma Goldman Papers editors maintain a day-by-day chronology detailing where Emma Goldman and her associates were and what they were doing. This chronology serves as an internal reference tool, allowing the editors to make inferences about when or where documents may have been produced and to check for inconsistencies in historical accounts. Starting with a text document for the years 1910 through 1916, we produced an RDF data set by parsing dates, geocoding place names, and disambiguating personal names by linking them to DBpedia. These 1,041 Emma Goldman events were modeled using the C4DM Event Ontology.

Issues Mapping Between Vocabularies. To combine these legacy event collections with our Wikipedia events we used the mappings defined between our event model and the BIO and EO vocabularies. We found that our mappings were not sufficient to achieve our goal of using a single generic view to browse all three data sets, as there is not yet widespread support for the `owl:equivalentClass` and `owl:equivalentProperty` predicates, upon which our mappings rely. However, we were able to achieve our goal by making additional mapping statements using `rdfs:subClass` and `rdfs:subProperty`. These mappings enable us to work with multiple event collections as a unified whole without re-modeling.

⁹ <http://vocab.org/bio/0.1/>

5 Conclusions and Future Work

There is a tremendous amount of timeline and chronology data on the web. There is also increasing interest in mining descriptions of historical events from narrative text, whether for temporal visualization of search results or for exploration of archival records. Historians and journalists are increasingly interested in presenting their work as structured data complementary to or in lieu of traditional narrative text. Yet, without some effort to bridge the various data models being developed and employed within these various applications, it will remain difficult to build the dense network of relations among them that could lead to new discoveries or novel modes of experiencing historical narrative. In this paper, we have presented a principled model for linking event-centric data that draws upon a close analysis of existing event ontologies. Our initial investigations show that it is useful for modeling a variety of timeline events and for mapping between events modeled using other vocabularies.

A number of questions remain to be answered. We have argued that a core event model should include only those relations about which a stable consensus has been reached, leaving more interpretive relations to a higher-level, application-specific models. But further application experience is needed before we can determine whether we have correctly identified those relations that are intersubjectively stable, or whether (for example) participation relations are interpretation-specific and ought to be moved outside the core model. A related problem is the question of event identification. In the applications discussed above, an event is identified with a single textual description. We have made no attempt to map multiple textual descriptions to the “same” event identifier. The reason for this is that it is not clear when (if ever) we should consider two textual descriptions to be of the “same” event. If we consider (as many contemporary philosophers of history do) events to be linguistic phenomena rather than objectively existing in the past, then there is no basis for arguing that two textual descriptions of an event refer to the same thing. At best we could say that they share a name, or that they refer to the same people, places, or spans of time. On the other hand, we clearly would like to say that two descriptions of past occurrences only differing in spelling or punctuation are the same event. These are deep philosophical questions about the nature of events that will likely only be answerable pragmatically, as we see which approaches are or are not useful for specific applications.

In future work, we plan on finding and working with more event collections modeled using the other ontologies discussed here, and putting these collections to use in a variety of applications. Current applications in development include event-centric searching and browsing of full-text historical scholarship, retrieval and display of historical context for documents by querying for related events, and interfaces for exploration, visualization, and comparison of events from a particular period or region.

6 Acknowledgments

The research leading to this paper was supported by the European Commission under contract FP6-027026, Knowledge Space of semantic inference for automatic annotation and retrieval of multimedia content – K-Space, and by the U.S. Institute of Museum and Library Services under a National Leadership Grant for Libraries (award number LG-06-06-0037-06).

References

1. R. Arndt, R. Troncy, S. Staab, L. Hardman, and M. Vacura. COMM: Designing a Well-Founded Multimedia Ontology for the Web. In *6th International Semantic Web Conference (ISWC'07)*, pages 30–43, Busan, Korea, 2007.
2. M. Doerr. The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata. *AI Magazine*, 24(3):75–92, 2003.
3. A. Gangemi and P. Mika. Understanding the Semantic Web through Descriptions and Situations. In *2nd International Conference on Ontologies, Databases and Applications of SEMantics (ODBASE'03)*, pages 689–706, Catania, Italy, 2003.
4. M. Hildebrand, J. van Ossenbruggen, and L. Hardman. /facet: A Browser for Heterogeneous Semantic Web Repositories. In *5th International Semantic Web Conference (ISWC'06)*, pages 272–285, Athens, Georgia, USA, 2006.
5. J. Hobbs and F. Pan. Time Ontology in OWL. W3C Working Draft, 2006. <http://www.w3.org/TR/owl-time>.
6. C. Lagoze and J. Hunter. The ABC Ontology and Model. *Journal of Digital Information (JoDI)*, 2(2), 2001.
7. Y. Raimond, S. Abdallah, M. Sandler, and F. Giasson. The Music Ontology. In *8th International Conference on Music Information Retrieval (ISMIR'07)*, Vienna, Austria, 2007.
8. A. Scherp, T. Franz, C. Saathoff, and S. Staab. F—A Model of Events based on the Foundational Ontology DOLCE+ Ultra Light. In *5th International Conference on Knowledge Capture (K-CAP'09)*, Redondo Beach, California, USA, 2009.
9. R. Shaw and R. Larson. Event Representation in Temporal and Geographic Context. In *12th European Conference on Research and Advanced Technology for Digital Libraries (ECDL'08)*, pages 415–418, Aarhus, Denmark, 2008.
10. R. Troncy. Bringing The IPTC News Architecture into the Semantic Web. In *7th International Semantic Web Conference (ISWC'08)*, pages 483–498, Karlsruhe, Germany, 2008.
11. W. van Hage, V. Malaisé, G. de Vries, G. Schreiber, and M. van Someren. Combining Ship Trajectories and Semantics with the Simple Event Model (SEM). In *1st ACM International Workshop on Events in Multimedia (EiMM'09)*, Beijing, China, 2009.
12. J. Wielemaker, M. Hildebrand, J. van Ossenbruggen, and G. Schreiber. Thesaurus-based search in large heterogeneous collections. In *7th International Semantic Web Conference (ISWC'08)*, Karlsruhe, Germany, 2008.