

# Linked Data and Time – Modeling Researcher Life Lines by Events

Johannes Trame, Carsten Keßler, and Werner Kuhn

Institute for Geoinformatics, University of Münster, Germany  
{johannes.trame, carsten.kessler, kuhn}@uni-muenster.de

**Abstract.** Most datasets on the Linked Data Web impose a static view on the represented entities and relations between them, neglecting temporal aspects of the reality they represent. In this paper, we address the representation of resources in their spatial, temporal and thematic context. We review the controversial proposals for the representation of time-dependent relations on the Linked Data Web. We argue that representing and using such relations is made hard through the direct encoding of inadequate conceptualizations, rather than through inherent limitations of the representation language RDF. Using the example of researcher life lines extracted from *curricula vitæ*, we show how to model sequences of activities in terms of events. We build upon the event participation pattern from the DOLCE Ultralite+DnS Ontology and show how places and social roles that people play during their careers relate to events. Furthermore, we demonstrate how scientific achievements can be related to events in a career trajectory by means of temporal reasoning.

**Keywords:** Event patterns, roles, knowledge representation, Linked Data, temporal reasoning

## 1 Introduction

Researcher careers are characterized by where and when the researchers have been active. The geographic mobility of researchers plays an important role in expanding scientific knowledge and in forming centers of “scientific gravity” [29]. Not surprisingly, there is also evidence that the mobility of researchers has considerable impact on their access to financial resources and networks [5].

In order to document the careers and achievements of its researchers, the University of Münster has introduced a *centralized research information management system* (CRIS) [19]. CRIS focuses on the representation of information on persons, publications, patents, prizes and projects and their interrelation to answer questions such as “Who authored which publications?” The system’s query capabilities do not go far beyond such simple questions. Yet, scientific achievements are made while people are engaged in projects and playing a particular role in an organization for a limited and well-defined time. One is therefore often interested in questions of the type “Who did what, when, where (and why)?”

**Table 1.** Extract of the curriculum vitæ of Prof. Dr. Andreas Pfingsten, available from <https://www.uni-muenster.de/forschungaz/person/10473>.

2004–2005	Visiting Professor at University of Illinois at Urbana-Champaign
2001	Visiting Professor at University of Calgary, Canada
1997–1999	Dean and Deputy Dean at Münster School of Business and Economics, University of Münster
since 1994	Professor of Business Administration in Banks and Director of the Institute of Banking at University of Münster
1992	Visiting Professor at University of Graz, Austria
1990–1994	Professor of Economics at University of Siegen

Table 1 shows a career path retrieved from the CRIS database. Assuming that we are able to extract the spatio-temporal information about such a career path by (semi)automatic means,<sup>1</sup> this paper addresses the question how to meaningfully represent the spatial, temporal and thematic references in an academic career on the Linked Data Web. By a meaningful representation we mean in the first place a semantic model that allows for answering the question “when, where and in which role did this person perform what activities during their professional career?”, independently of any specific encodings. The model should provide a sound ontological basis for temporal reasoning, so that one can ask which achievements have been made while a person was at a certain place playing a certain role. The model’s assumptions regarding space and time need to be anchored in a well-founded ontological theory to improve semantic interoperability with other systems, such as reporting tools or CV generators.

The modeling of researcher career paths serves here as an example to discuss the controversial viewpoints on how time-dependent relations should be modeled on the Linked Data Web and on the role that the Resource Description Framework (RDF) as a knowledge representation language can play. This paper provides an overview of the various existing proposals and of why they fall short of achieving meaningful data integration.

Event-based modeling approaches have attracted attention in several domains and a number of event-centered models have been proposed. While they all highlight the importance of events, they make slightly different assumptions. Additionally, none of them is anchored in an upper level ontological theory. We propose to build on the event participation pattern from the well founded DOLCE+DnS Ultralite Ontology (DUL)<sup>2</sup> instead of inventing yet another event model. We show how to model researcher career trajectories as sequences of events by extending only a few small constructs from the DUL ontology. In this work, we stick to a particular kind of scientific achievement, namely publications. Modeling the act of publishing as an event and separating it from the resulting

<sup>1</sup> Automatically extracting entities from textual career descriptions is a research subject in its own right. We have implemented a basic tool to semi-automatically fill our model with real data.

<sup>2</sup> See [http://ontologydesignpatterns.org/wiki/Ontology:DOLCE+DnS\\_Ultralite](http://ontologydesignpatterns.org/wiki/Ontology:DOLCE+DnS_Ultralite).

publication as an information object provides us with a sound basis for temporal reasoning, which enables us to relate achievements to events in career paths.

The following section reviews different approaches to handle time in knowledge representation languages, followed by an overview and discussion of the different accounts of objects and events in foundational ontologies (Section 3). Section 4 discusses the interplay of events and roles, followed by an overview of our application in Section 5 and conclusions in Section 6.

## 2 Time in Knowledge Representation

The problem of representing time-varying information on the Semantic Web has led to numerous controversial proposals in hundreds of papers [11]. While there is still no consensus and temporal aspects are largely neglected in practice [35], this section provides an overview of existing proposals.

### 2.1 Epistemological Knowledge Representation Languages: The Issue of Binary Relations and Time

The Resource Description Framework (RDF) was initially designed to ease the deployment and processing of meta-data about web resources. While RDF has an abstract data model and syntax, providing it with some basic modeling primitives (RDF(S), the RDF Schema extension) and giving it a model theoretic interpretation, it has developed into the de-facto standard language for knowledge representation on the Linked Data Web. The basic abstract concept of the RDF data model is that of a statement in the form of a triple consisting of a subject, predicate and object. The model theoretical interpretation of RDF(S) is strongly monotonic, so that new assertions cannot falsify old ones. Moreover, RDF(S) follows the *Open World Assumption*, which makes no inferences from the absence of a statement.

These properties of RDF(S) make the modeling of time-dependent relationships challenging. If we want to model a person's relation to an organization in the role of an employee, we may introduce two classes, **Person** and **Organization**, and a property **hasEmployee**. Two instances of these classes, *OrganizationXY* and *MisterX*, can be linked by the property **hasEmployee**. Treating the role as a named property or as a subclass of natural type (e.g., modeling **Employee** as a subclass of **Person**) is common practice on the Linked Data Web; however, this solution has some implications, especially concerning the question what happens if the relation changes or if we want to know when and how long the person played that role. In the following, we will review and discuss different approaches treating such time-dependent information in RDF.

### 2.2 Time as Meta-Information

Due to the limitation to binary relations of RDF(S), *RDF reification* and *named graphs* have been proposed as solutions to provide more information about a

statement (see Figure 1). In the following, we explain these two approaches and argue why they should not be used to account for changes in relationships that reflect changes in the real world.

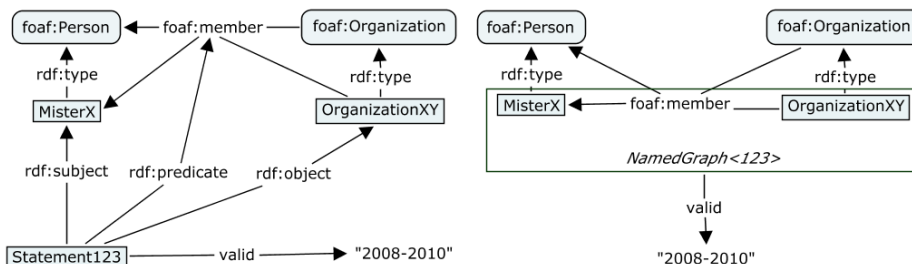


Fig. 1. RDF reification (left) and RDF named graphs (right)

**RDF Reification.** While neither the RDF XML syntax nor its model theoretic semantics support direct statements over statements, the RDF reification vocabulary has been introduced to provide meta information about single statements by turning these statements into subjects of other statements.<sup>3</sup> However, RDF reification is one of the most confusing and controversial constructs in the RDF specification, also since its meaning differs from the meaning of reification in conceptual modeling (representing n-ary relationships through binary ones) or linguistics (turning a verb into noun, see Section 3.1). RDF reification blows up the number of statements, makes querying cumbersome, and its model theoretic implications are ambiguous. Consequently, Semantic Web tools generally ignore RDF reification, and its use is discouraged [20].

**Listing 1.** RDF reification. Example taken from <http://spatial.linkedscience.org>

```

1 : acmgis/membership840 a rdf:Statement;
2   rdf:object :person/person666c1cdafc70d38a22b16775df20f004;
3   rdf:predicate foaf:member;
4   rdf:subject :affiliation/affiliation14d23c3ea0e0f08c852c0b329c8e3ee2;
5   dc:date "2008".

```

Listing 1 shows an example of RDF reification in practice [24]. However, in RDF model theory, it does not imply that the person is a member of the organization. Moreover, the property `dc:date` does not have any implications on the model theoretic interpretation of the statement. There is no way to communicate that there is even such information available, nor what it means. The meaning of the reified relation and its attached attribute values have to be defined by the application or by convention. We may query, for example, for all reified statements matching the property `dc:date` and its associated literal "2008":

<sup>3</sup> <http://www.w3.org/TR/rdf-mt/#Reif>

**Listing 2.** Querying RDF reification in SPARQL

```

1 SELECT * WHERE {
2   ?std a rdf:Statement;
3     rdf:subject ?a;
4     rdf:predicate foaf:member;
5     rdf:object ?b.
6   ?std dc:date "2008".}

```

But then, what does the result mean? Does it mean ?b was member of ?a *during* 2008? or was the statement that ?b is a member of ?a *created* in 2008? or was the statement *valid until* or valid from 2008?

**Named graphs.** Because of these semantic problems with RDF reification, metadata on statements is now widely provided through named graphs [6], extending triples to quad-tuples. This development has been driven by the various supersets of RDF syntax and grammar such as N3 or TriG, which extend the triple concept to quad-tuples (or sometimes called context-quads).<sup>4</sup>

**Listing 3.** Example for a NamedGraph encoded in TriG syntax

```

1 :acmgis/membership840 {
2   :person/person666c1cdafc70d38a22b16775df20f004
3     foaf:member affiliation/affiliation14d23c3ea0e0f08c852c0b329c8e3ee2. }
4 :acmgis/membership840 dc:date "2008".

```

Named graphs provide a lightweight but useful extension of the initial design encoding and query languages. Named graphs are widely accepted by the community to provide provenance information and they have been incorporated into the W3C recommendation for SPARQL 1.1.<sup>5</sup> Since named graphs are formally subgraphs, they have a clearer syntax and semantics. They are easier to handle and more widely supported by software tools than RDF reification. However, as pointed out by Carroll et al. [6] the semantics are intentionally limited and conform to the basic interpretation of RDF(S) in favor of simplicity. Named graphs are explicitly designed to ease the handling of collections of statements and to attach meta information to those collections, such as provenance or licensing information.

In summary, both RDF reification and named graphs can in principle be used to handle temporally varying information by attaching timestamps to single statements [24] or to (sub)graphs [40]. However, as the temporal dimension stays on a meta-level, it has no implication for the model theoretical interpretation and its existence cannot be stated in the vocabulary. Therefore, named graphs should only be used for temporal information belonging to the meta level, such as provenance data.

<sup>4</sup> See <http://www.w3.org/TeamSubmission/n3/> and <http://wifo5-03.informatik.uni-mannheim.de/bizer/trig/>.

<sup>5</sup> See <http://www.w3.org/TR/rdf-sparql-query/>.

### 2.3 Extensions of RDF(S)

Since both approaches to time are similar to traditional database techniques (time-stamping or temporal versioning), many researchers have taken up that direction, trying to provide a model-theoretic account of time. Within the database community, extensive research has been conducted on two notions of time: valid time (when a change occurred in the real world) and transaction time (when a change was entered to the database) [22]. Various proposals adapting the notion of valid time have been made by the Linked Data community, such as temporal RDF graphs (temporal reification vocabulary) [16, 15], multidimensional RDF (extended triple notion) [10], applied temporal RDF (named graphs) [43], stRDF (temporal quad) [25], RDF SpatialTemporalThematic (based on temporal graphs) [34], and temporal quintuples [26].

These approaches have in common that they either extend the RDF syntax or abuse RDF reification or the context quad in order to “label” a triple with a timestamp. The temporal label is then given a model theoretic interpretation modifying the truth value of the statement. This is done by extending the basic entailment rules of RDF(S) or even by moving to a different logic. However, these approaches are in many respects debatable, as they treat the representation of time as a feature of the encoding language. It is unlikely that all existing software implementations will adapt such syntactic and model theoretic extensions. Moreover, time-stamping triples provides no means to share the underlying conceptual model. We can only see changes between different versions, but we fail to explain where these changes in the real world come from. Finally, it is not clear how the notion of valid time interacts with the Open World Assumption, which supports contradicting statements.

### 2.4 N-Ary Relations

The natural way to deal with the restriction of binary relations in modeling languages is to use the n-ary design pattern. This approach is also known as *conceptual reification* of binary relations. The basic idea is to turn a property into a class and link it to the existing classes via two additional properties. Figure 2 shows an example suggested by the organization ontology.<sup>6</sup> From a practical point of view, n-ary patterns have been criticized because they increase the number of statements and lead to a proliferation of objects. Additionally, the introduction of new (anonymous) individuals through reified relationships causes a maintenance problem and limits the usefulness of RDF and OWL constructs such as domain and range restrictions or inverse property definitions.

The biggest problem, however, is that the potential application scope of the logical n-ary pattern is very broad. The n-ary pattern is often considered as an ad-hoc workaround and thus it is frequently used in an arbitrary fashion, lacking any design rationale. Various modeling solutions are possible, especially since RDF(S), as an epistemological KR language, should be neutral regarding

<sup>6</sup> <http://www.w3.org/TR/vocab-org/>

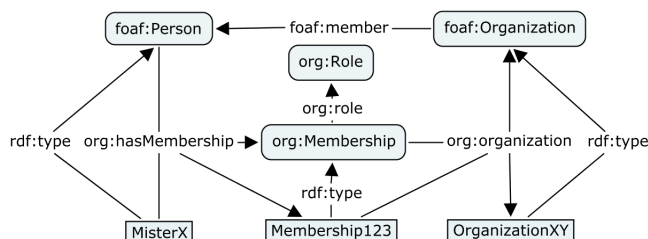


Fig. 2. Example of an n-ary relationship in RDF from the organization ontology.

metaphysical and ontological assumptions [13]. One may, for example, treat roles as binary properties or as instances of classes, making ontological commitments that remain implicit [13]. Without any explicit design rationale, reaching even a partial agreement on the conceptual level is unlikely.

### 3 Foundational Ontologies: Objects and Events

The use of foundational ontologies has been proposed to make ontological commitments in conceptual models explicit. With its cognitive bias [31], the Descriptive Ontology for Linguistic and Cognitive Engineering<sup>7</sup> (DOLCE), fits the need of the Semantic Web as a frame of reference for building domain ontologies [9]. It organizes particulars and axiomatizes them according to a set of meta-properties [14]. While DOLCE is relatively complex in its full extent, lightweight versions are provided and can be extended in a modular fashion, for example DOLCE-Lite-Plus<sup>7</sup> or DOLCE+DnS Ultralite (DUL).

#### 3.1 Linguistic View on Time

Linguists face the problem of how to capture adverbial modifiers (locative, temporal or instrumental) of action sentences in logic. Sentences such as “Johannes wrote a paper with a pen in the library at midnight” result in predicates of higher arity, such as `write(Johannes, a paper, a pen, the library, midnight)`, where the number of adjuncts for the predicate can become very large. As a solution to this *variable polyadicity* problem of action modifiers, Davidson [7] proposed that action predicates should explicitly range over an additional, normally hidden, event argument:  $\exists e [\text{write}(\text{Johannes}, \text{a paper}, e) \wedge \text{with}(e, \text{pen}) \wedge \text{in}(e, \text{library}) \wedge \text{at}(e, \text{midnight})]$ . This event argument can then be existentially quantified and bound to the whole sentence, whereas each adjunct will be attached as a separate conjunction clause, thus allowing for an arbitrary number of adjuncts.

Keeping the basic idea of the event argument, the so-called Neo-Davidson approaches have been established as the standard event semantics in linguis-

<sup>7</sup> See <http://www.loa.istc.cnr.it/DOLCE.html>.

tics [30]. Parsons [33] advocated to treat the event argument as the only argument of the predicate and link it with a set of thematic roles. Following the proposal by Parson, the structure of our example can be “. . . replaced by a truly compositional theory of predicates whose fundamental notion is that of event participant and whose fundamental predicates are [thematic roles]” [21, p.444]:  $\exists e [\text{write}(e) \wedge \text{agent}(e, \text{Johannes}) \wedge \text{location}(e, \text{library}) \wedge \text{time}(e, \text{midnight})]$ . The decomposition enables one to represent temporal relations directly in first order structures and to reason over them without the need to move, for example, to a temporal or modal logic.

### 3.2 Ontological Event Theories

A dualistic view, where objects and events complement each other [41], is reflected in most upper-level ontologies such as the Basic Formal Ontology (BFO)<sup>8</sup> or DOLCE [31]. The fundamental distinction is between occurrences and continuants, things that are primarily in space and things that are primarily in time [17, 4]. Objects get then located in time through their participation relation in events and events get located in space through their physical participants. The distinction also helps identifying and separating essential properties from temporally changing ones [38]. Spatial characteristics have an important function for identifying physical objects as well as for defining their unity criteria. Temporal characteristics play the same role for events [4]. We can distinguish two physical objects if they have different spatial locations at the same time. Events are different, if they occur at different times.

Events provide the temporal context in which some relations hold. Roles, in particular, depend on events to come into existence. They carry identity criteria, but do not supply them. Neglecting this may lead to the so-called *isA-overloading*, for example, placing a concept *Employee* under *Person* in the taxonomy) [12]. Extensive work on the representation of roles has been done by Steimann [42], who presented a list of fifteen different features that may apply to roles. In this list, Masolo et al. [32, pp.269–270] identified five characteristics that refer to the dynamic and temporal nature of roles:

- (1) an object can play different roles at the same time
- (2) an object can change its roles
- (3) an object can play the same role for multiple times simultaneously
- (4) some roles can be played by different objects, at the same or different times
- (5) some roles can only be played by an object depending on the previous roles it played

Loebe and August [28] propose to distinguish between three types of roles, namely relational roles, processual roles, and social roles based on their players and contexts. However, they emphasized that in some cases (for example, for the role *Student*) all three modeling approaches are valid solutions depending on the intention. Masolo et al. [32] argue that the problem with a contextual approach is that the notion of context is still quite fuzzy on its own and one contextual

<sup>8</sup> <http://www.ifomis.org/bfo>



approach may subsume the other. Consequently, they suggest to reify roles and separate them from their specification which accounts for the relational and contextual nature of roles. The object can then be classified for a certain time. This pattern rests on the so called *Descriptions and Situations* (DnS) extension of DOLCE and is also part of the DUL library. While the DnS pattern does not solve all problems of roles, it provides a practical modeling solution that accounts for their dynamic and relational nature.

### 3.3 Summary

Some researchers see the problem of capturing time in RDF analogously to that of representing time in traditional database systems, since RDF(S) is restricted to binary relations and does not allow to assert when a statement is or was valid in the “real world”. Consequently, many proposals have been made that treat time on the meta-level or by extending the syntax and semantics of RDF(S). However, an ontological view reveals that many problems regarding temporal relations result from conceptualizations and from the use of epistemological knowledge representation languages to encode them. Ad-hoc modeling solutions fail to capture the ontological commitments underlying conceptual models.

## 4 Career Trajectories – Events and Roles

This section focuses on our example problem of how to model researcher career trajectories. The basic idea is to model the career trajectory as a number of events, which relate to certain places and in which researchers play a particular social role. We will shortly discuss existing event models and explain why we do not adapt them, but build upon DOLCE+DnS UltraLite.

### 4.1 Existing Event Models

Event ontologies have gained attention in recent years. Examples include CIDOC CRM [8], the ABC Ontology [27] or the CultureSampo project [36], where event-based modeling approaches have proven useful to establish a common conceptual reference frame across applications. While these are domain-specific examples, the Event Ontology (EO)<sup>9</sup> aims for more generic event modeling patterns. All these approaches differ slightly in their terminology and regarding their conceptual assumptions. The authors of the LOD ontology, which provides a mapping between different event models, provide a good overview of the differences [39]. While LOD and EO do not account for roles, CIDOC CRM [1] realizes roles as *properties over properties* which cannot be implemented in RDF. Most recently, the Simple Event Model (SEM) [18] has been proposed, as a light-weight model that is neutral with respect to semantic commitments.

However, all concepts and relations in SEM are undefined primitives that are only informally and weakly specified. It is unclear what exactly distinguishes

<sup>9</sup> <http://motools.sourceforge.net/event/event.html>

events from objects and whether these concepts are disjoint or not. Knowing this is, however, essential for knowledge sharing, as different conceptualizations may only be understood by means of a shared understanding of some basic distinctions. Consequently we chose to build on the light-weight version DUL of DOLCE. DUL provides a good combination of flexibility and expressiveness and makes at least the basic distinctions explicit, such as that between events, physical objects and social constructions. In DUL the DOLCE distinction between endurants and perdurants is simplified to a distinction between objects and events, which suffices for most situations and avoids the heavy terminology. Classes have more intuitive names in DUL compared to DOLCE. Moreover, DUL is designed in a modular fashion from which many different ontological content design patterns can be extracted and then combined seamlessly.<sup>10</sup> As we will show, only small extensions are required to DUL to model researcher careers in terms of events. An alignment to DUL provides us with conceptual clarity and serves as a reference frame in order to communicate the basic ontological commitments, which remain ambiguous when using the models mentioned above. Our approach is similar to the F-Event model [37] which also builds on DUL and provides a number of specialized instantiations of descriptions and situations (DnS). By introducing certain types of events and situations, events can be composed to form more complex situations, such as mereological compositions of events or causality relations among events. While this is useful, these patterns add far more complexity than required here. Furthermore, the F-Event model does not suggest how to distinguish among different kinds of participants in an event.

#### 4.2 Basic Design Decisions and Alignment to DOLCE+DnS Ultralite

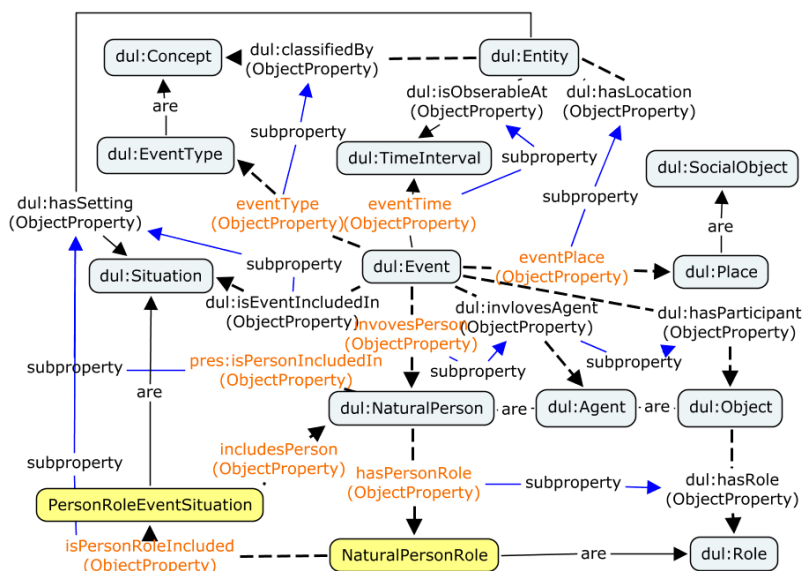
**Events.** Events unfold over time, which means that they can be located directly in a temporal region. In DUL, the region values can be directly added to the `Event` using the `hasEventDate`<sup>11</sup> property, or the temporal region, such as a `TimeInterval`, can be separated from the `Event` and linked via the property `isObservableAt`, which is a sub-property of `hasRegion`. While the former approach is simple on the query and application level, the latter provides a “cleaner” and more flexible solution in representing time values from other calendars. Particularly, in the case of temporal intervals we may have a temporal start and end value. These two values belong logically together denoting the boundaries of the interval, rather than the boundaries of the event. Furthermore, by representing the interval explicitly, the model could be extended to account for fuzzy temporal intervals [23].

We follow the second approach here and separate the region values from the event, even if it comes at the cost of simplicity. Since the `isObservableAt` property in DUL is in the domain of `Entity` (Objects, Events, Qualities, Abstracts),

<sup>10</sup> <http://ontologydesignpatterns.org/wiki/Category:ContentOP>

<sup>11</sup> The property ranges over data values of the XML date schema: <http://www.w3.org/TR/xmlschema-2/#date>

we introduce a new sub-property `eventTime`, explicitly linking the `Event` to its region `TimeInterval`. In order to link the `TimeInterval` to its region values we introduce the properties `hasIntervalStartDate` and `hasIntervalEndDate` ranging over values from the XML date schema. Both are specializations of the `hasIntervalDate` and `hasRegionDataValue` property in DUL.



**Fig. 3.** Overview of the classes and patterns taken from DUL. Extensions to DUL are colored. An RDF(S) encoding is provided online <http://vocab.lodum.de/pres/>.

Concerning the classification of events, we take a pragmatic approach and link the `Event` to an `EventType` which is a subclass of `Concept`. By this we separate the event from its interpretation, however, we do not provide a situational context. Since the property `isClassifiedBy` is the domain of all entities, we introduce a new sub-property `eventType` with a domain constraint on `Event` and range constraint on `EventTypes`. This allows us to dynamically type event concepts from external vocabularies as `EventTypes`.

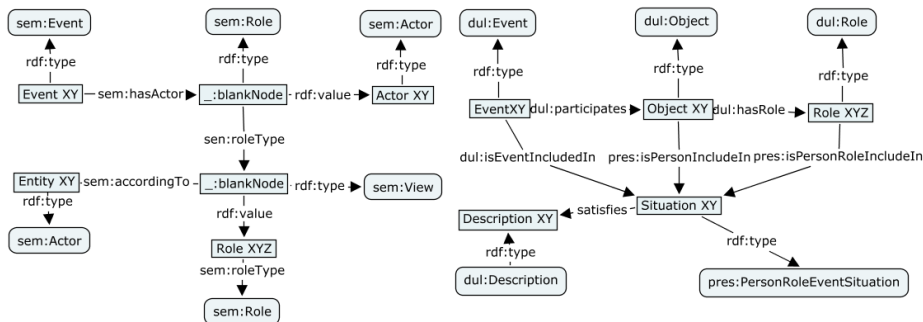
**Participants in Events.** While generally all objects can participate in events via the `isParticipantIn` relation, the primary participants in our case are persons and organizations. Since organizations are hard to extract consistently from CVs, we focus on persons at this point; however, organizations can easily be added later. A `NaturalPerson` in DUL belongs to the category of agentive physical objects because they have intentionality and a proper space region. We specialize the `involvesAgent` property of DUL by introducing a new property

`involvesPerson`, with domain constraints on `Event` and range constraints on `NaturalPerson`.

Because we aim to reuse the model in different contexts, we can add further participants as necessary. Participants can then structurally be distinguished based on their functional participation relation (thematic roles). These have already been encoded in the DOLCE Lite extension “Functional Participation Ontology”.<sup>12</sup> Thematic roles are realized as sub-properties of the `hasParticipant` property.

**Social Roles.** We assume that all roles played by researchers during their career belong to the category of social roles. A `Role` belongs to the class `Concept`, as it is externally defined in a description (which we simply assume to exist here). Concepts are social (mental) objects and consequently disjoint from physical agents. The property `classifies` relates the `Role` to any `Entity`. DUL provides a `hasRole` property which constrains the domain to objects. For more domain specific typing, we introduce a new sub-property `hasPersonRole` with domain constraints on `Person` and range constraints on a new sub-class of roles `NaturalPersonRole`. This provides us with a placeholder for the different classes of roles which people may hold, for example, a *Lecturer*.

We introduce a specialized situation called `PersonRoleEventSituation` in order to express that the classification holds in relation to the event, which is quite similar to the *Participant-Role* pattern.<sup>13</sup> Two additional properties are `isPersonRoleIncludedIn` and `isPersonIncludedIn`, so that one can assert OWL cardinality restrictions to the `PersonRoleEventSituation` class. The way we represent roles is the major difference to existing models like the Simple Event Model (SEM) [18].



**Fig. 4.** Comparison of role representations between the SEM model (left) and our role model (right)

<sup>12</sup> <http://www.loa.istc.cnr.it/ontologies/FunctionalParticipation.owl>

<sup>13</sup> <http://ontologydesignpatterns.org/wiki/Submissions:ParticipantRole>

SEM shares our idea that social roles of participants in events depend on a particular external view. However, the way SEM models this leads to implementation problems. SEM suggests to replace the actor in an event by a blank node, which can then be typed as a role and links to particular role-type instances, which is again a blank node (see Figure 4). SEM then links from the blank node to the actual participant using the `rdf:value` property, which has no meaning on its own and cannot be constrained with domain and range restrictions. What SEM actually tries is to imitate an n-ary relation, however, by using blank nodes and the `rdf:value` it makes things unnecessarily complicated. Blank nodes are difficult to handle, and using them as an intermediate anonymous node also changes the graph structure. Using the situation setting in DUL provides us with a flexible view on the event, but does not change the logical structure. Furthermore, it clearly separates the event and its participants from the social interpretation.

**Place.** Since events do not have spatial qualities, they can only be located indirectly through the location of their participants [4]. This can be difficult, however, if participant locations cannot easily be identified. In our example, there is no physical participant that is located inherently. Even if we were able to extract and identify organizations from CV’s as participants in an event, the problem would remain that an organization is a social object that by itself does not have a proper space region. Postal addresses might be resolvable to a location in a generic sense (e.g. a political entity). We thus use the social concept `Place` and specialize the generic `hasLocation` by introducing a property `eventPlace`. As the socially constructed concept of place needs to be distinguished from its (approximated) physical region, we introduce a new property `approxGeoLoc`. This is not to say that the place has a region, but that it is approximately located somewhere in a region. While the place might cease to exist, the space region remains.

## 5 Application

In this section we apply our event model to (semi)automatically extracted data on curricula vitæ of researchers. Exemplary queries show that the model fulfills the basic requirements. Furthermore, we demonstrate how the distinction between objects and events is used to reason about researcher achievements and career events.

### 5.1 Information Extraction from Researchers’ Curricula Vitæ

In the following, we briefly explain how we extracted the relevant pieces of information from the university research database (compare Section 1, Table 1) and how we “filled” our model with this data.

**Temporal Information.** For each database entry, we create a new event and an interval resource. The person entity to which the database entry belongs is already available through LODUM.<sup>14</sup> We link the event resource to the person resource using the `involvesPerson` property. Since we do not know when specifically a position started or ended, we assume maximal boundaries for the interval (first/last day of the year or month).

**Social Roles.** Extracting the temporal information is straightforward, but extracting event information from the textual description is more challenging. We implemented some rules according to which we split the description into smaller chunks (punctuations and prepositions). We assume that the role is usually named first and followed by the name and address of the organization or place. For the identification of role concepts, we use DBpedia Spotlight,<sup>15</sup> a tool for automatically annotating textual descriptions with DBpedia.<sup>16</sup> In the case that a role has been identified by DBpedia Spotlight, we take the labels from the DBpedia resource to create new role concepts. Out of the concepts, we create a role resource. At the same time a new instance of the class `PersonEventRoleSituation` is created, which relates the event, the person and the role to each other (see Section 4).

**Places.** While the Geonames Gazetteer<sup>17</sup> is the most widely used hub for geographic information on the Linked Data Web, its natural language processing capabilities turned out to be too limited for our purpose. Instead, we have used the Google Geocoder API.<sup>18</sup> It is primarily an API for resolving addresses to geographical coordinates, but it also returns a structured list of named entities and their associated types, which have been identified from the input. Out of this list we create a basic place hierarchy, where places are related to each other via the informal `parentPlaceOf` property.

## 5.2 Querying and Reasoning Capabilities

The core model provides answers to the following questions:

1. In which events does a particular person participate?
2. What are the social roles that a given person plays in an event?
3. Where did the event take place?
4. When did the event start and end?
5. Which concept is assigned to the event?

We formalized and combined several of these questions into one SPARQL query that our prototypical application builds on (see Figure 6).

<sup>14</sup> Linked Open Data University of Münster; see <http://lodum.de>.

<sup>15</sup> <https://github.com/dbpedia-spotlight/dbpedia-spotlight/wiki/Web-service>

<sup>16</sup> <http://dbpedia.org/>

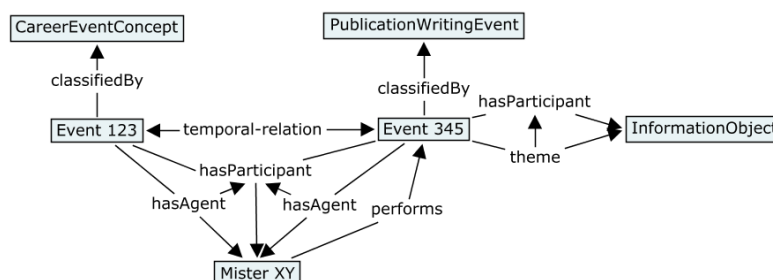
<sup>17</sup> <http://www.geonames.org/>

<sup>18</sup> <https://developers.google.com/maps/documentation/geocoding/>

Since events are time-based and clearly distinguished from objects, we can establish temporal relations between them. Based on Allen’s interval calculus [3], some basic relations have already been encoded as OWL properties by others.<sup>19</sup> In order to infer these relations automatically, we have formalized them by inference rules using the Jena Rule Syntax.<sup>20,21</sup>

Generally, temporal relations exist between all events, independently of their participants. However, here we are interested in temporal relations between events involving the same participant. Therefore, we restrict the inference rules to those situations, reducing the computational overhead. The relations enable simple queries such as “which event precedes this event?” or “which events happen during this event?”.

In the context of our use case, this means that we can relate the career events temporally to other kinds of events, for example, to events through which scientific achievements have been made. In order to exemplify this idea, we construct some new events from the meta-data about publications. We call them **PublicationWritingEvents** and create them by making the simplified assumption that an article is written within at most a year before it is published. For representing these events, we apply the basic Event-Participation Pattern. We separate the publication as an **InformationObject** from the event and link it as a participant via the thematic role **theme-of**<sup>12</sup> to the event. The person participates also by a thematic role **performs**<sup>12</sup> in the event.



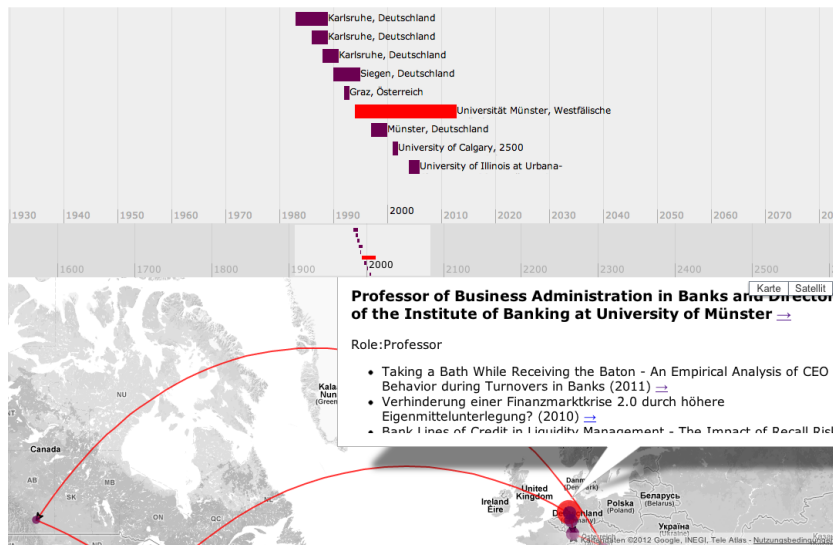
**Fig. 5.** InformationObjects as functional participants (thematic roles) in events.

Temporal relations between the “Career Event” and the “Publication Writing Event” can be derived from the aforementioned set of temporal inference rules. This way we can retrieve, for example, all publications that have been written while a person played a certain role at a certain place. As a demo for the querying capabilities of our model, we have implemented an explorative user-interface which combines a timeline view with a map interface, see Figure 6.

<sup>19</sup> <http://www.loa.istc.cnr.it/ontologies/TemporalRelations.owl>

<sup>20</sup> <http://jena.apache.org/documentation/inference/index.html#RULEsyntax>

<sup>21</sup> <http://vocab.lodum.de/pres/temporal.rules>



**Fig. 6.** Prototype application to visualize a researcher career trajectory on a timeline, along with a map of associated places. The circle sizes on the map approximate the number of publications that have been written by the person during the time period the person played a role at a place; see <http://data.uni-muenster.de/php/timemap/timemap.html>

## 6 Conclusions

We have discussed a variety of proposals regarding the representation of time-dependent relations on the Linked Data Web. Many of these proposed solutions are inspired by the database community, which traditionally treats time at the time stamp level, for example, by assigning validity time stamps to facts. However, doing so fails to provide the means for sharing conceptual assumptions about events with other agents. One can only try to infer real world changes from version changes.

As stated by Allen and Ferguson [2, p.535] “... events are primarily linguistic or cognitive in nature. That is, the world does not really contain events. Rather, events are the way by which agents classify certain useful and relevant patterns of change.” This creates all sorts of semantic interoperability problems if event conceptualizations are not made explicit. Treating events as first-class citizens is an essential requirement to make changes traceable [44], as events are the observables in which the semantics of many relationships can be grounded.

As a practical application of our approach, we have shown how researcher career trajectories can be modeled by means of events. We explained how these events relate to places and to social roles that researchers play while participating in events. We built our approach upon the Dolce+DnS Ontology. DOLCE UltraLite does not only provide conceptual clarity, but also servers to commu-



nicate basic ontological distinctions (like those between objects and events, or different kinds of event participants) to users of temporal data. These distinctions create a sound basis for temporal reasoning that relates different career events, but also to relate these events to achievements during a career. By extracting relevant data from *curricula vitæ* and feeding them into our model, we have demonstrated that the event centered modeling approach provides a solution to represent the spatial, temporal, and thematic references in academic life lines meaningfully and usefully. Compared to underspecified event models and workarounds such as RDF reification or named graphs, our solution provides both unambiguous semantics and straightforward querying of time-dependent properties.

The next steps in this research will be to address vague intervals. A clean modeling pattern for such intervals is especially relevant for events where the exact time frame is unknown, but the ordering relations are known – e.g., a paper is always written *before* it is published. On the implementation side, we plan to evolve data annotations in LODUM and `spatial@linkedsience` [24] to the event model introduced in this paper.

## Bibliography

- [1] Alexiev, V.: Types and Annotations for CIDOC CRM Properties. Digital Presentation and Preservation of Cultural and Scientific Heritage (Report) (2011), <http://www.ontotext.com/sites/default/files/publications/CRM-Properties.pdf>
- [2] Allen, J.F., Ferguson, G.: Actions and events in interval temporal logic. *Journal of logic and computation* 4(5), 531–579 (1994)
- [3] Allen, J.F.: Maintaining knowledge about temporal intervals. *Communications of the ACM* pp. 832–843 (1983)
- [4] Borgo, S., Masolo, C.: Foundational choices in DOLCE. In: Staab, S., Studer, R. (eds.) *Handbook on ontologies*, pp. 361–381. Springer Berlin / Heidelberg, second edn. (2009)
- [5] Cañibano, C., Otamendi, J., Andújar, I.: Measuring and assessing researcher mobility from CV analysis: the case of the Ramón y Cajal Programme in Spain. *Research Evaluation* 17(1), 17–31 (2008)
- [6] Carroll, J., Bizer, C., Hayes, P., Stickler, P.: Named graphs. *Web Semantics: Science, Services and Agents on the World Wide Web* 3(4), 247–267 (2005)
- [7] Davidson, D.: The logical form of action sentences. *Essays on actions and events* (1967), [http://ontologics.net/download/dot/Publication/sub\\\_Davi1967a.pdf](http://ontologics.net/download/dot/Publication/sub\_Davi1967a.pdf)
- [8] Doerr, M.: An Ontological Approach to Semantic Interoperability of Metadata 24(3), 75–92 (2003)
- [9] Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., Schneider, L.: Sweetening Ontologies with DOLCE. In: Gómez-Pérez, A., Benjamins, V. (eds.) *Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web*, Lecture Notes in Computer Science, vol. 2473, pp. 223–233. Springer Berlin / Heidelberg (2002)

- [10] Gergatsoulis, M., Lilis, P.: Multidimensional RDF (Epeaek Ii), 1188–1205 (2005)
- [11] Grandi, F.: An Annotated Bibliography on Temporal and Evolution Aspects in the Semantic Web. Tech. rep., Universita di Bologna, Bologna, Italy (2012), <http://timecenter.cs.aau.dk/TimeCenterPublications/TR-95.pdf>
- [12] Guarino, N.: Formal Ontology and Information Systems. In: International Journal of Human-Computer Studies - Special issue: the role of formal ontology in the information technology. pp. 3–15. IOS Press (1998)
- [13] Guarino, N.: The ontological level : Revisiting 30 years of knowledge representation. In: Conceptual Modeling: Foundations and Applications, pp. 52–67. Springer (2009)
- [14] Guarino, N., Welty, C.: A formal ontology of properties. Knowledge Engineering and Knowledge Management Methods, Models, and Tools pp. 191–230 (2000)
- [15] Gutierrez, C.: Introducing time into RDF. Knowledge and Data Engineering 19(2), 207–218 (2007)
- [16] Gutierrez, C., Hurtado, C., Vaisman, A.: Temporal RDF. The Semantic Web: Research and Applications pp. 167–199 (2005)
- [17] Hacker, P.M.S.: Events and Objects in Space and Time. Mind XCI(361), 1–19 (1982)
- [18] Hage, W.V., Malaisé, V., Segers, R.: Design and use of the Simple Event Model (SEM). Web Semantics: Science, Services and Agents on the World Wide Web 9(2), 128–136 (2011)
- [19] Herwig, S., Höllrigl, T., Ag, A.: All roads lead to Rome: Establishing Best Practices for the Implementation and Introduction of a CRIS: Insights and Experiences from a CRIS Project at the University of Münster. In: Jeffery, K.G., Dvoák, J. (eds.) Proceedings of the 11th Int. Conference on Current Research Information Systems. pp. 93–102. Prague, Czech Republic (2012)
- [20] Hogan, A., Umbrich, J., Harth, A., Cyganiak, R., Polleres, A., Decker, S.: An empirical survey of Linked Data conformance. Web Semantics: Science, Services and Agents on the World Wide Web 14, 14–44 (Jul 2012)
- [21] Hornstein, N.: Events in the Semantics of English: A Study in Subatomic Semantics. Mind & Language 8(3), 442–449 (1993)
- [22] Jensen, C.S., Snodgrass, R.T.: Temporal data management. Knowledge and Data Engineering, IEEE Transactions on 11(1), 36–44 (1999)
- [23] Kauppinen, T., Mantegari, G., Paakkarinen, P., Kuittinen, H., Hyvönen, E., Bandini, S.: Determining relevance of imprecise temporal intervals for cultural heritage information retrieval. International journal of human-computer studies 68(9), 549–560 (2010)
- [24] Keßler, C., Janowicz, K., Kauppinen, T.: spatial@linkedscience – Exploring the Research Field of GIScience with Linked Data. In: Xiao, N., Kwan, M.P., Goodchild, M., Shekhar, S. (eds.) Proceedings of the 7th Int. Conference on GIScience 2012, Columbus, OH, Lecture Notes in Computer Science, vol. 7478, pp. 102–115. Springer (2012)

- [25] Koubarakis, M., Kyzirakos, K.: Modeling and querying metadata in the semantic sensor web: The model strdf and the query language stsparql. *The semantic web: research and applications* pp. 425–439 (2010)
- [26] Krieger, H.u.: A Temporal Extension of the Hayes/ter Horst Entailment Rules and a Detailed Comparison with W3Cs N-ary Relations 49(681) (2011)
- [27] Lagoze, C., Hunter, J.: The ABC ontology and model. In: *DC-2001: International Conference on Dublin Core and Metadata Applications 2001*. vol. 2, pp. 1–18. British Computer Society and Oxford University Press (2001)
- [28] Loebe, F., August, N.: *An Analysis of Roles* (2003)
- [29] Mahroum, S.: Scientific Mobility: An Agent of Scientific Expansion and Institutional Empowerment. *Science Communication* 21(4), 367–378 (2000)
- [30] Maienborn, C., Wöllstein, A.: *Event arguments: foundations and applications*, vol. 501. De Gruyter (2005)
- [31] Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A., Horrocks, I.: *WonderWeb Deliverable D18 Ontology Library, WonderWeb Project (final)*. Tech. rep. (2003)
- [32] Masolo, C., Vieu, L., Bottazzi, E., Catenacci, C.: Social roles and their descriptions. *Procs. of KR* 4, 267–277 (2004)
- [33] Parsons, T.: *Events in the Semantics of English*. MIT Press, Cambridge; Massachusetts (1990)
- [34] Perry, M.S.: *A framework to support spatial, temporal and thematic analytics over semantic web data*. Ph.D. thesis, Wright State University (2008)
- [35] Rula, A., Palmonari, M., Harth, A., Stadtmüller, S., Maurino, A.: On the Diversity and Availability of Temporal Information in Linked Open Data. In: Philippe, C.M., Jeff Heflin, J., Sirin, E., Tudorache, T., Euzenat, J., Hauswirth, M., Xavier Parreira, J., Hendler, J., Schreiber, G., Bernstein, A., Blomqvist, E. (eds.) *ISWC 2012 – 11th Int. Semantic Web Conference*. pp. 492–507. Springer, Boston, MA, USA (2012)
- [36] Ruotsalo, T., Hyvönen, E.: An event-based approach for semantic metadata interoperability. *The Semantic Web* pp. 409–422 (2007)
- [37] Scherp, A., Franz, T., Saathoff, C., Staab, S.: F-A model of events based on the foundational ontology dolce+ DnS ultralight. In: *Proceedings of the 5th Int. Conference on Knowledge Capture – K-CAP*. pp. 137–144 (2009)
- [38] Semy, Salim K. Pulvermacher, Mary K. Obrst, L.J.: *Toward the Use of an Upper Ontology for U.S. Government and U.S. Military Domains: An Evaluation*. Tech. rep. (2004)
- [39] Shaw, R., Troncy, R., Hardman, L.: LODE: Linking Open Descriptions of Events. In: Gómez-Pérez, A., Yu, Y., Ding, Y. (eds.) *The Semantic Web, Lecture Notes in Computer Science*, vol. 5926, pp. 153–167. Springer Berlin / Heidelberg (2009)
- [40] Sheridan, J., Tennison, J.: Linking UK Government Data. In: Bizer, C., Heath, T., Berners-Lee, T., Hausenblas, M. (eds.) *WWW2010 workshop: Linked Data on the Web. CEUR Workshop Proceedings*, vol. 628. CEUR-WS.org (2010)

- [41] Simon, J.: How to be a Bicategorialist. In: Varzi, A., Vieu, L. (eds.) *Formal Ontology in Information Systems: Proceedings of the Third International Conference (Fois-2004)*. p. 60. Ios PressInc, Ios PressInc (2004)
- [42] Steimann, F.: On the representation of roles in object-oriented and conceptual modelling. *Data & Knowledge Engineering* 35(1), 83–106 (Oct 2000)
- [43] Tappolet, J., Bernstein, A.: Applied temporal RDF: Efficient temporal querying of RDF data with SPARQL. In: *ESWC 2009 Heraklion Proceedings of the 6th European Semantic Web Conference on The Semantic Web: Research and Applications*. pp. 308 – 322. No. 200021 (2009)
- [44] Worboys, M., Hornsby, K.: From objects to events: Gem, the geospatial event model. In: Egenhofer, M., Freksa, C., Miller, H. (eds.) *Geographic Information Science, Lecture Notes in Computer Science*, vol. 3234, pp. 327–343. Springer Berlin Heidelberg (2004)