

Ontological Analysis of Observations and Measurements

Florian Probst

Institute for Geoinformatics, University of Münster, Robert-Koch Str. 26-28,
48149 Münster, Germany
f.probst@uni-muenster.de

Abstract. Geographic information is based on observations or measurements. The Open Geospatial Consortium (OGC) has developed an implementation specification for observations and measurements (O&M). It specifies precisely how to encode information. Yet, the O&M conceptual model does not specify precisely which real-world entities are denoted by the specified information objects. We provide formal semantics for the central O&M terms by aligning them to the foundational ontology DOLCE. The alignment to a foundational ontology restricts the possible interpretations of the central elements in the O&M model and establishes explicit relations between categories of real world entities and classes of information objects. These relations are essential for assessing semantic interoperability between geospatial information sources.

Introduction

Geographic information is based on observations or measurements. The Open Geospatial Consortium (OGC) has developed an implementation specification for observations and measurements (O&M). The purpose of this specification is to provide “a conceptual model and encoding for observations and measurements” [1, p.1].

This purpose indicates that two fundamentally different things are specified. First, a conceptual model specifies the relevant concepts in the domain of interest [2]. Second, an information encoding model specifies how to produce information items *representing* concepts and their instances specified in the conceptual model. The O&M specification provides an information encoding model in UML. It specifies precisely the internal structure of information objects. Yet, the conceptualization of the geospatial real-world entities is described only in natural language and reflected only implicitly in the UML diagrams.

Textual descriptions allow a wide scope of interpretation of what the specified information objects denote in geospatial reality. The current O&M conceptual model is not suitable for assessing and achieving semantic interoperability between information sources. It specifies the conceptualization of an information object which *represents* an observation; however, it makes not formally explicit the conceptualization of an observation as such. Observations are central to the creation of geographic information. Thus, it seems desirable for the GI community to have a formal and explicit

method at hand to express the meaning of terms related to observation events, independent of the information encoding.

We follow Guizzardi [3] in defining ontology modeling as special type of conceptual modeling. Ontology modeling focuses on structuring the entities existing in the domain of interest. Thus, one can distinguish an ontology-based conceptual model for domain entities and a conceptual model for information objects.

The work is driven by the need for a framework enabling information providers to specify precisely which real world entities are denoted by the information objects they provide.

The authors of the O&M specification request that “an ontology of observable phenomena must be available” [1 p.15] in order to improve discovery and retrieval of O&M conformant data sources. We present, as a first step towards this goal, an ontology of the core O&M terms which is aligned to the foundational ontology DOLCE [4]. The alignment is performed in two steps:

- Model internal analysis: The central elements of the O&M conceptual model are interpreted in the DOLCE context. Model internal inconsistencies are identified and resolved.
- Model external analysis: To support semantics-based information discovery, the concepts specified in a conceptual model should identify precisely the set of intended real world entities. Thus, overly general meanings of the central O&M terms are identified and remodeled in order to restrict the scope of interpretation.

The work presented continues the work of turning the O&M model into an ontology-based representation format¹ and is a step towards an ontology-based semantic reference system [5, 6] which enables information providers to semantically annotate their O&M-conformant information models in a consistent and reproducible manner.

Throughout the article, the following formatting is used. Categories and feature types start with an upper-case letter and are written in *italics*; concepts start with a lower-case letter and are written in *italics*. When we refer to a term or notion or to an individual entity no additional formatting is used. For example, the category (or feature type) *Observation* is conceptually based on the concept *observation* which provides semantics to the term observation.

The remainder of the paper is structured as follows. Section 2 gives an introduction to DOLCE and the process of ontology alignment. In section 3, we introduce the central O&M terms, analyze the definitions with respect to internal conceptual inconsistencies and suggest alignments to DOLCE categories. Section 4 combines the results of chapter 3 into a first cut of a DOLCE-aligned ontology for observations. Section 5 draws conclusions and points to future work.

Background

In this section we introduce the general process of aligning domain ontologies to a foundational ontology. This is followed by an introduction to the central categories of the foundational ontology DOLCE.

¹ See <http://seres.uni-muenster.de/> for a discussion paper and OWL-based ontologies.

Ontology Alignment Process

“Foundational ontologies are axiomatic theories of domain-independent top-level notions such as object, attribute, event, parthood, dependence, and spatio-temporal connection” [7, p.1]. Foundational ontologies provide a rigorous formal semantics for the top-level notions. They are intended to serve as conceptual foundation for domain ontologies, which in turn provide specialized views on a certain domain. The semantics of the vocabulary used in domain ontologies are specified using top-level concepts.

Approaches to ontology alignment are presented in [8-11]. The alignment process to a foundational ontology can be seen as committing to the foundational ontology’s “view” on how entities are categorized. In other words, the domain ontology engineer commits to the conceptualization underlying the foundational ontology and further refines it.

Generally, existing (informal) domain ontologies (e.g. the O&M model) were not intended to be aligned to a foundational level at the time they were developed. Thus, a careful analysis of the domain concepts is needed before an alignment can be established. In Fig. 1, three possible alignment situations are depicted and explained in the following.

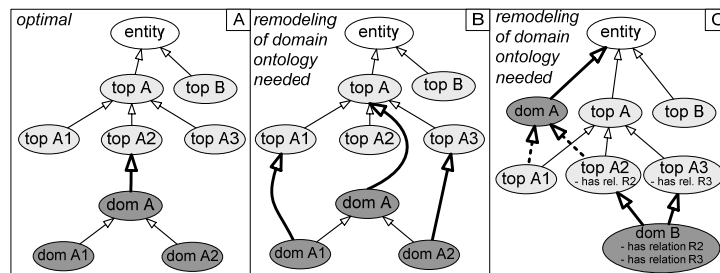


Fig. 1. Three general situations can be encountered when an existing domain ontology is intended to be aligned to a foundational ontology. Thick arrows indicate the *kind-of* relation between a domain concept (dom) and foundational concept (top). Remodeling of domain concepts is required in situations B and C in order to achieve a consistent alignment (situation A).

- A: The most general concepts of the domain ontology are aligned to the most specific concepts of the foundational ontology. This is the optimal alignment situation, and should be the result of an alignment process. It is unlikely that this situation is encountered without previous remodeling steps.
- B: Taxonomic relations of the domain ontology overlap with taxonomic relations of the foundational ontology. The domain ontology itself provides concepts which are more general than the foundational ontology’s most specific concepts. Domain concepts might be removed to attain situation A.
- C: A domain ontology concept might act as superconcept of several foundational ontology concepts (see: dom A in Fig. 1C). Since this generalization is not considered in the foundational ontology it might indicate an overgeneralization in the domain ontology. Another suboptimal situation arises if a domain concept is speci-

fied to be the subconcept of two top-level concepts. This implies that the sets of characteristics (relations to other concepts) of two or more top-level concepts are all assigned to a single domain concept. In both cases a consistent alignment is impossible and thus a remodeling is required.

The Foundational Ontology DOLCE

Currently only a few full-fledged foundational ontologies are available. Among those are the General Formal Ontology (GFO) [12], the Object-Centered High-level Reference Ontology (OCHRE), [7]), OpenCyc [13], the Suggested Upper Merged Ontology (SUMO) [14] and the Basic Formal Ontology (BFO) [4, 15]. For this investigation, we have chosen the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) [4] as foundational ontology². The central categories of DOLCE are introduced in the following and depicted in Fig. 2. This introduction is based mainly on [4, 16]. The main purpose of DOLCE is the negotiation of meaning. It enables to sort the particulars (also called entities) encountered in the domain of interest into categories. Categories are understood as cognitive artifacts, which are more or less dependent on human perception, cultural imprints and social conventions [4]. We follow Sloman [17, p.192] stating that “[c]oncepts and categories are, to a large extent, flip sides of the same coin. Roughly speaking, a concept is the idea that characterizes a set, or category, of objects”. In this sense, the categories in DOLCE reflect concepts and the concepts in turn provide the meaning to terms used in information systems. We are aware that the use of category in this sense has shortcomings as discussed in [18-20].

Endurants and Perdurants. Endurants and perdurants belong to the four top categories of DOLCE. Perdurants embrace entities generally classified as events, processes, phenomena, activities and states. Endurants can be divided into physical and non-physical endurants. Only physical endurants have physical and thus spatial qualities.

The difference between endurants and perdurants is related to their behavior in time. Endurants are wholly present at any time they exist. Their parts move with them in time. Perdurants, on the contrary, extend in time by accumulating different temporal parts. At any time that they are present, they are only partially present in the sense that some of their proper parts are not present anymore or are not yet present [16]. An endurant “lives” in time by participating in some perdurant(s). For example, a tree (endurant) participates in its lifespan (perdurant).

² Additionally some concepts are taken from the DOLCE extension DOLCE2.1-Lite-Plus (DLP) <http://www.loa-cnr.it/DOLCE.html>

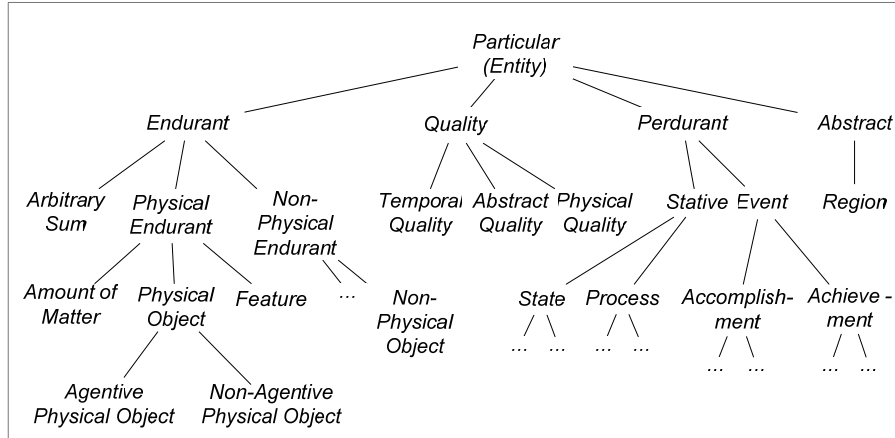


Fig. 2. The basic categories in DOLCE (reduced, for more details see [4]).

Qualities, Qualia and Quality Region. Qualities are seen as the basic entities we can perceive or measure, for example the volume of a certain lake, the color of a certain rose, or the length of a certain street [4]. The main characteristics of qualities are that they are observable and that they are inherent in other entities. Every (geospatial) entity comes with certain qualities, which exist as long as the entity exists. The (geospatial) entity itself can not be observed but only its qualities and the qualities of its parts. For example, we do not perceive an apple but its (indirect) qualities like weight, shape, color, taste or volume.

In O&M the notion property refers to what is called quality in DOLCE. However, DOLCE puts strong emphasis on the distinction between quality and property. A quality is a particular which is understood as individual entity. For the philosophical argument we refer to [4].

DOLCE distinguishes between physical, temporal and abstract qualities. Physical endurants can have only physical qualities. Non-physical endurants can have only abstract qualities and perdurants can have only temporal qualities. DOLCE defines a strict distinction between a quality, e.g., the depth of a specific lake, and its value (also called quale) which can be approximated with a depth measure, e.g. 10 m. To bring the above into a geospatial context: Geospatial entities are either conceptualized as perdurants or endurants. Endurants and perdurants are characterized through their qualities. Endurants and perdurants have unique qualities which are entities by themselves. Only endurants have spatial locations, whereas perdurants have indirect spatial locations via the endurants participating in them. Each quality of an entity has a quale which is represented as an atomic quality region. All regions at which qualities of a certain type can be located form together the quality space of that quality type. Quality regions, quality spaces and the values of qualities (qualia) are abstract entities.

Aligning the Central O&M Terms to DOLCE

This chapter introduces the notions feature, feature of interest, phenomena, event, procedure and instrument which are the notions employed by the O&M model to specify the central notion observation. Aligning these notions to DOLCE yields the following benefits:

- The central terms of the O&M model inherit the formal semantics provided by the foundational ontology. (DOLCE is fully axiomatized in the modal logic S5 plus the Barcan Formula [21].) When different GI information sources inherit the same top-level semantics, the benefit emerges that further refined terms become comparable.
- Internal ontological inconsistencies are identified when the elements of the O&M conceptual model are interpreted in the DOLCE context. In information encoding specifications such as the O&M, a model element (e.g. a feature type) may have multiple relations to other elements. This model element represents a set of real world entities. A model element is inconsistent, if the set of represented entities needs to be changed in order to make sense of all the relations that the model element entertains (see section *Feature of Interest* and section *Procedure and Instrument*). If the conceptual model is mainly based on natural language, this inconsistency may not be obvious, since most natural language terms change the set of entities they denote dependent on the context in which they are used.
- Overly general definitions in the current O&M model are identified. A model may be internally consistent, yet the concepts are too general to support efficient semantics-based discovery of information sources.

On the left side of figures 4-7, the original O&M definitions are interpreted in terms of DOLCE. On the right side suggestions are given on how to arrive at an optimal alignment situation (see Fig. 1A).

The O&M categories of each figure are depicted in light grey. DOLCE concepts and DOLCE Lite Plus concepts are depicted in dark grey. Black open arrows represent taxonomic relations (kind-of). Black filled arrows represent non-taxonomic relations. Thick arrows indicate the alignment between domain and foundational ontology.

Feature

Feature is one of the core notions in all OGC specifications. It is remarkable that the definition given in most of the OGC standards' terminology section (a feature is an abstraction of real world phenomena) appears to be understood in two different senses.

- *Sense 1*: A feature can be any entity that exists in physical and social reality, for example the town you live in or your favorite national park. Thus, the process of abstracting real world entities into categories results in the definition of feature types, for example, the feature type town. Feature types are categories to which individual features belong.

- *Sense 2*: Feature type and feature are understood as information objects. Such an information object can be used to *represent* either some category or some individual of a category.

The difference between those senses boils down that one can either understand feature as pure information object with the sole purpose of *representing* a real-world entity or understand a feature to *be* a (geospatial) real world entity.

Congruent with sense 1 is for example the section on conceptual modeling in ISO 19101 [22, p.11]. “Conceptual modeling is the process of creating an abstract description of some portion of the real world and/or a set of related concepts. As an example, a set of features such as watercourses, lakes, or islands might constitute a portion of the real world being modeled. [...] The abstract description of these *real-world features* is called a *conceptual model*. Conceptual models may exist only in the minds of people who communicate them to each other verbally and often imprecisely”.

Concepts are mental constructs according to which human agents form categories. We understand “abstract descriptions” in the quote above as concepts, if these descriptions exist in the mind of an information provider only. As soon as these “abstract descriptions” are made partially explicit via some representation language, we talk about models of concepts. Information system ontologies are concerned with providing such models of concepts. We understand concepts as providing the rules according to which we classify the entities in physical and social reality into categories. In this sense, the notion feature type is equivalent to the notion category.

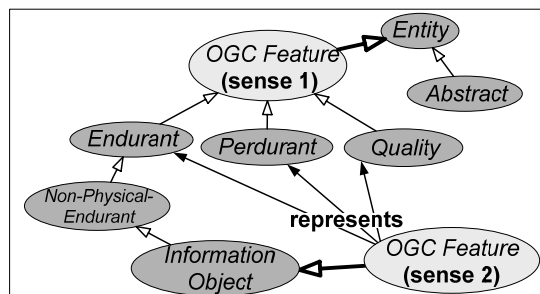


Fig. 3. The two possible interpretations of the notion feature are depicted in terms of DOLCE. The alignment situation for sense 1 corresponds to Fig. 1C. A feature in sense 2 is an information object that is used to represent perduring entities (e.g. processes) and enduring entities (e.g. physical objects). An information object itself is an enduring entity.

Congruent with sense 2 is for example ISO 19109 [23, p.8]. It states that the feature is “[a] fundamental unit of geographic information”.

In this definition, a feature is understood as information carrying unit or object. Thus, the feature’s purpose is to model and to represent some entity. The definition restricts the notion of feature to denote exclusively information objects. According to this definition, a feature type can be used to *represent* an abstraction of real world phenomena (e.g. the category *Lake*) and an individual feature can be used to *represent* some entity (e.g. Lake Constance).

Semantic problems are unavoidable if the notion feature is used within the community to denote the real world entity as well as its representations. We suggest that

in an OGC context, the meaning should be restricted to sense 2; a feature is an information object used to represent individuals of other categories.

Feature of Interest

According to the O&M specification, “the **feature of interest** is the object regarding which the observation is made, and whose value must be a feature instance” [1, p.12]. “[A] phenomenon is associated with an identifiable object, which is the feature of interest of the observation. [...] The key idea is that the observation result is an estimate of the value of some property of the feature of interest” [1, p.11].

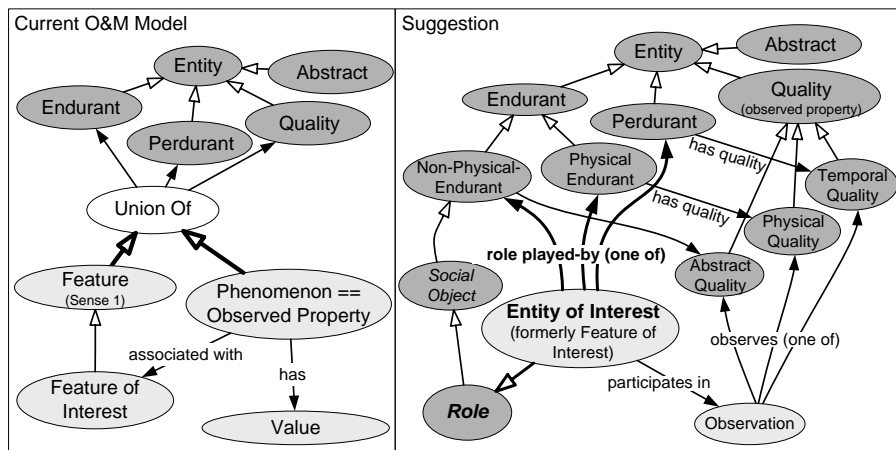


Fig. 4. Left: The model is consistent with the current definition of *Observed Property* only if *Feature of Interest* is understood in sense 1 (see Fig. 3). Right: In order to keep a consistent use of the notion feature, we suggest renaming the class *Feature of Interest* to *Entity of Interest*.

Ontological Interpretation and Suggestions. In the O&M model, a *feature of interest* is the target of an observation. This implies that *feature* is understood in sense 1 as real-world entity (see Fig. 3). Understanding *feature of interest* in sense 2, implies that *only* properties³ of information objects can be observed. For the discovery of information sources it is important to know whether the observation is performed on an individual real world entity or on an information object. For example, the width of a lake can be observed directly or it can be “observed” on a representation of a lake, e.g. in a GIS.

The O&M model is consistent only if *feature of interest* is understood in sense 1 (see Fig. 3 and Fig. 4). However, this conceptualization of *feature* leaves the model underspecified. The extension of the category *Feature* is equivalent to the extension of the category *Phenomenon* which may indicate that one is redundant. Further, the

³ In DOLCE terminology the notion quality is used instead of the notion property.

model does not specify which phenomenon can be associated with which feature of interest. For example, a subcategory of *Phenomenon* called *Speed* could be associated with a subcategory of *Feature of Interest* called *Lake*.

We suggest to emphasize the ontological distinction between a feature of interest and the entity whose properties are observed. We introduce the example of observing the depth of a lake to illustrate why this ontological distinction is useful. According to the O&M definition, the observed phenomenon (*depth*) is associated with an identifiable object (*lake*). The identifiable object (*lake*) is seen as an individual of the category *Feature of Interest*. According to [24, 25] an entity should entertain a IS-A (taxonomic) relation to a category only, if this relation holds as long as the entity exists. It is plausible that an entity can stop being the feature of interest and still exist. Therefore we suggest that *Feature of Interest* is *not* to be modeled as subcategory of *Feature* (in neither sense 1 nor 2).

DOLCE introduces the category *Role* being a kind of *Social Object* which in turn is a kind of *Non-Physical Endurant*. Any entity can entertain the non-taxonomic relation *plays-role* to individuals of the category *Role*. For example, a lake plays the role of a recreation area. Related to the O&M model, *feature of interest* is conceptualized as a *role* which any entity that has an observable property can play. In our example, a lake *plays* the role of a feature of interest only if it is the target of an observation. If it is not the target of an observation, the lake still is-a (individual of) *Lake* which is a kind-of *Physical Endurant* but the lake *is not* an individual of *Feature of Interest* anymore. The feature of interest stops to exist as soon as the observation event is finished. Modeling the semantics of the term lake as denoting an entity which stops to exist as soon as it is not the feature of interest of a certain observation anymore is not helpful for a semantics-based search environment.

The O&M model has the purpose of specifying information encodings. Thus, the classes which use the stereotype <<feature type>> are understood in sense 2. However, in the case of the class *Feature of Interest*, the O&M model implicitly uses the notion feature in sense 1. In the context of standards for information encoding, we suggest to use the notion feature only in sense 2 in order to avoid inconsistencies explained in the next section.

Phenomenon

According to the O&M specification, “the observed property is usually a phenomenon, the estimate of whose value is the primary result of the observation” [1, p.14]. “[T]he value of the observed property identifies or describes the phenomenon for which the observation result provides an estimate of its value “ [1, p.12].

Ontological Interpretation and Suggestions. The O&M model intentionally gives a rather vague definition of what an observed property is (personal communication with O&M editors). Phenomenon, the second important term in the definition is also not precisely defined. We rely on the general OGC usage of the notion “real world phenomenon”, assuming that the O&M notions phenomenon and observed property are intended to denote the same category. This would imply that any entity can be the observed property of another entity. Interpreting the O&M definition in terms of DOLCE implies that an entity belongs to the category Phenomenon if it also belongs to one of the categories Endurant, Perdurant or Quality. In the context of semantics-based discovery of information sources, the O&M definition of phenomenon does not allow to sufficiently specify a phenomenon.

The category *Observed Property* in the O&M model subsumes the category *Quality* in DOLCE. For the formal ontological distinction between property and quality we refer to [4]. In DOLCE, only qualities are entities that can be observed. In this sense, a physical endurant, a lake for example, is not observable. In turn, the sum of all observed quality values allows the cognition of physical endurants. DOLCE provides a generic categorization of qualities, restricting which qualities an entity can have. We want to stress that it is important to enable the identification of the quality *without* referring to the information encoding with which the observation result is encoded.

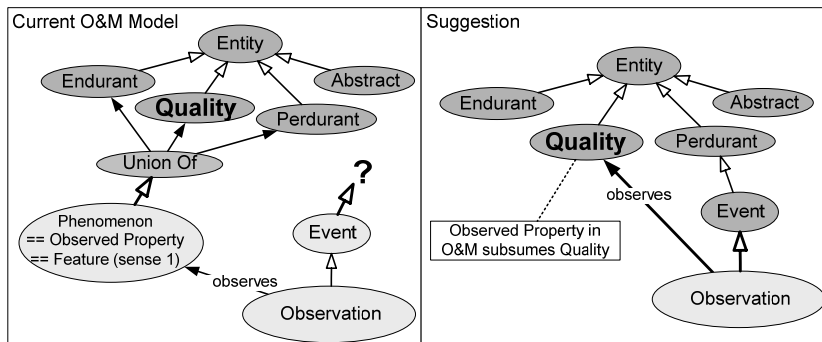


Fig. 5. Left: *Phenomenon* has the same extension as *Feature* in sense 1 (see Fig. 3). This leads to an underspecified model. The O&M category *Event* can not be aligned consistently to DOLCE. Right: According to DOLCE, *only* individual qualities are entities which can be observed.

Event

According to the O&M specification, an event “is a feature type characterized by a time whose value is a temporal object (TM_Object), a location whose value is either a spatial object described using coordinates (GM_Object) or a named place (EX_GeographicDescription). A number of persons or organizations (CI_ResponsibleParty) may be identified as responsible for the Event. “[1, p.11].

Ontological Interpretation and Suggestions. The textual definition of Event in the O&M model states that the category *Event* is a kind of feature type. Thus any individual event is a feature. This definition can be understood in two ways.

1. If *Feature* is understood in sense 2, then *Event* refers only to the specification of information objects representing real world events. This implies that the O&M model provides no definition of the underlying conceptualization of real world events. An event understood as information object has relations only to other information objects. This is reflected in the relations to *GM_Object*, *TM_Object* or *EX_GeographicDescription*, individuals of these categories are all information objects encoding time and location.
2. If *Feature* is understood in sense 1, then the semantics of the term event stay rather unspecific (see Fig. 3) since the direct supercategory of *Event* is *Feature* and almost all entities can be considered as features according to this definition.

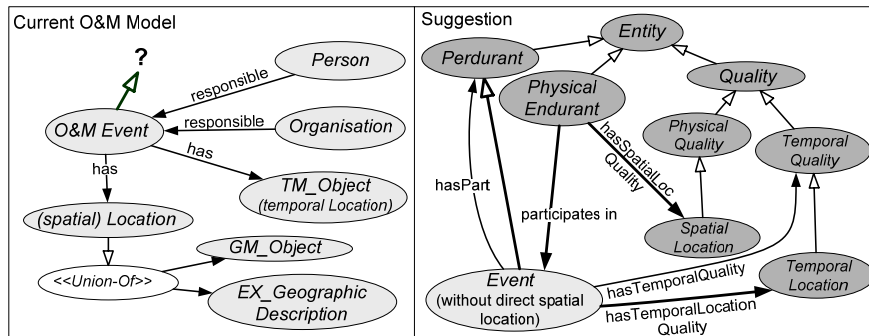


Fig. 6. Left: An event is defined to have both, a direct *spatial* and a *temporal* location. This prevents an alignment to DOLCE. Right: An event is a perduring entity, which has only temporal qualities. A physical object participates in an event and thus provides indirectly a spatial location for the event.

In the O&M model, any event has a relation to a time individual and a location individual. The conceptualization of this relation is not further specified. For the categories *Time* and *Location* no conceptualization is provided, they are defined in terms of information objects. Additionally any event can have a relation *responsible* which relates it to some person or some organization. The relation *responsible* is not further specified, too.

The current model of *Event* can not be aligned to DOLCE since an event has both, a direct spatial location and a direct temporal location (Fig. 6). This is not possible in DOLCE. In terms of DOLCE, an event is a *Perdurant*; it is an entity that happens in time (e.g. a soccer match, a departure, or an observation). An event entity is never fully present at a certain time. Take as example a particular soccer match event. The first chip pass is not present any longer when the first goal happens. In this sense, a soccer match is never completely present. Both, chip pass and goal are temporal parts of the event soccer match. Events are characterized by having only temporal qualities and by having at least one temporal location. An event can not have physical qualities. This leads to an important difference to the O&M *Event*. An event in the

DOLCE sense can not have a spatial location, since a spatial location is a physical quality. Any perdurant can have physical qualities only indirectly via a physical object which participates in an event. For example a physical instrument participating in an observation event spatially locates the observation event. Events can have other perduring entities as parts.

Separating temporal and physical location has the advantage that one can ask for the spatial location of different physical entities which participate in an event. An observation event for example has two important but indirect spatial locations; the spatial location of the entity of interest, for example urban areas, and the spatial location of the physical instrument, for example a satellite, with which a quality of the entity of interest is observed. In the search for suitable data sources, it is useful to be able to specify in which of the two spatial location one is interested.

Procedure and Instrument

According to the O&M specification, a procedure is used to generate an observation [1, p. 16]. An "observation uses a procedure, which is often an instrument or sensor but may be a process chain, human observer, a computation or simulator"[1, p. 11]. "The value of the procedure property is the description of a procedure" [1, p. 16].

Ontological Interpretation and Suggestions. The O&M textual section defines implicitly an anonymous category as direct super category of *Procedure* (see Fig. 7). This anonymous category is the union of the categories *Human Observer*, *Process Chain*, *Sensor*, *Computation*, *Simulator* and *Instrument*. In the UML diagrams, *Instrument* is specified to be the subcategory of *SimpleProcedure*. None of the other supercategories are specified in the UML diagrams of the O&M model. Circular taxonomic relations for *Instrument* result, when textual and UML-based definitions of the O&M model are viewed together (see Fig. 7). This is probably not intended.

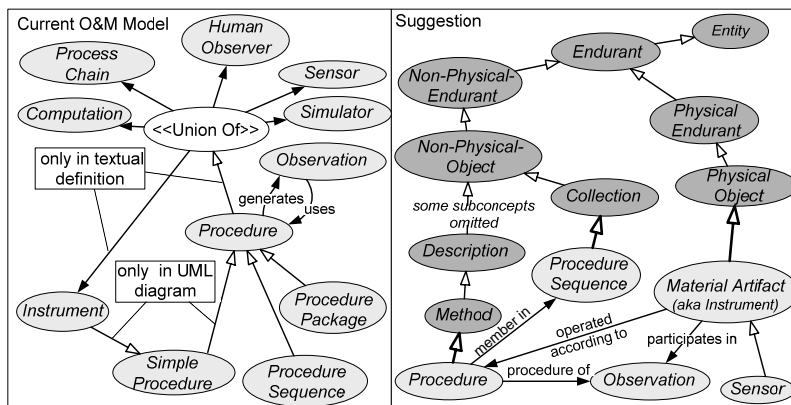


Fig. 7. Left: Textual definitions and UML-based definitions result in circular definitions. We suggest to define *Procedure* as a kind of *Method* according to which a physical instrument is involved in an observation.

According to [1, p. 16], the “value of the procedure property is the description of a procedure”. According to this definition individuals of *Procedure* are descriptions. DOLCE provides a category *Description* which in turn has a subcategory *Method*. Descriptions and methods are non-physical, social objects. We consider *Method* as suitable supercategory for *Procedure*. A method is a description according to which a process or event can be performed. However, this would be inconsistent with the current O&M understanding that a procedure (actively) *generates* an observation. We suggest to replace the *generates*-relation by a newly introduced relation *procedure of*, which we specify as subrelation of the DOLCE relation *method of*.

If *Instrument* is understood as subcategory of *Physical Entity*, then the definition of *Procedure* as subcategory of *Method* causes an inconsistency with *Instrument*. Physical entities and methods are ontologically distinct categories and should not be confused. However, a physical instrument does operate according to a procedure or method. We suggest to remodel *Instrument* as *MaterialArtifact* which is a subcategory of *PhysicalObject* and as supercategory of *Sensor* and *Simulator* or any other newly introduced category accounting for measurement or observation devices. In the O&M model, *Instrument* has an attribute *SerialNumber*, indicating that instruments in O&M are understood as physical entities and not as procedures or methods.

Analogously to *Instrument*, the individuals of *HumanObserver* do act according to a method or a procedure but should not be modeled as supercategory of *Procedure*. According to DOLCE, *HumanObserver* can be seen as a kind of *Role* which is played by an individual of the category *Person*.

According to DOLCE, a simulation or a computation is a process which in turn is a perduring entity. Since observations, seen as perduring entities, can have other perduring entities as parts, we suggest to model simulation and computation processes to be temporal parts of observation events. Similar to observations, simulations and computations can be described by methods or procedures, yet they are no methods or procedures.

To summarize the above, simulations and computations are individuals of the category *Processes*, sensors and simulators are individuals of the category *Material Artifacts* and human observers can be seen as individuals of the category *Role* some person can play. Unifying the above categories into one anonymous super category of *Procedure* (see also Fig. 7) hinders a precise semantic annotation of observation encoding models drastically.

Another inconsistency in the O&M model is that currently the category *Station* is modeled as subcategory of *Feature of Interest*. This modeling decision is partly inconsistent with the textual descriptions in the O&M specification. On page 11 it is stated that “[t]he phenomenon is associated with an identifiable object, which is the feature of interest of the observation”. For example, a thermometer observes the phenomenon air temperature. In this case, the surrounding air mass plays the feature of interest, not the station at which the thermometer is hosted. We suggest that *Station* never to be modeled as feature of interest. The *hosts* association which relates *Station* and *Procedure* is inconsistent with the new and partly with the old definition of *Procedure*. A station understood as physical entity which is not a physical instrument does not follow nor does it host a procedure or method. We suggest to model stations to *host* physical instruments, which in turn operate according to a procedure.

2. The process of assigning a symbol (number, term, complex information item) to the approximation entity identified in the previous process. In order to communicate which approximation entity was identified in the previous step, symbols are needed.

A procedure describes an observation and thus provides the context needed to interpret the observation results. However, care should be taken when the procedure or the employed instrument is used to characterize the quality. For example, it seems odd to define the ontological nature of the quality temperature by stating that a temperature is a temperature because it is observed by a thermometer.

The categories Value, Quale and Abstract are not further investigated in this article and are subject to future work.

Result and Conclusion

Geospatial information sources can be discovered successfully if they are based on two explicitly specified types of conceptual models:

1. A conceptual model of the domain of interest for which representations are supplied. This includes specifications of the entities of interest, the observed qualities of these entities as well the procedures applied to achieve observation results which can be communicated.
2. A conceptual model of the information encoding in which the representations of the real world entities are supplied.

The O&M implementation specification provides only the second type explicitly. However, only if one knows which kinds of real world entities are represented, it is helpful to know how they are represented. In this paper we presented the formal and explicit basis of the missing first type of conceptual model. It is specified as an ontology. This ontology reflects as closely as possible the implicit O&M conceptual model and is at the same time ontologically consistent with the foundational ontology DOLCE.

The remodeling involved to

- restrict *Feature* to be a subcategory of DOLCE:*Information Object*.
- restrict *Feature of Interest* to be a subcategory of DOLCE:Role
- turn *Feature of Interest* into Entity of Interest.
- restrict *Observable Property* to be equivalent to DOLCE:*Quality*.

The ontological analysis of O&M showed that the notion of feature needs to be interpreted in two senses in order to keep the model consistent. This however leads to a model where the categories *Feature* and *Phenomenon* are equivalent on the ontological level leaving one notion redundant. In the current O&M model, the category *Procedure* is modeled with circular taxonomic relations. Remodeling it as subcategory of *Method* and adding non-taxonomic relations between the categories *Procedure* and the categories *Instrument*, *Sensor* and *Simulator* resolved this inconsistency.

The current O&M model relies on the human user to understand the meaning of the textual descriptions and the terms used in the UML descriptions. The presented

ontology relates these terms to the rigorous formal semantics provided by the foundational ontology DOLCE.

Smith and Mark [18] claim that one benefit of developing and applying ontology in geospatial application is that “geographic information systems need to manipulate representations of geographic entities, and ontological study of the corresponding [categories], especially those at the basic level, will provide default characteristics for such systems”. The results presented in this article are practical examples supporting this claim. The above mentioned benefit can be further specified: The risk of misinterpretation of O&M-conformant information sources is drastically restricted if they commit to the presented ontology or a specialization of it. The discovery of geospatial information sources is facilitated since reasoning about entities of interest and their potential observable qualities as well as the procedures employed to observe them becomes possible. A further benefit is that the occurrence of inconsistencies introduced unintentionally by the developer is decreased since automatic consistency checking becomes available.

Much work has been done to produce foundational ontologies like DOLCE as well as to produce domain specifications like the O&M model. A combination of these efforts by undergoing the alignment process is highly beneficial for current and future challenges in information discovery and retrieval.

Future Work

The attempt to align observation and measurement concepts to the foundational level has produced further questions regarding the modeling decisions taken in DOLCE. Further investigations are required on the concept *quality*. Especially for information sources dealing with observations and measurements, a precise specification of the observed qualities of the entities of interest is important. Questions regarding the combination of qualities forming compound or complex qualities need to be solved. Individuals of which quality categories can be combined and in which category of entity can the different qualities be inherent? In this context a detailed analysis on how to further combine the theory of conceptual spaces [26] with the quality spaces provided in DOLCE is required. We envision a theory of semantic reference spaces and regions which allow to combine the ontologically important structures of quality spaces with spaces serving the solely purpose of approximating the generic value (quale) of a quality, thus allowing to specify the semantics of unit of measure, scale, reference datum and other conventionally agreed structures needed to perform observations and measurements. We consider this as the basis for methods enabling semantic translation between terms of different domain ontologies.

With respect to practical application, the task of annotating geospatial information sources needs to be facilitated. A framework is required guiding and assisting the information provider through the annotation process, by suggesting and explaining the formal structure of the ontology.

Acknowledgements

Discussions with Michael Lutz, Eva Klien, Claudio Masolo, Stefano Borgo and Werner Kuhn have greatly influenced and improved the ideas presented here. The work has been supported by the German Research Foundation (DFG) through the Semantic Reference Systems Project (grant KU 1368/4-1).

References

1. OGC: Observations and Measurements. Open GeoSpatial Consortium, OpenGIS Recommendation Paper OGC 05-087r1 (2005)
2. Guizzardi, G., Herre, H., Wagner, G.: On the General Ontological Foundations of Conceptual Modeling. In: Proc. 1st International Conference on Ontologies, Databases and Applications of Semantics (ODBASE 2002) (2002)
3. Guizzardi, G.: Ontological Foundations for Structural Conceptual Models. Enschede, Netherlands (2005)
4. Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A. WonderWeb Deliverable D18, Ontology Library (final). Available: <http://wonderweb.semanticweb.org/deliverables/documents/D18.pdf> (2003)
5. Kuhn, W.: Semantic Reference Systems. International Journal of Geographical Information Science 17 (2003) 405-409
6. Kuhn, W.: Geospatial Semantics: Why, of What, and How? Journal on Data Semantics (2005) 1-24
7. Schneider, L.: Designing Foundational Ontology - The Object-Centered High-level Reference Ontology OCHRE as a Case Study. In: Proc. Conceptual Modeling - ER 2003, 22nd International Conference on Conceptual Modeling (2003) 91-104.
8. Gangemi, A., Guarino, N., Masolo, C., Oltramari, A.: Sweetening WordNet with DOLCE. AI Magazine (2003)
9. Oberle, D., Mika, P., Gangemi, A., Sabou, M.: Foundations for service ontologies: Aligning OWL-S to DOLCE. In: Proc. World Wide Web Conference (WWW2004), Semantic Web Track (2004)
10. Noy, F. N., Musen, M. A.: SMART: Automated Support for Ontology Merging and Alignment. In: Proc. 12th Workshop on Knowledge Acquisition, Modelling and Management (KAW'99) (1999)
11. Borgo, S., Leitão, P.: The Role of Foundational Ontologies in Manufacturing Domain Applications. In: Proc. OTM Confederated International Conferences, ODBASE 2004 (2004) 670-688
12. Degen, W., B., H., H., H., Smith, B.: GOL: Towards an axiomatized upper level ontology. In: Proc. 2nd International Conference of Formal Ontologies and Information Systems (FOIS 01) (2001)
13. Lenat, D. B., Reed, S. L.: Mapping Ontologies into Cyc. In: Proc. AAAI workshop on Ontologies and the Semantic Web (2002)
14. Niles, I., Pease, A.: Toward a Standard Upper Ontology. In: Proc. 2nd International Conference of Formal Ontologies and Information Systems (FOIS 01) (2001) 2-9
15. Grenon, P., Smith, B.: SNAP and SPAN: Towards Dynamic Spatial Ontology. Spatial Cognition and Computation 4 (2004) 69-104

16. Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., Schneider, L.: Sweetening Ontologies with DOLCE. In: Proc. International Conference on Knowledge Engineering and Knowledge Management. AAAI (2002)
17. Sloman, S. A., Love, B. C., W.-K., A.: Feature centrality and conceptual coherence. *Cognitive Science* 22 (1998)
18. Smith, B., Mark, D.: Ontology and Geographic Kinds. In: Proc. 8th Int. Symposium on Spatial Data Handling, SDH'98 (1998) 308-320
19. Lakoff, G.: *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*. The University of Chicago Press, Chicago (1987)
20. Rosch, E.: Principles of Categorization. In: E. Rosch and B. Lloyd, (eds.): *Cognition and Categorization*. Lawrence Erlbaum Associates (1978) 27-48
21. Hughes, G. E., Cresswell, M. J.: *A New Introduction to Modal Logic*. Routledge, London (1996)
22. ISO/TC-211, OGC: 19101 Geographic information - Reference model. ISO TC 211, Draft International Standard (2002)
23. ISO/TC-211, OGC: Text for DIS 19109 Geographic information - Rules for application schema Vs. 2.0. Draft Version. ISO, (2001)
24. Guarino, N.: The Role of Identity Conditions in Ontology Design.
25. Welty, C., Guarino, N.: Supporting Ontological Analysis of Taxonomic Relationships. *Data and Knowledge Engineering* 39 (2001) 51 - 74
26. Gärdenfors, P.: *Conceptual Spaces - The Geometry of Thought*. Bradford Books, MIT Press, Cambridge, MA (2000)