Separating Ontological and Informational Concerns:

Towards a Two-Level Model-Driven Approach

Roberto Carraretto and João Paulo A. Almeida
Ontology and Conceptual Modeling Research Group (NEMO),
Federal University of Espírito Santo (UFES)
Vitória, ES, Brazil
rc.carraretto@gmail.com, jpalmeida@ieee.org

Abstract— **Many authors have discussed the importance of ontological concerns in the development of information systems, emphasizing the benefits of ontology-based approaches to conceptual modeling tasks. A principled ontology-driven approach typically relies on the definition of a domain ontology and its use in subsequent phases of information system design and integration. Many of the challenges in the application of such an approach are related to addressing ontological concerns (defining the nature of phenomena of interest) and addressing information modeling concerns (defining the information demand about the phenomena of interest). In this paper we argue that ontological concerns should be clearly separated from information modeling concerns. We have observed that ontology-based approaches and information modeling approaches have been treated mostly in isolation, with the consequence that the relation between a domain ontology and an information model is still in need of clarification, despite the efforts of the formal ontology and information modeling communities. We strive to provide the initial steps for a principled two-level approach in which a domain ontology addressing ontological concerns is used as a starting point for the definition of an information model, according to a certain information demand. We identify some important decisions that should guide the construction of information models.**

Ontology-driven conceptual modeling, ontology-driven information systems design, information modeling.

I. Introduction

There has been significant interest in the role of formal ontologies to support conceptual modeling. The argumentation put forward by many authors is that ontological concerns support the understanding and communication about a domain of interest, in an effort that is, in principle, largely independent of information systems development.

In this view, an *ontological level* for conceptual modeling [1] aims at building agreement on metaphysical aspects of a domain of interest, capturing the categories of being which are assumed to exist in that domain. The ultimate artifact of the ontological level is a domain ontology, i.e., a conceptual specification about a domain, independent of epistemic states and states of affairs [2].

The importance of addressing ontological concerns in conceptual modeling is clearly defended by Guarino in [1], where he argues that, when conceptual modeling languages take into account formal distinctions at the ontological level, the potential misunderstandings and inconsistencies in conceptual models are reduced, thus improving the quality of conceptual models. Since conceptual models are often used as a basis for the construction and integration of information systems, the production of principled ontology-based conceptual models should affect positively the development of information systems. In the course of such a development trajectory, the role of developers is to bridge the gap between ontology-based conceptual models and an implementation of the information system.

A significant portion of this gap involves addressing what we call here *informational concerns*, which focus on the information about the phenomena of interest. Information modeling concerns arise as information systems (and agents alike) are not omniscient and obtain most information through observation (measurement and perception) as well as communication (or interaction). In addition, agents are susceptible to the flaws of knowing, thus potentially obtain and hold incorrect, inaccurate and incomplete information. Further, the capacity to hold and manipulate information is not unlimited, and thus information may be stored when deemed relevant and discarded when deemed irrelevant.

We believe that ontological concerns that shape the ontological level should be clearly separated from the concerns that arise from addressing information demands. This should provide a starting point for a principled two-level approach for conceptual modeling, in which the ontological level addresses what types of entities and relations are considered to *exist* in a domain of interest, while the *information level* addresses, at type level, all which a system may record, infer and communicate about the entities of the domain under consideration. Every system may have particular ways of dealing with the various aspects of information handling, e.g., acquirement, communication, relevance, and accuracy. Thus, the strict separation would beneficially affect the generality (and thus reusability) of conceptual models at the ontological level, while addressing unavoidable information modeling concerns at the information level.

We have observed that ontology-based approaches and information modeling approaches have been treated in isolation, with the consequence that the relation between a domain ontology and an information model is still in need of clarification, despite the efforts of the formal ontology (e.g., [1][2]) and information modeling communities (e.g., [3][4][5]).

This paper aims at addressing this relation by identifying the information modeling concerns that arise from differences in information demand and from the various means to perceive or measure properties, identify/name/refer to entities, and keep track of history and time. We strive to provide the initial steps for a principled two-level approach in which a domain ontology addressing ontological concerns is used as a starting point for the definition of an information model, according to a certain information demand.

This paper is further structured as follows: section 2 discusses some important features of the information level; section 3 presents what we consider here as the core elements of the ontological level; section 4 discusses how information modeling decisions can be used to guide the definition of an information model based on the elements of the ontological level; section 5 discusses related work; and, finally, section 6 provides concluding remarks.

II. Features of the Information Level

The information level concerns the information demand about the phenomena of interest, i.e., what kinds of information may be stored and exchanged by agents. Our usage of the term "agent" extends the traditional reference to human and artificial intelligence agents, including information systems as well. We do such on the premise that information systems manipulate data with underlying semantic content, i.e., data is interpreted in order to establish some correspondence to propositions about phenomena of interest. Hence, by "information" we mean factual semantic content or "propositions" [6]. We assume agents somehow carry propositions about the world (e.g. "George Edward Moore wrote The Nature of Judgment"). In the case of human and artificial intelligence agents, those propositions are the objects of belief. In the case of information systems, those propositions are the semantic content encoded in data (i.e., information). We adopt a sense of "proposition" akin to Moore [7] and Frege [8] (the latter uses the term "thought"). That is to say, propositions are considered to be immaterial entities that are the meaning of sentences and the objects of belief. Additionally, we consider them to be the primary bearers of truth, i.e., propositions are the entities that are ultimately true or false, which means sentences and beliefs only possess an indirect truth value. The notion of truth enables us to account for flaws of knowing: to make a mistake is akin to believe a false proposition or to store data with false semantic content.

An important characteristic of the information level is that propositions about the same phenomena of interest can be structured in different manners. For example, the propositions "John studied in a university in the past" and "John studied in Stanford between 2006 to 2010" are true by virtue of a single *enrollment* phenomenon, which is the truthmaker of the two different propositions. Likewise, the propositions "John is married to Mary" and "John is married" are true by virtue of a particular *marriage* phenomenon. We consider that this difference in *structure* (i.e., types and relations underlying propositions) occurs because every agent has a particular *information demand* that is used to accomplish a strategic purpose. It is the goal of the information level to define one among the many possible structures of propositions (information) about a domain and to characterize a certain information demand. In order to characterize the information demand, the agent has to make *informational decisions*.

In the following, we briefly characterize the informational decisions that will be discussed in section 4.

In terms of observation and communication, we identify informational decisions concerning *reference*. More specifically, at the information level, one must formalize how agents refer to the perceived entities, through symbols that we call *identifiers*. For instance, agents may refer to people using different types of identifiers such as names, national identification numbers or arbitrary internal codes. Moreover, agents may be concerned about aspects related to the origin of identifiers, e.g., when an identifier was attributed to a certain individual and who made such attribution.

We also identify informational decisions concerning the *measurement* of properties, which are related to considering the measuring instrument, the abstract data types adopted (unit dimensions, granularity) and qualitative aspects of the obtained result. As an illustration, there are multiple ways of measuring height, e.g., using a ruler with centimeter precision or a measuring tape with 1/32 inch precision. Besides that, agents may be concerned about aspects related to the origin of the measured values, e.g., when the value was obtained, who performed the measurement, what were the environmental conditions.

At the information level, concerns about what is relevant to remember arise. Thus, we identify informational decisions involving *history tracking*. For example, consider the domain of organizations and employees. For human resource management, one may be concerned with both current and past employments; for pay roll, one may be interested exclusively on current employments. In another example, consider the domain of persons and their weights. If one is interested in information for controlling weight loss, one may be interested in the history of values assumed by a person's weight; in the context of a boxing championship, one may simply be interested in the latest weight value measured in a predefined point in time; in the context of drug administration, one may simply be interested in the current weight value in order to calculate appropriate dosage.

Finally, we identify informational decisions regarding *time tracking*. Timing aspects of things (such as start time, end time and duration) are often implicit in domain ontologies and are addressed by the theories of the ontological level. Time tracking decisions determine whether agents require to know about the timing of processes and lifecycles of entities in the domain. Consider, for example, a domain about allocation of resources. When allocation is treated as a (non-atomic) event in a domain ontology, allocations will possess underlying aspects such as start time, end time and duration (as a direct consequence of being considered non-atomic events). Nonetheless, which of those aspects are relevant to be known depends on an information demand. Some agents may be interested when resources were allocated and deallocated (i.e., interest in both start time and end time), while others may be interested in how long the allocations lasted (i.e., sole interest in the duration).

The ultimate artifact of the information level is the so-called *information model*, which many refer to as conceptual schema, usually expressed in ER diagrams, UML class diagrams and ORM fact-based models.

An example of information model is presented in Figure 1. The figure shows an ORM model addressing the information demand of knowing the history of patient weights, where for each patient at most one weight measurement is performed per day. All the four aforementioned informational concerns are addressed in this model. Concerning reference, the model determines that each patient is identified by a patient number (see the reference mode "(.nr)" inside the Patient entity type rounded rectangle). Concerning measurement, the model determines that Weight is to be measured in Kilograms. Concerning history tracking, the model includes an entity type "WeightMeasurement" in order to store measurement records. Concerning time tracking, the model explicitly represents an interest on dates of weight measurements. Also, it defines that those dates are represented in year-month-day format, specifying thus a particular granularity of time (a measurement concern). Figures 2 and 3 depict, respectively, the corresponding UML model and ER diagram for the same information demand.

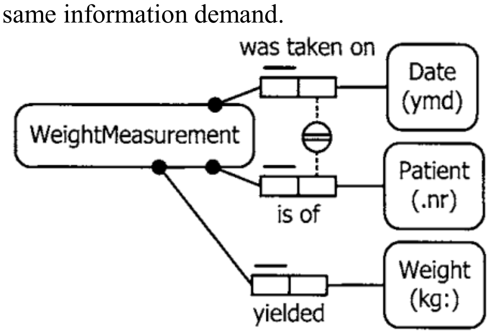

Figure 1. Informational concerns addressed in ORM (from [5])

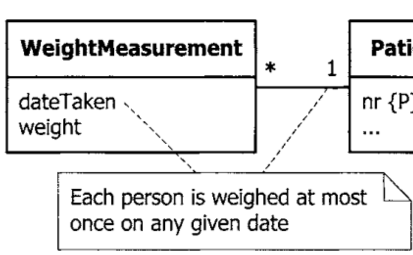

Figure 2. Informational concerns addressed in UML (from [5])

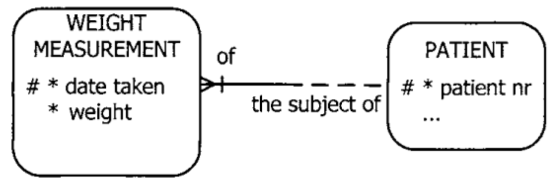

Figure 3. Informational concerns addressed in ER (from [5])

The UML and ER versions are slightly different from the ORM one w.r.t. structure and constraint specification. In UML and ER, Date and Weight are represented as attributes instead of entity types. Also, in the UML model, the constraint that guarantees that each patient's weight is measured only once per day has to be declared apart.

III. The Ontological Level

Differently from the information level, the ontological level concerns the phenomena of interest, addressing the categories of being which are assumed to exist in a certain domain independently of particular information demands.

In order to characterize the ontological level and to help us articulate about informational decisions, we adopt here some concepts from a foundational ontology, namely, the Unified Foundational Ontology (UFO) [2]. UFO is a domain-independent system of categories dealing with formal aspects of objects, addressing ontological aspects such as identity and unity, types and instantiation, rigidity, mereology and so forth. It has been developed from a combination of the GFO (Generalized Formal Ontology) underlying GOL (General Ontology Language) [9] and the Ontology of Universals underlying OntoClean [10]. We describe here only the top-level concepts which are relevant to the scope of this paper.

The first fundamental distinction adopted is that between universals and individuals. ***Universals*** are patterns of features that can be realized in a number of different ***individuals***.

Individuals can be distinguished in terms of their behavior w.r.t. time. ***Endurants*** are said to be wholly present whenever they are present, i.e., they are in time (e.g., a person, the brightness of the Sun). Perdurants (henceforth, ***events***) are individuals composed of temporal parts, i.e., they happen in time (e.g., a birth, a marriage, an enrollment). Endurants can be divided into substantials and moments. ***Substantials*** are existentially independent individuals (e.g., a person, a forest, a lump of clay). ***Moments***, in contrast, are individuals that can only exist in other individuals, and thus they are existentially dependent on other individuals (e.g., a table's height, a person's headache, a covalent bond between atoms).

Here, we are particularly interested in two types of moments, namely, qualities and qua individuals. ***Qualities*** are moments that depend on a single individual and "can be seen as the basic entities we can perceive or measure" [11] (e.g., a person's weight, an apple's color). A ***qua individual*** is the moment that bears all moments of a thing (say, John) that share the same dependencies and the same foundation (e.g., John qua student, John qua husband). In other words, a qua individual is the way an object participates in a certain relation.

For every category of individual discussed so far, UFO assumes a corresponding category of universal. For instance, substantial universal is the category of universal whose individuals are substantial individuals. The first distinction among universals is based on the notion of a *principle of identity* which supports the judgment whether two individuals are the same (i.e., in which circumstances the identity relation holds). Substantial universals that *carry* a principle of identity for the individuals they collect are called ***sortal universals*** (e.g., Apple). ***Mixin universals***, on the other hand, are substantial universals that represent an abstraction of properties that are common to multiple disjoint types (e.g., Red Thing) and thus do not carry a principle of identity.

The specialization of sortal universals is based on a meta-property called *rigidity*. An example of rigid universal would be Person, since instances of Person cannot cease to be so without ceasing to exist. Conversely, Child is an example of anti-rigid universal, since for every instance of Child it is possible to imagine a world in which that instance no longer instantiates the universal but continues to exist, e.g., as a Teenager or an Adult. ***Kind universals*** are rigid sortal universals that *provide* a principle of identity for its individuals (e.g., Person). ***Role universals*** are anti-rigid sortal universals that represent participations in events and depend on extrinsic (relational) properties (e.g., Student participating an Enrollment along with University, Husband and Wife participating a Marriage).

The categories of individuals and universals that were previously discussed are represented in Figure 4 (the terms 'universal' and 'individual' have been abbreviated). The so-called domain ontology is a specification about universals (not individuals) following the rules of a foundational ontology.

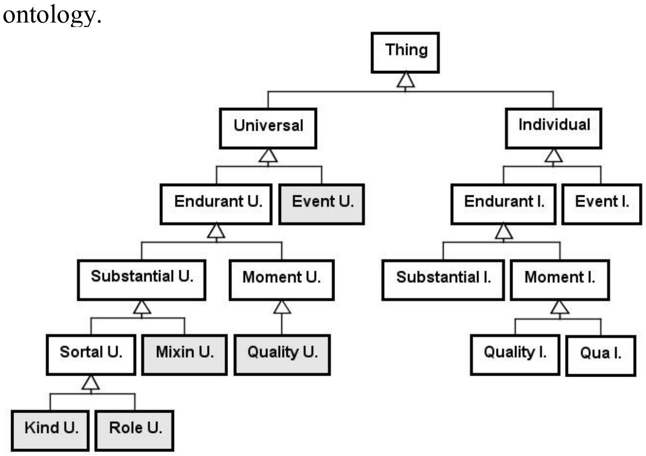

Figure 4. Categories of universals and individuals

IV. Examining Informational Decisions

In this section, we discuss some of the most prominent informational decisions we have identified so far. We discuss informational concerns independently of a particular language for the information level. In order to establish the relation between the entities in the domain ontology and the decisions concerning information on these entities, we explicitly relate these decisions with the elements of the domain ontology.

A. Measurement

We start with measurement concerns, which arise from the fact that information on *qualities* may be obtained in different ways with different instruments according to particular information demands.

According to the DOLCE foundational ontology, ***qualities*** "can be seen as the basic entities we can perceive or measure" and they inhere in specific individuals ("no two [individuals] can have the same quality") [11]. As a consequence DOLCE (and UFO alike) distinguishes "between a quality (e.g., color of a specific rose), and its 'value' (e.g., a particular shade of red)" [11]. The latter is called ***quale***, and describes the position of an individual quality within a certain ***quality dimension***. So "when we say that two roses have (exactly) the same color their *two* colors have the same position in the color space (they have the same *color quale*), but still the two roses have numerically distinct color qualities" [11]. Each quality dimension is "endowed with certain geometrical structures" and is supposed "to satisfy certain structural constraints" [13]. For example, for the weight quality, there is a quality dimension "which is one-dimensional with a zero point and thus isomorphic to the half-line of nonnegative numbers" [13]. As another example, "our cognitive representation of colors can be described by three dimensions: hue, chromaticness, and brightness" [13] (those dimensions form the geometrical structure of the color spindle). Finally, "there is, in general, no unique way of choosing a dimension to represent a particular quality but a wide array of possibilities" [13].

Information on the quale (value) of a quality is obtained through a *measurement event* involving a *measuring instrument*. Measurement events determine attributes of the obtained information such as its accuracy, precision and granularity. Accuracy (i.e., being near to the true value) and precision (i.e., being reproducible) reflect our incapability to perfectly grasp reality through instruments. For instance, the accuracy of the information on a quale (value) is just as accurate as the measuring instrument can be, along with other circumstances of the measurement event. Granularity (i.e., the level of detail) may reflect ignorance about further details of a certain quality or a scope limitation that corresponds to an information demand. For example, a coarse-grained granularity for weight may be the result of the instrument's granularity, or an information demand of having values with coarse detail.

Considering the variety of measurement aspects, there are several informational decisions for each ***quality universal*** in the domain ontology. These decisions relate to the information about the measurement event, which include the aforementioned attributes (e.g., accuracy, precision and granularity) and information on the participants of the measurement event (e.g., the measuring instrument, the measurer). Information on the measurement event corresponds to the so-called meta-attributes in the data quality literature [12].

As an illustration, consider a domain ontology specifying a kind universal Person, characterized by a weight quality universal that is related to a quality dimension isomorphic to the half-line of nonnegative numbers. At the information level, there may be different information demands concerning weight values. In the context of experimental physics, one may be interested in various sorts of information about weight measurement. For a certain weight value, one may be interested in which physicist performed the measurement and which weight scale was used, along with experimental uncertainties (e.g., "John obtained a value of 60,000±0,003 kg for Mary's weight using the weight scale XYZ123"). On the other hand, in the context of gym classes in a high school, one may simply need to record weight values without knowing any further detail about the measurement event (e.g., "Mary's weight is 60 kg"). At the information level, there is freedom to choose the data structure representing a certain quality dimension, as long as there is a correspondence between the data structure and the geometrical structure of the quality dimension. For example, weight values might be encoded, e.g., in real numbers (e.g., for experimental physics), in integer numbers (e.g., for high school gym classes) and in lexical spaces (e.g., "thin", "average", "overweight" in a survey).

B. *Reference*

Reference is a topic in philosophy concerned with the relation between "expressions and what speakers use expressions to talk about" [1]. Here, we are particularly interested in reference to ***individuals*** (not universals or other sorts of things). As an illustration, this type of reference is the main topic of Strawson's "On Referring" [14]: "we very commonly use expressions of certain kinds to mention or refer to some individual person or single object or particular event or place or process (…)". According to Strawson, the classes of expressions that are most commonly used for reference are: singular demonstrative pronouns ("this" and "that"); proper names (e.g., "Venice", "Napoleon", "John"); singular personal and impersonal pronouns ("he", "she", "I", "you", "it"); and phrases beginning with the definite article followed by a noun, qualified or unqualified, in the singular (e.g., "the table", "the old man", "the king of France"). We are especially interested in proper names, but also in things such as national identification number, GS1 country code and MAC address. The latter were not discussed in philosophy but are alike proper names and occur frequently in conceptual modeling. As a consequence, we call all those sorts of things (including proper names) *identifiers*.

We argue that identifiers should not be treated undiscerningly with the attributes that are the subject of measurement. First of all, identifiers exist for a very specific purpose, namely, denotation. This is similar to the view advocated by [15] for proper names: "Proper names are not connotative: they denote the individuals who are called by them; but they do not indicate or imply any attributes as belonging to those individuals". Secondly, we could say an identifier is *attributed* to a thing in a *baptism ceremony* involving one or more baptizers (the term "baptism ceremony" was taken from Kripke [16]). Finally, there is no measurement event or measuring instrument to obtain (the value of) an identifier and the symbols that constitute an identifier always remain unchangeable, as opposed to the quale (value) assumed by a quality.

As a language issue, reference has to be explicitly taken into account at the information level. Agents may agree on ontological aspects (categories of being) of a domain, but refer to individuals in a different way. For instance, a country may be referred to via a language specific name (e.g., "Brasil"), an ISO 3166 country code (e.g., "BR"), a GS1 country code (e.g., 789) and so forth. Those are simply agent-specific ways of referring to (presumably) the same country individual. In the following, we use the categories of the ontological level to discuss what sorts of things can be referred to by an identifier.

First, identifiers can be attributed to ***substantial individuals***, e.g., social security numbers for US citizens, "Allies" and "Axis" for the (non-extensional) rival groups of the World War II, numbers for blood samples in a laboratory, numbers for bank accounts. In this case, they refer to instances of ***kind universals***.

Furthermore, identifiers can also be attributed to ***event individuals***, e.g., names for important historical events ("D-Day", "Prise de la Bastille"), numbers for medical exams, sequential numbers for sub-events (laps of a race). In these circumstances, they refer to instances of ***event universals***.

Other identifiers, such as student enrollment number, driver's license number and passenger id, refer to ***qua individuals***. For example, a student enrollment number is an identifier specifically created to refer to the many attributes of a person while playing the role of student in certain relation. If John studies in two universities at the same time and he has two different student enrollment numbers for each, then his particular behavior in each university can be tracked by each identifier. For instance, if John is enrolled in Stanford and in Harvard, one identifier may refer to "John qua student of Stanford" (to keep track of his grades and class attendance records in Stanford) and the other to "John qua student of Harvard" (to keep track of his grades and class attendance records in Harvard). Ergo, we say those identifiers refer to qua individuals inhering in instances of ***role universals*** [17].

The same type of identifier can be bound to individuals with different principles of identity. For example, items in a pet store that have different principles of identity (e.g., cats, cages and bird feed) may share the same type of identifier. As another illustration, a sales agent may identify customers (people and companies) using the same type of identifier. That is to say, the type of identifier may have a correspondence not to the ***kind universal*** that the individuals instantiate (e.g., Person, Company), but rather to the ***mixin universal*** common to all individuals (e.g., Customer).

As a conclusion, identifiers may refer to individuals that are instances of kind, event, role and mixin universals. In a domain ontology, for each universal pertaining to one of

[1] http://plato.stanford.edu/entries/reference

those categories, it is an informational decision whether to attach an identifier to it. Informational decisions may also concern the structure of each identifier type (defining symbols, syntax, check digits, etc.) in order to address certain pragmatic issues of information manipulation. Lastly, we also consider informational decisions concerning the baptism ceremony. There may be cases where it is helpful to explicitly represent the ceremony when one has to know specific details about it. For example, one may need to know the issuing date of national identification numbers (i.e., baptism ceremony date) and which governmental organization issued them (i.e., baptizer). There are also cases where the simple representation of the identifier will suffice.

Some information modeling languages deal explicitly with reference. For instance, ORM includes the so-called “reference schemes”, addressing aspects such as compound reference, disjunctive reference, context dependent reference, preferred identifiers, rigid identifiers, information bearing identifiers, etc. [5]. Such sophisticated support indicates that the variety of informational decisions involved in dealing with reference should be treated separately from the ontological level.

C. History Tracking

The world constantly changes: events start and stop happening, substantials begin and cease to exist. Things that are currently happening or existing may not be the only targets of information demand. In many circumstances, one is interested in things from the past, i.e., substantials that ceased to exist and events that happened. We consider the choice of which sorts of things are relevant to remember to be a decision related to an information demand.

For example, consider a domain ontology that specifies enrollments of students in universities as events. At the information level, systems may have different information demands with respect to the history of enrollments. For instance, the Ministry of Education may need to track all enrollments (current and past) while a University Library may simply need to keep the current ones.

We now use the categories of the ontological level to support the discussion of what sorts of things can be target of history tracking. ***Events***, particularly, play a central role in this discussion. We consider that, for every ***event universal*** in the domain ontology (e.g., Enrollment, Marriage, Flight, Business Process), one has a decision about which events are relevant to keep track of. This decision is two-fold: one has to decide if the past events are relevant and also decide if the current events are relevant. As a consequence, one may be interested in: (i) exclusively past events (or a selection thereof), (ii) exclusively current events or (iii) both sorts of events. For instance, an air traffic controller may only be interested in current flights, while an airline administrator may be interested in the history of flights.

Substantials are also related to history tracking, since they begin and may cease to exist at some point in time. In this case, we consider that history tracking decisions target every ***kind universal*** in a domain ontology (e.g. Person, Organization). Similarly to the case of events, a history tracking decision for a kind universal is two-fold and may result in tracking: (i) exclusively the past, i.e., the substantials that have ceased to exist, (ii) exclusively the present, i.e., the currently existing substantials, or (iii) both the present and the past, i.e., substantials that have ceased to exist and those that exist in the present.

Determining when a substantial begins and ceases to exist is a competence of the ontological level and is captured in the principle of identity of a kind universal. Therefore, one should be aware of the definition of existence adopted for a certain substantial universal before committing to a history tracking decision. Consider for example the information demand of a cemetery which is interested exclusively on maintaining records of the deceased. Varying the principle of identity associated with Person in the domain ontology affects directly the history tracking decision concerning persons. If the domain ontology associates the concept of Person to the notion of living organism, people cease to exist when they die (viz. they stop being in space-time and only their body, which is a different concept, remains). In this case, one is interested solely on the past (persons that no longer exist according to the ontology). However, if the domain ontology adopts a principle of identity for the concept of Person that is based on a theological or spiritualistic perspective, then people never cease to exist (viz. when people die, they become some sort of everlasting disembodied souls). In this case, one is interested in a portion of the existing people (namely, those that have passed away).

For ***qualities***, history tracking can be done in terms of their assumed qualia (values). Alternatively, this could be seen as the history tracking of *measurement events*. For each quality, it is a decision whether to keep only its current quale value (the last *measurement event*) or a history of values. For example, consider a car dashboard where one can see the various values assumed by qualities of a car, e.g., fuel level, speed, coolant temperature and odometer distance. An on-board diagnostic system may be interested in the current values of all those qualities but only some of them may actually require history tracking.

History tracking has been treated by some approaches purely at the information level. For instance, [5] provides modeling guidelines in ORM when one needs to maintain full or partial history of things. For that, it suggests modelers should introduce additional objects for history tracking (cf. the “WeightMeasurement” object in Figure 1). Since these are ordinary objects, history tracking is not explicitly separated from the things being tracked, which is the case in the two-level approach discussed here. Furthermore, information modeling approaches do not discuss what ontological categories of things can be the target of history tracking (events, substantials, qualities).

D. Time Tracking

Time tracking concerns the interest on timing aspects of things, i.e., properties corresponding to the time dimension. Here, we relate timing aspects to ***events*** and we focus on three *basic timing aspects*, namely: (i) start time of events (e.g., date of admission in a job), (ii) end time of events (e.g., graduation year), and (iii) duration of events (e.g., duration of a phone call).

We consider the basic timing aspects (viz. start time, end time and duration) to be built-in features of the ontological level, since by definition all non-atomic events possess them. Thus, for every event universal in the domain ontology, one has the informational decision of tracking the relevant basic timing aspects.

For instance, consider a domain ontology specifying enrollments as events. One system may be interested exclusively on the start time of enrollments (e.g., to check for current students that are about to exceed the maximum duration of a course). Another system may be interested exclusively on the duration of past events (e.g., to know how long past students took to graduate). Besides that, a system could be interested in none of the basic timing aspects (e.g., a librarian could be interested if someone is enrolled or not, but unconcerned about any timing aspect of enrollments).

Another built-in feature of the ontological level refers to the lifecycle of each endurant being encompassed by an implicit *event of existence*. Such feature is acknowledged in the Basic Formal Ontology (BFO) foundational ontology: "One type of relation is of special importance (…) this is the relation between each [endurant] and that unique [event] which is its *life*" [18] (terminology has been adapted). Ergo, for every endurant universal in the domain ontology, one has the informational decision of tracking timing aspects of the underlying event of existence.

Although we have dealt here only with basic timing aspects, there are other timing aspects that may also be relevant to information modeling. For example, one may be interested in the frequency of events (e.g., how many accidents have occurred in the last year), prospective information (e.g., when certain events are expected to occur) and derived information (e.g., age).

E. Summary

We believe the value of this paper lies on clarifying the distinction between modeling phenomena of interest and addressing an information demand. To the best of our knowledge, this broad issue is not explicitly addressed in the literature. A great portion of this effort is achieved through the identification of informational concerns and of their relations with ontological aspects. A summary of these relations is presented in Table I.

TABLE I. SUMMARY OF INFORMATIONAL CONCERNS AND CORRESPONDING ONTOLOGICAL ASPECTS

Informational Concern	**Ontological Aspect**	**Informational Decision**
Measurement	Quality Universals	Required information on measurement events (e.g., time, instrument, measurer, meta-attributes).
Reference	Kinds, Event Universals, Roles, Mixin Univerals	Required information on the baptism ceremony (e.g., time, baptizer) and on the identifiers used.
History Tracking	Kinds, Event Universals, Quality Universals	Required information on past, present or both.
Time Tracking	Kinds, Event Universals	Required information on start time, end time, duration, etc.

V. RELATED WORK

In a seminal paper [1], Guarino revisits a classification of the various primitives used by knowledge representation systems, originally provided by Brachman [19]. Brachman argued that primitives could be grouped into four levels (namely, implementational, logical, conceptual and linguistic) – each level corresponding to an explicit set of primitives offered to the knowledge engineer. At the logical level, primitives are propositions, predicates, logical functions and operators, which are extremely general and ontologically neutral. For instance, one may use predicates in arbitrary ways, to represent "a property of a thing, the kind the thing belongs to, a role played by the thing, among other possibilities" (e.g., both Apple and Red are admissible predicates) [2].

To improve the "flatness" of the logical languages, Brachman proposed the introduction of an *epistemological level* between the logical and conceptual levels. According to Guizzardi [2], the rationale behind the design of languages for the epistemological level is the following: (i) the languages should be designed to capture interrelations between pieces of knowledge that cannot be smoothly captured in logical languages; (ii) they should offer structuring mechanisms that facilitate understanding and maintenance, they should also allow for economy in representation, and have a greater computational efficiency than their logical counterparts; (iii) finally, modeling primitives in these languages should represent structural connections in our knowledge needed to justify conceptual inferences in a way that is independent of the meaning of concepts themselves. Guizzardi points as examples of epistemological level languages Brachman's KL-ONE and its derivatives (including the semantic web languages such as OWL) as well as object-based and frame-based modeling languages (such as EER and UML).

On top of the epistemological level, Guarino introduced an *ontological level*. While the epistemological level is the level of structure, the ontological level is the level of *meaning*. At the ontological level, knowledge primitives satisfy formal meaning postulates, which restrict the interpretation of a logical theory on the basis of formal ontology, intended as a theory of *a priori distinctions*: (i) among entities of the world (physical objects, events, processes…), (ii) among the meta-level categories used to model the world (concepts, properties, states, roles, attributes, various kinds of part-of relations…) [1].

Foremost, we agree with the main *purposes* of both levels (meaning and structure) as discussed by Guarino and Guizzardi. Since the epistemological level focus on structure, rather than meaning, it resembles the information level discussed here. While Guarino and Guizzardi have focused mostly on defending the ontological level (i.e., defending the use of ontological theories at an ontological level in conceptual modeling), we focus on informational decisions that can be treated separately, but in conformance with the ontological level. We have observed that little attention has been given in the literature to how a domain ontology will be used to derive information models, so we address informational decisions referring to elements of the ontological level.

With respect to approaches that divide conceptual modeling in two similar levels, it is worth to mention a work by Ashenhurst [4]. In Ashenhurst's approach, information modeling is regarded as having two interacting realms, namely, the objective and the subjective realms, which resemble the ontological and information levels discussed here, respectively. According to him: "Objective representation involves the notions of being and belonging (what exists, and how related), while subjective representation is concerned with referring and inferring (reference to what exists and how related, and inference of additional beings and belongings)" [4]. Further on, he stresses this idea: "much of the information used in practice has only an indirect objective basis. For example, the average value of some attribute of interest over a set of objects is a derived quantity not corresponding to any object itself (…)" [4]. In this regard the author also defends certain relations between the levels, he says that "a 'direct' aspect of the subjective representation can be based on 'presentations' of the objective representation (…)" [4], although he does not identify explicit relations or decisions required to establishing these relations.

Also in line with a two-level approach, Halpin and Morgan [5] stress the difference between the "real world" and the "recorded world". Conceptual schemas written in ORM represent the "recorded world" and their constraints should be interpreted "as applying to the database, not necessarily to the real world" [5]. Because the "real world" and the "recorded world" share some correspondence, the constraints in ORM conceptual schemas "should be at least as strong as those that apply in the real world" [5]. For instance, things that are necessary in the "real world" (e.g., ORM mandatory roles) may be specified as optional information in the "recorded world" (e.g., ORM optional roles), due to our ignorance (or scope) about the domain. Those aspects enforce the information modeling nature of ORM.

Finally, many of the informational concerns we have discussed here are addressed in the scope of information modeling, and thus, previous work in this area supports us in characterizing the information level. For instance, ORM deals with reference and measurement (through "reference schemes"), derived attributes and, in addition, [5] provides guidelines for history and time tracking. Nonetheless, languages such as ER, UML and ORM are not ontologically grounded. Moreover, they usually deal undiscerningly with both ontological concerns (viz. determining what types of things exists in the domain) and informational concerns (viz. determining what is relevant to know about those types). Most of the time, ontological and informational concerns are blurred in modeling approaches using those languages, and informational concerns are not properly defined in terms of ontological aspects.

VI. Conclusions and Future Work

This paper has contributed to clarifying the role of the ontological and the information levels in information systems development. By establishing relations between the ontological and the information levels, we have aimed at leveraging the benefits of ontological level distinctions, while addressing the unavoidable informational decisions of information systems development.

In terms of methodology, the separation into levels can be further elaborated to provide a systematic way to develop an information model based on a domain ontology (in a top-down approach) or to develop a domain ontology that represents the underlying semantics of an information model (in a bottom-up approach). In the top-down approach, the advantage of the separation into levels is that an unbiased domain ontology is likely to be shared by a larger community, which may agree on ontological concerns but may disagree on information demands. This can be the basis for the establishment of semantic interoperability; in this case, the domain ontology should be considered as a reference model for interoperation.

We have discussed a number of informational decisions, which entail several aspects of information handling and are related to the information demand of an agent (or a community of agents). We have shown that informational decisions are far from trivial, and deserve separate treatment, as they involve observation and measurement events, attribution of identifiers, considerations regarding the limited capacity for handling information, etc.

We have not aimed at an exhaustive set of informational decisions and concerns, and many relevant informational decisions will be reported in our future work. These include those decisions that arise from inferred information (which may be obtained deductively and inductively with different consequences), from the limited, incomplete knowledge about phenomena of interest, and from the scope of interest of agents in the phenomena represented at the ontological level.

For future work, we also consider a more prescriptive method for constructing an information model using a domain ontology as a starting point. This should result in the specification of model transformation rules in a model-driven approach. This model transformation should be highly parameterizable to enable designers to capture the results of informational decisions. Another area of investigation concerns the definition of data implementation models and data interchange formats in conformance with information models. This would result in a full model-driven approach; the use of conformance relations [20] between the various levels would support us in leveraging the benefits of the ontological level to the whole information systems development lifecycle.

Acknowledgments

This research is funded by the Brazilian Research Funding Agencies FAPES (grant number 52272362/2011) and CNPq (grants number 483383/2010-4 and 310634/2011-3). Roberto Carraretto is funded by FAPES. We would like to thank Giancarlo Guizzardi for his comments on earlier drafts of this paper. Any errors or omissions remain the responsibility of the authors alone.

REFERENCES

[1] N. Guarino, The Ontological Level, Philosophy and the Cognitive Science, Hölder-Pichler-Tempsky, Vienna, 1994.

[2] G. Guizzardi, Ontological Foundations for Structural Conceptual Models, Telematica Institute Fundamental Research Series, The Netherlands, 2005.

[3] W. Kent, Data and Reality, 2nd ed., 1st Books Library, 2000.

[4] R.L. Ashenhurst, Ontological Aspects of Information Modelling, Minds and Machines, Vol. 6, No. 3, Springer, 1996, pp. 287-394.

[5] T. Halpin and T. Morgan, Information Modeling and Relational Databases, Morgan Kaufmann, 2008.

[6] L. Floridi, Information: A Very Short Introduction, Oxford University Press, 2010.

[7] G.E. Moore, The Nature of Judgement, Mind, New Series, Vol. 8, No. 30, Oxford University Press, 1899, pp. 176-193.

[8] G. Frege, The Thought, Mind, New Series, Vol. 65, No. 259, Oxford University Press, 1956, pp. 289-311.

[9] B. Heller and H. Herre, Ontological Categories in GOL, Axiomathes, Vol. 14, No. 1, Kluwer Academic Publishers, 2004, pp. 57-76.

[10] N. Guarino and C. Welty, Evaluating ontological decisions with OntoClean, Communications of the ACM 45, 2002, pp. 61-65.

[11] A. Gangemi, N. Guarino, C. Masolo, A. Oltramari and L. Schneider, Sweetening Ontologies with DOLCE, Knowledge Engineering and Knowledge Management Ontologies and the Semantic Web, Lecture Notes in Artificial Inteligence, Vol. 2473, Springer, 2002, pp. 166-181.

[12] Y. Wand and R.Y. Wang, Anchoring data quality dimensions in ontological foundations, Communications of the ACM, Vol. 39, No. 11, ACM, 1996, pp. 86-95.

[13] P. Gärdenfors, Conceptual Spaces: The Geometry of Thought, MIT Press, 2000.

[14] P.F. Strawson, On Referring, Mind, New Series, Vol. 59, No. 235, Oxford University Press, 1950, pp. 320-344.

[15] J.S. Mill, A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation, Harper & Brothers, 1843.

[16] S.A. Kripke, Naming and Necessity, Harvard University Press, 1980.

[17] R. Wieringa, W. De Jonge and P. Spruit, Using dynamic classes and role classes to model object migration, Theory and Practice of Object Systems, Vol. 1, No. 1, John Wiley & Sons, 1995, pp. 61-83.

[18] P. Grenon, B. Smith and L. Goldberg, Biodynamic ontology: applying the BFO in the biomedical domain, Ontologies in Medicine, IOS Press, 2004, pp. 20-38.

[19] R.J. Brachman, On the Epistemological Status of Semantic Networks, Associative Networks Representation and Use of Knowledge by Computers, Academic Press, 1979, pp. 3-50.

[20] J.P.A. Almeida, M. Iacob and P. Eck, Requirements traceability in model-driven development: Applying model and transformation conformance, Information Systems Frontiers, Vol. 9, Springer, 2007, pp. 327-342.