On the Application of Ontological Patterns for Conceptual Modeling in Multidimensional Models

Glenda Amaral[0000−0003−0460−2271] and
Giancarlo Guizzardi[0000−0002−3452−553X]

Free University of Bozen-Bolzano, Bolzano BZ 39100, Italy
{gmouraamaral,giancarlo.guizzardi}@unibz.it

Abstract. Data warehouses (DW) play a decisive role in providing analytical information for decision making. Multidimensional modeling is a special approach to modeling data, considered the foundation for building data warehouses. With the explosive growth in the amount of heterogeneous data (most of which external to the organization) in the latest years, the DW has been impacted by the need to interoperate and deal with the complexity of this new type of information, such as big data, data lakes and cognitive computing platforms, becoming evident the need to improve the semantic expressiveness of the DW. Research has shown that ontological theories can play a fundamental role in improving the quality of conceptual models, reinforcing their potential to support semantic interoperability in its various manifestations. In this paper we propose the application of ontological patterns, grounded in the Unified Foundational Ontology (UFO), for conceptual modeling in multidimensional models, in order to improve the semantic expressiveness of the models used to represent analytical data in a DW.

Keywords: Multidimensional Modeling · Data Warehouse · Conceptual Modeling · Ontological Patterns.

1 Introduction

Multidimensional modeling is the foundation for building data warehouses (DW). Data warehouses were initially designed to support business intelligence applications in the internal context of the organization. In the latest years, the explosion in the volume of data on the web and in social networks, together with the accumulation of data generated by mobile devices, sensors and other semi-structured and unstructured data sources brought a challenge to the traditional analysis model, based on the DW, giving rise to the need of an approach that is suited to deal with the complexity of this new type of information, such as big data, data lakes and cognitive computing platforms. In this scenario, the necessity of integrating the Data Warehouse with new heterogeneous sources of information (most of which external to the organization) emerges. In addition, with the open data phenomena, the data stored in the DW was made available outside the organization, becoming evident the need to make explicit the meaning of the information disclosed. In the light of the above, there is the need to improve the semantic expressiveness of the multidimensional models used to represent DW analytical data, making explicit the worldview to which they are committing (i.e., their ontological commitments), thus providing *intra-worldview consistency* and *inter-worldview interoperability*. In this paper, we move towards addressing this issue by means of the application of *ontological patterns* for conceptual modeling in the design of multidimensional models.

Conceptual modeling is the activity of formally describing some aspects of the physical and social world for the purposes of understanding and communication [23]. It plays a fundamental role, helping us to understand, elaborate, negotiate and precisely represent subtle distinctions in our multiple conceptualizations of reality. The discipline of conceptual modeling is supported by a wide range of methods and tools for representing the conceptualization of subject domains of interest. In this paper, we focus on a set of conceptual modeling techniques, which can be applied to address recurrent multidimensional modeling issues and to improve the semantic expressiveness of multidimensional models. This set includes three techniques related to the notion of ontological patterns that are grounded in the Unified Foundational Ontology (UFO)[11], namely, *Foundational Ontology Patterns* [29], *Reification and Truthmaking Patterns* [10] and the *Powertype Pattern* [3].

UFO is an axiomatic formal theory based on theories from Formal Ontology in Philosophy, Philosophical Logics, Cognitive Psychology and Linguistics. For an in-depth discussion, empirical support and formalization see [11, 15]. UFO is the theoretical basis of OntoUML, a language for ontology-driven conceptual modeling that has been successfully employed in several projects in different domains [14]. A recent study shows that UFO is the second-most used foundational ontology in conceptual modeling and the one with the fastest adoption rate [33].

Several approaches have been proposed to multidimensional modeling in the conceptual level, either as extensions to the Entity-Relationship model [6, 30], as extensions to UML [1, 21], or ad hoc models [8, 17]. The past decade has seen an increasing interest in ontology-driven approaches for multidimensional modeling, which led to a number of research initiatives in this area, most of them using domain ontologies for representing shared conceptualizations [32, 25, 18, 31, 28, 27, 16]. Different from other approaches that use domain ontologies to provide more semantics to the information stored in the data warehouse, we have focused in this paper on improving the semantic expressiveness of multidimensional models by applying ontological patterns in their design.

The remainder of this paper is organized as follows. In Section 2, we give a brief review on multidimensional modeling and introduce the reader to the main notions on ontological patterns. Section 3 presents our approach for applying ontological patterns in the design of multidimensional models. In Section 4, to validate and demonstrate the contribution of our approach, we apply it to model a case study on education, extracted from [20]. We finalize the paper in Section 5 with some final considerations and directions.

2 Multidimensional Modeling and Ontology Patterns

2.1 Multidimensional Modeling

Multidimensional modeling is the process of modeling data in a universe of discourse, under the multidimensional paradigm. This is widely accepted as the preferred technique for modeling analytic data [20].

Multidimensional models categorize data either as facts with associated measures, which correspond to events occurred in the business domain, or as dimensions that characterize the facts and are mostly textual [26]. For example, in financial sector payment systems, money is transferred between financial institutions in certain amounts and at certain times. A typical fact would be a payment. Typical measures would be the debited and the credited amounts. Typical dimensions would be the debited financial institution, the credited financial institution, the currency and the time of the money transfer. Queries aggregate measure values over ranges of dimension values to produce results, such as the total value credited per financial institution, per month.

Traditionally, a cube metaphor is used to represent the multidimensional data view. The cells of the data cube contain the measures describing the fact. The axes of the cube, called dimensions, represent different ways of analyzing the data [2]. Classification hierarchies containing levels are used for the structuring of dimensions. A hierarchy level contains a distinct set of members and different levels correspond to different data granularities. Another orthogonal way of structuring dimensions from a users point of view is the use of dimension level attributes. These attributes describe dimension level members but do not define hierarchies (e.g. the name and address of a financial institution).

Multidimensional models implemented in relational databases are referred to as star schemas because of their resemblance to a star-like structure [19]. Basically, the star schema represents each dimension as a dimension table and each fact as a fact table with a many-to-many relationship with all the dimensions. Fig. 1 shows an example of a star schema. In this particular schema, the fact is the Payment table. Measures are the non-foreign keys in the PAYMENT fact table (e.g. amount). Dimensions (Time, Credited Financial Institution, Debited Financial Institution and Currency) are all the tables connected to the fact table in a one-to-many relation-ship. Note that in this example the Financial Institution is referenced multiple times in the fact table, with each reference linking to a logically distinct role for this dimension (Credited and Debited Financial Institution), what is commonly referred to as role-playing dimension [19].

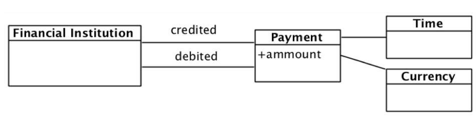


Fig. 1: Star Schema Payments

Although these schemas provide some level of modeling abstraction that is understandable to the user, they are not proper conceptual models in the sense of [13], given that they assume an underlying relational model implementation choice and contain further decisions that are proper of a physical design phase.

2.2 Ontological Patterns as Tools for Conceptual Modeling

Foundational Ontology Patterns

Foundational Ontology Patterns are reusable fragments of foundational ontologies. As foundational ontologies span across many fields and model the very basic and general concepts and relations that make up the world, Foundational Ontology Patterns can be applied in any domain [2]. They are reused *by analogy*, i.e., by establishing a structural correspondence (or structural transfer) between the structure of the pattern and the one of the problem at hand. In this article, we focus on the use of some of the Foundational Ontology Patterns that constitute the OntoUML Pattern Grammar [29].

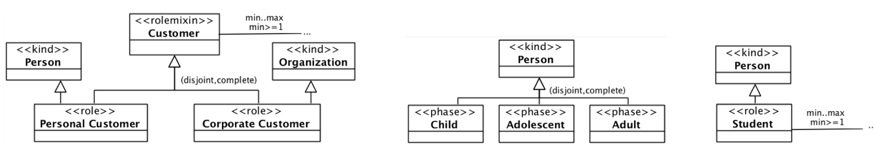
Over the past decade, a number of Foundation Ontology Patterns have been derived from UFO, using OntoUML as a pattern language. Given the objectives of this paper, we focus here on four examples extracted from [29], selected for their applicability in the scope of multidimensional modeling: the *RoleMixin*, the *Phase*, the *Role* and the *Collective* Patterns. For a detailed description of these and other OntoUML Patterns, one should refer to [29].

The RoleMixin Pattern has been extracted from UFO's theory of sortal universals and addresses the problem of specifying *roles with multiple disjoint allowed types* [11].

UFO makes a fundamental distinction between Sortal and Non-Sortal types. A sortal is a type that either provides or carries a uniform principle of identity for its instances. A principle of identity supports the judgment whether two individuals are the same or, as a special case, what changes an individual can undergo and still be the same. A Kind is a sortal that is rigid, meaning that all its instances cannot cease to be so without ceasing to exist. In contrast with rigidity is the notion of anti-rigidity that characterizes a type whose instances can move in and out of its extension without altering their identity. A Role is a sortal, anti-rigid and relationally dependent type. Therefore, every Role in UFO must be connected to an association representing this relational dependence condition. Moreover, the association end connected to the depended type in this relation must have a minimum cardinality ≥ 1.

A RoleMixin is an anti-rigid and relationally dependent non-sortal that aggregates properties that are common to different Roles. Different from Roles, RoleMixins classify entities that instantiate different kinds (and that obey different principles of identity). Fig. 2(a) shows an example the RoleMixin Pattern. In this picture, the abstract class CUSTOMER is the RoleMixin that covers different Role types. Classes PERSONAL CUSTOMER and CORPORATE CUSTOMER are the disjoint subclasses of Customer that can have direct instances, representing the Roles (i.e., sortal, anti-rigid and relationally dependent types) that carry the principles of identity that govern the individuals that fall in their extension. Classes Person and Organization are the ultimate Kinds that supply the principles of identity carried by Personal Customer and Corporate Customer, respectively.

The Phase Pattern consists of a *phase partition*, i.e., a disjoint and complete set of two or more complementary phases that specialize the same sortal and that are associated with the same *dividing principle* (e.g., gender, life status, developmental state). Phases in UFO are relationally independent, anti-rigid types, defined as a partition of a sortal. This partition is derived based on an intrinsic property of that sortal (e.g., Child is a phase of Person, instantiated by instances of person who are less than 12 years). Fig. 2(b) presents an instance of the Phase Pattern. In this picture, class Person is the sortal and classes Child, Adolescent and Adult represent the different phases that specialize this sortal. The sortal instances can move in and out of the extension of the phases, due to a change in the intrinsic properties of these instances. Analogous, in the Role Pattern we have one or more roles that specialize a sortal (Fig. 2(c)).



(a) RoleMixin (b) Phase (c) Role

Fig. 2: Foundational Ontology Patterns

The Collective Pattern, exemplified in Fig. 3, describes a Collective Universal and the universals whose instances are members of these collectives. The unity principle of collectives is a uniform relationship (i.e., a relation instance) that holds between all parts and only those parts [12]. Because of the uniformity of this relationship, the collective has a uniform structure, i.e., all its members are undifferentiated with respect to (w.r.t.) the whole. In other words, they can be said to play the same role w.r.t. the whole. Take for example collectives such as a crowd or a forest with their corresponding instances of the *member of* relation (i.e., person-crowd, treeforest). In all of these cases, the wholes have a uniform structure provided by a uniform unity principle (e.g., a crowd is a collective of persons all which are positioned in a particular topologically self-connected spatial location) and their parts are all considered to play the same role w.r.t. the whole (e.g., all persons are equally considered to be *members Of* the crowd).

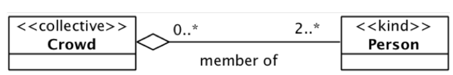


Fig. 3: Collective Pattern

Reification and Truthmaking Patterns

Reification is a standard technique in conceptual modeling, which consists of including in the domain of discourse entities that may otherwise be hidden or implicit [10]. Classic examples are the reification of relationships [10, 9, 24, 4] and events [7, 5]. Recent work on formal ontology suggests that entities that should be put in the domain of discourse are those responsible for the (alleged) truth of our propositions. These are called truthmakers [22].

In [10], the authors propose a systematic analysis of truthmaking patterns (TMP) for properties and relations, based on the ontological nature of their truthmakers (TM) and present a number of Truthmaking Patterns for properties and relations at different levels of expressivity. In this paper we focus on two Truthmaking Patterns proposed in [24], which are more relevant in the context of multidimensional modeling.

The first one is the TMP proposed for intrinsic descriptive properties. Regarding the concept of intrinsic property, [10] states that a property holding for x is extrinsic iff it requires the existence of something else external to x in order to hold, and intrinsic otherwise. As for descriptive property, [10] defines that a property P is descriptive iff, for every x, P(x) holds in virtue of (at least) a quality q being existentially dependent on x.

The second TMP considered here was proposed in [10] for descriptive relations. Analogously to the case of descriptive properties, a descriptive relation is defined as a relation that holds in virtue of some qualities that are existentially dependent on one or both its relata. Following is a brief description of these two TMP, extracted from [10]. For a formal definition of them, as well as for additional TMP not mentioned here, the reader should refer to [10].

Before proceeding, there is an important notion that should be defined, namely the distinction between strong and weak truthmakers. In the strong version of truthmakers t is a truthmaker of the sentence ϕ if the existence of t is sufficient to make ϕ true. By contrast, t is a weak truthmaker of ϕ if it makes the proposition true not just because of its existence, but because of the way t contingently is.

Intrinsic descriptive properties. Intrinsic descriptive properties rarely correspond to classes, because they do not carry a principle of identity [11]. So, for example, the property of being red for a rose is typically expressed as an attribute-value pair within the class Rose (Fig. (4a)), where the attribute name implicitly denotes the color quality [3]. We have three reification options, corresponding to different Truthmaking Patterns. A weak TMP emerges when the quality is reified as a separate class (Fig. (4b)). Note the 1-1 cardinality constraint, showing that a quality inheres in exactly one object, and an object has exactly one quality of a given kind. A strong TMP is exemplified in Fig. (4c), where an event of color occurrence is reified. The first option is generally more flexible, making it possible to describe the way the quality interacts with the world (Mary likes the color of this rose), or further information about the quality itself (the color of a rose is located in its corolla). The second option is however necessary when we need to account for temporal information (e.g., how long the redness lasted), or for the spatiotemporal context (what happened meanwhile and where...). To achieve the maximum expressivity, a third option is that of a full TMP, including both strong and weak TMs plus the relationship among them (Fig. (4d)). Concerning the latter, note that there is a formal ontological connection between qualities and events, discussed in [9]: events can be seen as manifestations of qualities, and qualities as the focus of events.

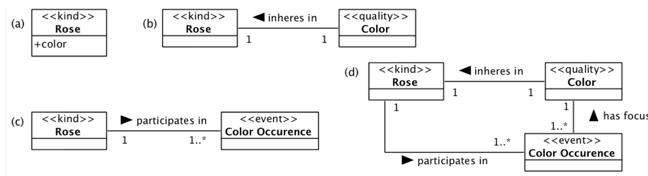


Fig. 4: Truthmaking patterns for an intrinsic descriptive property [10]

External descriptive relations. External descriptive relations hold in virtue of at least one relational quality inhering in at least one relatum. We distinguish two main cases: single-sided relations holding in virtue of one or more qualities inhering in just one relatum, and multi-sided relations holding in virtue of at least two qualities, each inhering in a different relatum. The reification of multi-sided relations is often necessary to model social and legal relationships, such as marriages, economic contracts, employment relationships, and so on. An example of the first kind is an attitudinal relation such as desires, represented in Fig. (5a). A weak TMP is shown in Fig. (5b), where a desire quality inhering in an agent and depending on some resources is reified. Note that we have represented it as a quality, but it could be seen as well as a relator consisting of just one quality. The addition of a strong TM, resulting in a full TMP, is shown in Fig. (5c). The event labeled DesireEvolution describes whatever happens in reality whose focus is that particular desire, such as the arising of the desire and its satisfaction.

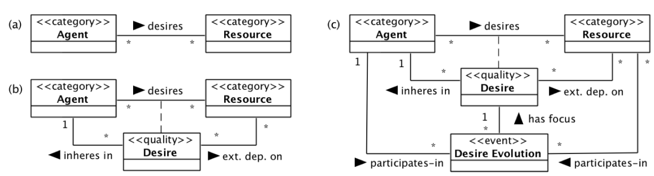


Fig. 5: Weak and full truthmaking patterns for a single-sided relation [10]

The Powertype Pattern

In several subject domains there is the need to deal with multiple classification levels. In such domains, the occurrence of situations in which instances of a type are specializations of another type is recurrent [3]. This phenomenon is known in the conceptual modeling community as the Powertype Pattern [3].

The Powertype Pattern is an example of an early approach for multi-level modeling in software engineering. This approach is used to model situations in which the instances of a type (the power type) are specializations of a lower-level type (the base type), and both power types and base types appear as regular classes in the model.

In [3], the authors address multi-level modeling from the perspective of the Powertype Pattern. They propose an axiomatic well-founded theory called MLT (for Multi-Level Theory) and apply it to revise the powertype support in UML. In their approach, they propose to mark the association between the base type and the higher order type with the ≪instantiation≫ stereotype, in order to distinguish it from other domain relations that do not have an instantiation semantics. An association stereotyped ≪instantiation≫ represents that instances of the target type are instantiated by instances of the source type and, thus, denote that there is a *characterization relation* (in the technical sense of [3]) between the involved types (regardless of possible generalization sets). The multiplicities of the "target" side of an ≪instantiation≫ association can be used to distinguish between the different variations of characterization. Whenever the lower bound multiplicity of the target association end is set to one, each instance of the base type is instance of, at least one instance of the powertype (e.g., every instances of person is necessarily either a living person or a deceased person). Thus, the higher order type *completely characterizes* the base type. In contrast, if the lower bound multiplicity of the target association end is set to zero, the inferred characterization relation is not a complete characterization. Analogously, if the upper bound multiplicity of the target association end is set to one, each instance of the base type is instance of, at most one instance of the higher order type. Thus, in this case, the higher order type *disjointly characterizes* the base type (again, no person can be both an instance of living person and of deceased person). In contrast, if the upper bound multiplicity of the target association end is set to many (*), the inferred characterization relation is not a disjoint characterization. Fig. 6 shows the application of the Powertype Pattern proposed in [3]. As the authors show, there are non-trivial interactions between the semantics of the ≪instantiation≫ relation and the meta-properties of a given generalization set. In this example the generalization set is *incomplete* and *disjoint* meaning that: (i) there are instances of Employee which are not instances of any instance of Management Role (as a consequence of the semantics of the instantiation association) ; and (ii) there are instances of Employee which are neither Organization President nor Department Dean (as a consequence of the semantics of incomplete generalization sets).

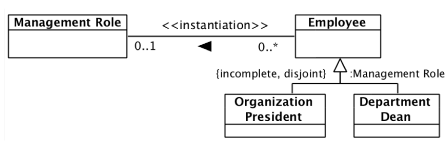


Fig. 6: Using ≪instantiation≫ [3]

The UML extensions proposed in [3] go beyond the ≪instantiation≫ stereotype and the lower/upper bound multiplicities. Further details of their approach fall outside the scope of this paper and are not presented here. For a complete description of the approach just described the reader should refer to [3].

3 Piecing it all together

3.1 Applying to Dimensions

Foundational Ontology Patterns (FOP) can be used to improve the expressiveness of multidimensional models, thus, facilitating activities, such as communication and meaning negotiation, as well as the semantic interoperability regarding the domains represented therein. The application of FOPs in the modeling of dimensions provide more semantics for the concepts represented.

For example, the modeling of role-playing dimensions can benefit from the use of the Role Pattern, as it can be used to represent the different roles played by a dimension, at the same time that it makes it explicit that the same entity plays different roles in that specific context. Fig. 7 illustrates the application of the Role Pattern in the dimension FINANCIAL INSTITUTION of the star-schema illustrated in Fig. 1. In a payment event (fact), FINANCIAL INSTITUTION (dimension) plays two different roles: CREDITED FINANCIAL INSTITUTION (the Financial Institution whose account should be credited) and DEBITED FINANCIAL INSTITUTION (the Financial Institution whose account should be debited).

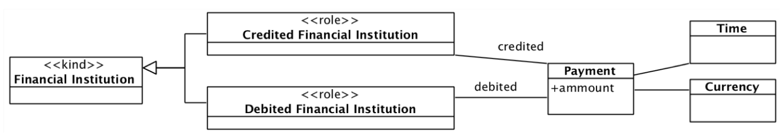


Fig. 7: Application of the Role Pattern in the modeling of Role-Playing Dimensions

When the role played by a dimension aggregates properties that are common to different Roles, the RoleMixin Pattern can be applied. Again, at the same time that the pattern reinforces the truthfulness of the concepts represented, it makes explicit the nature and the restrictions applicable to the entity represented by the dimension. The OntoUML model presented in Fig. 8 illustrates the application of the RoleMixin Pattern in the modeling of dimensions that represent borrowers, in the context of Finance. In this case, a borrower may be defined as a person or an organization that obtains a loan from a Financial Institution. In the figure, LOAN represents the fact table about the loans.

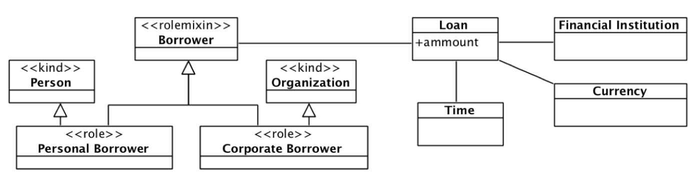


Fig. 8: Application of the RoleMixin Pattern

In the figure, BORROWER is the RoleMixin that covers different role types. CORPORATE BORROWER and PERSONAL BORROWER are the disjoint subclasses of BORROWER that can have direct instances, representing the sortal roles that carry the principles of identity that govern the individuals that fall in their extension. Dimensions ORGANIZATION and PERSON are the ultimate substance sortals (kinds) that supply the principles of identity carried by Corporate Borrower and Personal Borrower, respectively. The application of the RoleMixin Pattern preserves the unity of the concept borrower at the same time that clarifies the distinction between different types of borrowers (personal borrower and corporate borrower), satisfying both the modeling of facts related to all types of borrowers and the modeling of facts related only to a specific type of borrower (person or organization).

Analogously, when it is necessary to relate a dimension to a fact whose instances apply only to a subset of the dimension instances (corresponding to a phase partition) the Phase Pattern may be applied. Fig. 9 depicts an example of the Phase Pattern applied to the modeling of a fact representing exams taken by applicants for a driver's license. As only persons over 18 years are eligible to a driver's license, the Person dimension related to the fact Driver License Exam should be restricted to people meeting the minimum age requirement. The Phase Pattern was applied to create three phase-partitions specializing the dimension Person (Child, Adolescent and Adult). Then it was possible to relate the fact Driver License Exam to a subset of the Person dimension representing only adults (Phase Adult).

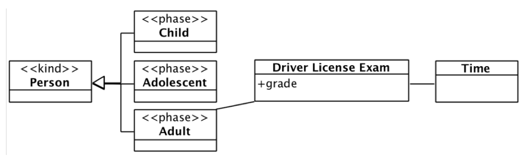


Fig. 9: Application of the Phase Pattern

Finally, the Collective Pattern is applicable to dimensions that represent entities as integral wholes, composed by members that play the same role in the collective. In many cases, in multidimensional models, it is important to distinguish the conceptualization of the whole from the conceptualization of the parts, because it is necessary to relate the whole to a fact that applies to the collective and the parts to a fact applicable only to the individuals. At the same time, it is important to make explicit the existence of a uniform relationship that holds between all parts (and only those parts). Fig. 10 presents an example of the Collective Pattern applied to a multidimensional model in the context of product manufacturing systems, which work with the concepts of Lot and Item. In this case, a Lot is defined as a group composed of a definite quantity of some product, manufactured under conditions of production that are considered uniform, while the Item corresponds to each product in the Lot. In the model, the dimension Lot represents the collection, while Item represents its members. In this approach, it is possible to relate the dimension Lot to the fact Delivery containing information applicable to the collective (for example, the lot weight), as well as to relate the dimension Item to the fact Sell whose granularity is the individual product (for example, unit price).

Turning now to Truthmaking Patterns, this technique can be applied in the modeling of dimensions to improve the expressivity of attributes describing di-

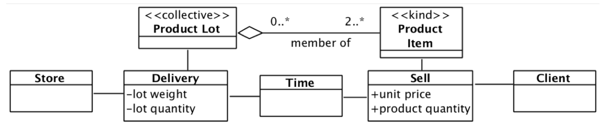


Fig. 10: Application of the Collective Pattern

mension level members. These attributes are mostly intrinsic descriptive properties that can be reified as previously discussed.

Take as example the dimension HOTEL illustrated in Fig. 11(a). The property "star rating", used to classify hotels according to their quality, is typically expressed as an attribute-value pair within the dimension HOTEL (Fig. 11(a)), where the attribute name implicitly denotes the hotel star rating quality. The first option is to reify this quality (weak truthmaker) as separate class (Fig. 11(b)), making it possible to describe the ways the quality interacts with the world (e.g., people prefer hotels rated from four to five stars), or further information about the quality itself (e.g., the hotel star rating is reviewed annually). The second option is to reify the event of "star rating occurrence" (strong truthmaker), which allows to account for temporal information (e.g., how long the hotel has been rated as five stars), or for the spatiotemporal context (what happened when the rating changed from five to four stars). The third option, which gives maximum expressivity, is that of a full TMP, including both strong and weak TMs plus the relationship among them (Fig. 11(d)).

Fig. 11: Hotel Dimension with a "star rating" attribute

Finally, there is another modeling issue that, despite being often neglected in the design of multidimensional models, should be addressed in the models to reinforce truthfulness to the reality. This is the case of dimensions that represent entities of different classification levels. For example, let us take the case of Financial Institutions and their types. Consider that FINANCIAL INSTITUTION can be specialized in BANK, INSURANCE COMPANY, INVESTMENT COMPANY and BROKERAGE FIRMS. In this case, "Bank A" and "Bank B" are particular BANKS, both instances of FINANCIAL INSTITUTION. Data analysis under the perspective of the type of FINANCIAL INSTITUTION are particularly common in this context, then the TYPE OF FINANCIAL INSTITUTION should also be considered as an entity, whose instances are "Bank", "Insurance Company", "Investment Company" and "Brokerage Firm". Traditionally, entities like FINANCIAL INSTITUTION and TYPE OF FINANCIAL INSTITUTION are represented as unrelated dimensions in multidimensional models and the relationship between the different classification levels is not explicit in the models. We propose the use of the Powertype Pattern previously mentioned (section 2.2) to address this issue. Fig. 12 presents an example of the application of the Powertype Pattern to the scenario of Financial Institutions and their types. In the example, the association stereotyped ≪instantiation≫ has both the lower and the upper bound multiplicity set to one, meaning that the target dimension (Type of Financial Institution) disjointly and completely characterize the source dimension (Financial Institution). Thus, the model in Fig. 14 represents that: (i) every instance of Financial Institution must be either an instance of Bank, an instance of Insurance Company, an instance of Investment Company, or an instance of Brokerage Firm and that (ii) "Bank" and "Insurance Company", "Investment Company" and "Brokerage Firm" are the only admissible instances of Type of Financial Institution.

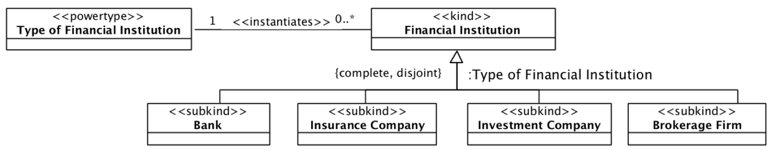


Fig. 12: Application of the Powertype Pattern

3.2 Applying to Facts

Conceptual modeling tools and techniques can also be applied to provide more semantics for the concepts represented by fact tables.

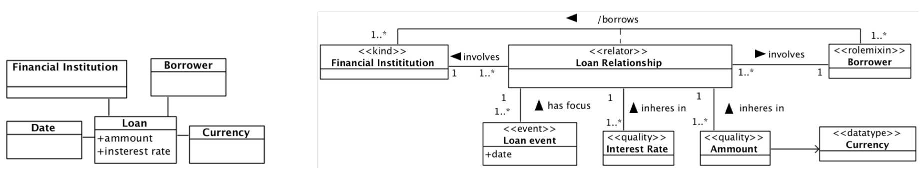
In [20], Kimball defines fact tables as many-to-many relationships with the dimensions. In the same book, Kimball states that fact tables in multidimensional models store measurements resulting from organizations' business processes events. In one of his examples he illustrates a shipment process and states that each movement of product onto an outbound truck generates performance measures or facts, such as the shipment quantity. In this way, it seems that Kimball is committed to the view that fact tables are relationships, but he also admits that a fact table corresponds to a physical observable event.

In [9], the authors propose a view in which events emerge from scenes as a result of a cognitive process that focuses on relationships: relationships are therefore the focus of events, which in turn can be seen as manifestations of relationships. Further in the paper, they state that referring to the relationship (which maintains its identity during the event) is unavoidable when we need to describe what changes in time, while referring to the event is unavoidable when we need to describe contextual aspects that go beyond the relationship itself.

In the light of what has been discussed in [20] and [9] regarding relationships and events, a reasonable approach would be to consider two elements w.r.t. fact tables: the fact as a relationship involving multiple participants (dimensions) and, on the other hand, the event that is the sum of the manifestations of the qualities constituting this relationship (measures). According to [26], not only the relationships should be reified but also the events.

Following the terminology for kinds of relationships defined in [10], we may classify fact tables as external descriptive relations, as they hold in virtue of relational qualities (measures) inhering in their relata (dimensions). Thus, both the relationship and the event (whose focus is the relationship) can be reified by applying the TMP for external descriptive relations previously mentioned.

An example of the application of the full TMP is presented in Fig. 13, where the TMP was applied to the fact table Loan represented in the star schema of Fig. 13(a). The example describes a loan relation holding between a Financial Institution and a Borrower. The relator is shown as a Loan Relationship composed of the amount, which has a value in a Currency conceptual space, and of the loan interest rate. Because the amount was reified as a ≪quality≫ (whose instances inhere in the loan), it is possible to express further information about it, for instance: (1) "This was the highest loan amount so far" or (2) "The amount borrowed did not reach the credit limit. It is still possible to grant new loans". In addition, the application of the TMP allows to explicitly represent other relevant information regarding the Loan Relationship, such as the reciprocal commitments and claims inhering in the financial institution or the borrower (and externally dependent on each other). The event labeled Loan Event describes the loan date as well as whatever happens in reality whose focus is that particular loan, such as the occurrence of loan disbursements, repayments and credit risk assessments.



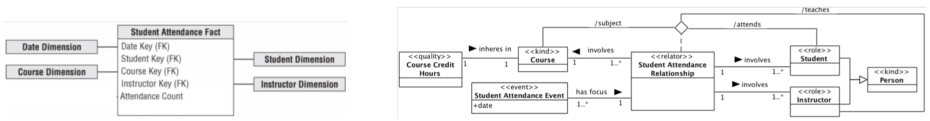
(a) Star schema (b) OntoUML Diagram

Fig. 13: Application of Truthmaking Patterns to Fact Tables

The reification of measures as individual qualities represents an interesting improvement in the semantic expressiveness of measures in multidimensional models. It allows to express the correlated units of measures, magnitudes, and scales, which are generally overlooked in multidimensional approaches. This empowers multidimensional models because each scale type defines a mathematical structure on which the permissible statistics and scale transformations are allowed. It also provides a better understanding about the nature of additivity constraints, as many statistic functions may be used to aggregate data cells of measures, though their use depends on which sort of measure and aggregation criteria are involved. Identifying these concepts in the multidimensional models, based on their ontological foundations, enables designers to describe properly what is being modeled, and therefore, to elucidate how data should be analyzed.

4 Case Illustration on Education: Student Attendance

To validate and demonstrate the contribution of our proposition to the multidimensional modeling practice, we have applied it to model a tangible example: a case study on education, extracted from [20], designed to track student attendance in a course. In this model, the grain is each occurrence of a student walking through the course classroom door each day, considering multi-instructor courses (that is, co-taught courses). The original multidimensional model, depicted in Fig. 14(a) [20], allows business users to answer questions concerning student attendance at courses, including: which courses were the most heavily attended, which students attended which courses, and which faculty member taught the most students.



(a) Star schema (b) OntoUML Diagram

Fig. 14: Applying Ontology Patterns to the Multidimensional Model

We applied ontology patterns to reengineer the original model (Fig. 14(a)) and produced the OntoUML model depicted in Fig 14(b). By applying the Role Pattern, we elucidated that both Student and Instructor are roles played by Persons. Consequently, Person instances can move in and out of the extension of these roles (due to changes in their relational properties), without any effect on their identity. For example, a Student is a role that a Person plays when related to an education institution, and it is the establishment (or termination) of this relation that alters the instantiation relation between an instance of Person and the type Student. The application of the ontological pattern not only provides more clarity and expressiveness to the model, but also favors the reuse of encoded experiences and good practices. Considering the existence of a property Course Credits Hours in the dimension Course, we have applied the TMP for Intrinsic Properties to reify the Course Credits Hours as separate class, thus making it possible to describe the ways in which this quality interacts with the world (e.g., this amount of credit hours can also be earned by taking part on a summer school), or further information about the quality itself (e.g., course credit hours are specified in the course regulation). In this case, the TMP contributes to enrich the expressivity of the model. Finally, we applied a full TMP for External Descriptive Relations on the original table Student Attendant Fact to reify the truthmaker of the Student Attendant Fact Relationship (between student, course and instructor) by means of the Student Attendant Event. The application of the TMP allows to explicitly represent relevant information regarding the Student Attendant Relationship, as well as to explicit represent the Student Attendant Event, which describes the date of the student attendance, as well as whatever happens in reality, whose focus is that particular student attendance, such as late arrivals or early leaves from the classes. By applying the TMP we improve the model expressivity, conceptual clarity as well as its truthfulness to reality. The aforementioned benefits seem to corroborate the fact that the use of ontological patterns in multidimensional modeling helps domain experts to externalize the knowledge about the domain, making the ontological commitments explicit and the models more truthful to the domain being represented.

5 Conclusions

This paper described our approach to systematically apply ontological patterns in the design of multidimensional models. We have discussed how conceptual modeling techniques can be applied in combination for building consistent multidimensional models. In our approach we focused on the application of Foundational Ontology Patterns, Reification and Truthmaking Patterns and the Powertype Pattern to improve the semantic expressiveness of multidimensional models. The case illustration in the area of Education exemplified how our propositions contribute to improve the quality of multidimensional models, enhancing their quality as artifacts to support communication, problem solving, meaning negotiation and, principally, semantic interoperability in its various manifestations.

Conceptual modeling is a fundamental discipline to several communities in computer science. In the future we plan to extend our work by conducting an analysis of the role played by conceptual models and philosophically grounded foundational ontologies in the scope of other technologies used for data analytics.

Acknowledgment

CAPES (PhD grant# 88881.173022/2018-01) and OCEAN project (UNIBZ).

References

1. Abelló, A., Samos, J., Saltor, F.: Yam2: a multidimensional conceptual model extending uml. Information Systems **31**(6), 541–567 (2006)
2. de Almeida Falbo, R., Guizzardi, G., Gangemi, A., Presutti, V.: Ontology patterns: Clarifying concepts and terminology. In: WOP (2013)
3. Carvalho, V.A., Almeida, J.P.A., Guizzardi, G.: Using a well-founded multi-level theory to support the analysis and representation of the powertype pattern in conceptual modeling. In: CAISE (2016)
4. Dahchour, M., Pirotte, A.: The semantics of reifying n-ary relationships as classes. In: ICEIS (2002)
5. Davidson, D.: The individuation of events. In: Essays in honor of Carl G. Hempel, pp. 216–234. Springer (1969)
6. Franconi, E., Kamblet, A.: A data warehouse conceptual data model. In: SSDBM (2004)
7. Galton, A.: Reified temporal theories and how to unreify them. In: IJCAI. pp. 1177–1183. Citeseer (1991)
8. Golfarelli, M., Maio, D., Rizzi, S.: The dimensional fact model: A conceptual model for data warehouses. International Journal of Cooperative Information Systems **7**(02n03), 215–247 (1998)
9. Guarino, N., Guizzardi, G.: Relationships and events: Towards a general theory of reification and truthmaking. In: AI*IA (2016)

10. Guarino, N., Sales, T.P., Guizzardi, G.: Reification and truthmaking patterns. In: ER (2018)
11. Guizzardi, G.: Ontological foundations for structural conceptual models. CTIT, Centre for Telematics and Information Technology (2005)
12. Guizzardi, G.: Ontological foundations for conceptual part-whole relations: The case of collectives and their parts. In: CAiSE (2011)
13. Guizzardi, G., Halpin, T.: Ontological foundations for conceptual modelling. Appl. Ontol. **3**(1-2), 1–12 (2008)
14. Guizzardi, G., Wagner, G., Almeida, J.P.A., Guizzardi, R.S.S.: Towards ontological foundations for conceptual modeling: The unified foundational ontology (UFO) story. Applied Ontology **10**(3-4), 259–271 (2015)
15. Guizzardi, G. et al.: Towards ontological foundations for the conceptual modeling of events. In: ER (2013)
16. He, L., Chen, Y., Meng, N., Liu, L.Y.: An ontology-based conceptual modeling method for data warehouse. In: ICM (2011)
17. Hüsemann, B., Lechtenbörger, J., Vossen, G.: Conceptual data warehouse design. Universität Münster. Angewandte Mathematik und Informatik (2000)
18. Khouri, S., Ladjel, B.: A methodology and tool for conceptual designing a data warehouse from ontology-based sources. In: Proceedings of the ACM 13th international workshop on Data warehousing and OLAP. pp. 19–24. ACM (2010)
19. Kimball, R., Ross, M.: The data warehouse toolkit: the complete guide to dimensional modeling. John Wiley & Sons (2011)
20. Kimball, R., Ross, M., Thornthwaite, W., Mundy, J., Becker, B.: The data warehouse lifecycle toolkit. John Wiley & Sons (2008)
21. Luján-Mora, S., Trujillo, J., Song, I.Y.: A uml profile for multidimensional modeling in data warehouses. Data & Knowledge Engineering **59**(3), 725–769 (2006)
22. MacBride, F.: Truthmakers. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University (2019)
23. Mylopoulos, J.: Conceptual modeling and telos. In: ER (1992)
24. Olivé, A.: Relationship reification: A temporal view. In: CAiSE (1999)
25. Pardillo, J., Mazón, J.N.: Using ontologies for the design of data warehouses. arXiv preprint arXiv:1106.0304 (2011)
26. Pedersen, T.B.: Multidimensional modeling. In: Encyclopedia of Database Systems, pp. 1777–1784. Springer (2009)
27. Romero, O., Abelló, A.: Automating multidimensional design from ontologies. In: Proceedings of the ACM tenth international workshop on Data warehousing and OLAP. pp. 1–8. ACM (2007)
28. Romero, O., Abell, A.: A framework for multidimensional design of data warehouses from ontologies. Data Knowl. Eng. **69**(11), 1138–1157 (2010)
29. Ruy, F.B., Guizzardi, G., Falbo, R.A., Reginato, C.C., Santos, V.A.: From reference ontologies to ontology patterns and back. Data & Knowl. Eng. **109**, 41–69 (2017)
30. Sapia, C., Blaschka, M., Höfling, G., Dinter, B.: Extending the e/r model for the multidimensional paradigm. In: ER (1998)
31. Selma, K., Ilys, B., Ladjel, B., Eric, S., Stphane, J., Michael, B.: Ontology-based structured web data warehouses for sustainable interoperability: requirement modeling, design methodology and tool. Computers in Industry **63**(8), 799 – 812 (2012)
32. Thenmozhi, M., Vivekanandan, K.: A tool for data warehouse multidimensional schema design using ontology. International Journal of Computer Science Issues (IJCSI) **10**(2), 161 (2013)
33. Verdonck, M., Gailly, F.: Insights on the use and application of ontology and conceptual modeling languages in ontology-driven conceptual modeling. In: ER (2016)