Ontologies and data models: essential properties and data modeling for petroleum exploration

Abel, M.[1]; Perrin, M.[2] ; Carbonera, J.L. [1]; Garcia, L. [1]
[1]Informatics Institute – Universidade Federal do Rio Grande do Sul, Brazil
[2] Geosiris, France.

Petroleum exploration and production rest on reservoir models that integrate a large set of data of various kinds. The common backbone of these data is the object of the modeling itself: the reservoir and the geological properties attached to it. Each category of professionals involved in the reservoir study views this reality according to some specific field of knowledge. These specialists thus generate various sets of data, each resting on a different conceptualization of one same object: the petroleum prospect. The resulting data models can be efficient in attending a particular application, but they are hardly interoperable and thus difficult to use in federate software environments. In view of this situation, petroleum exploration appears to be a domain rich in challenges related to conceptual modeling and data integration, in which ontologies can play a central role.

Ontology definition

Ontology is a branch of Philosophy that studies the nature of existent beings and their mutual relationships. In Computer Science, the term ontology has been used to designate an artifact (a file, a description, a representation) that formally describes, in a computer language, a set of concepts, whose meaning is shared by a community of practitioners. Significant progress was made in the field of ontology studies in the late 90's, when Nicola Guarino analyzed the various meanings in which the word ontology was being used (Guarino 1998). He insisted on the idea that an ontology is, primarily, a logical theory accounting for the *intended meaning* of the formal vocabulary utilized by a community for naming the elements of its domain. Guarino introduced a few meta-properties based on philosophical notions, such as *identity, unity, rigidity*, and *dependence* (Guarino & Welty 2000), which greatly help to clarify the meaning of the concepts that are currently expressed by means of domain ontologies in the various fields. We intend to demonstrate here, by a gentle introduction of two of these metaproperties – rigidity and dependence - that analyzing information through the view of ontological metaproperties, as proposed by Guarino, can be helpful for reducing both the complexity and the ambiguity of data models.

The use of ontological metaproperties in modeling

The first useful ontological notion is *essence*. According to (Guarino & Welty 2004), a property attached to an entity is essential to this entity if it must hold for it in every possible world. For example, being crystalline is an essential property for a mineral but it is not for a gemstone, since we can produce gemstone from non-crystalline material, like amber. When a property is essential for every instance that can exhibit it, we say that this property is *rigid*. The notion of *property,* in Logic, refers to every predicate that can be applied to a given instance, like "being a horse", "being a mineral" or "having a brain". In our example, "being crystalline" is essential for minerals, but not for other substances, like glass, so it is not a rigid property. Considering another example, a human being is an instance of the concept person and a human being is a person along all his life (and even after). Then the quality of "being a person" is rigid since there is no instance of human being that can stop being a person. Conversely, being a student is not a rigid property, since someone can stop being a student without stopping existing. A piece of mineral cannot stop being a mineral, but an entity which we consider being a gemstone, has not been a gemstone all along its existence since it was not one before having been cut and polished in order to be used in jewelry. Student and gemstone are defined by *anti-rigid* properties that define *roles,* like a student related to a person, or *phases,* like gemstones related to some mineral piece.

The notions of essence and rigidity help in identifying the concepts in the domain that provide the identity to individuals and can be tracked in the models. It thus allows one to identify vocabulary practices that may cause ambiguity like denominating instances of a domain according to anti-rigid properties and building models over anti-rigid concepts. For example, naming a person as a "client", a geographic area as a "prospect", a geological unit as an "economic target" hardly help in producing long term integrable models.

In the field of data models, considering essential properties allows one to correctly identify entities and to produce a more precise representation, which facilitates further integration and interoperability. We will analyze here a simple example related to petroleum exploration: the modeling of the entity *reservoir*[1].

In the context of petroleum exploration, a reservoir is a volume inside a prospect, which may contain petroleum and water. For modeling it, we must examine whether the property of "being a reservoir" is rigid or not. In other words, we should decide whether some entity called "reservoir" may stop being a reservoir and still exist. The answer strongly depends on the modeler's conceptualization of a reservoir. Some geologists may simply define a reservoir as a portion of rock having high porosity/permeability. This definition is rooted in some *intrinsic* properties of the entity (porosity and permeability) that cannot be lost[2]. In this case, "being a reservoir" is a rigid property. This first conceptualization will produce the model showed in the Figure 1(a).

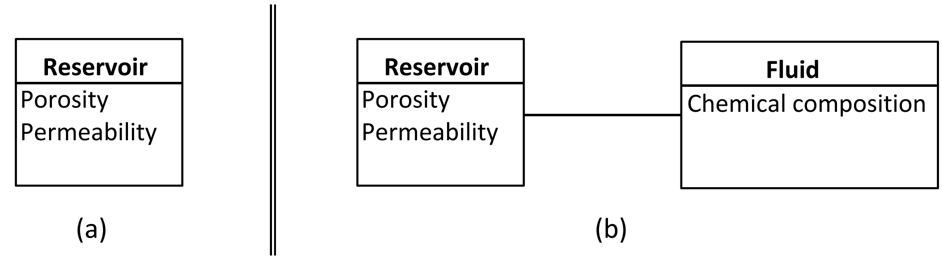

Figure 1 Alternative models for the entity reservoir based on intrinsic essential properties (a) or on external dependence (b).

However, some other geologists will consider that a portion of rock with high porosity and permeability is not a reservoir until its voids actually contain petroleum or water. This second definition implies that an instance will stop *being a reservoir* if it stops having water or petroleum inside its empty voids. If a reservoir is exposed to air, it will lose its content of petroleum or water, but the volume of rock to which it corresponds will not disappear. But, according to our second model definition, it will stop being a reservoir. The property of *being a reservoir* in this second model is anti-rigid. It is just a role of some existent portion of rock that should be considered as an entity of another concept, such as Rock body, and modeled in this way in the data model.

As shown in Figure 1 (b), this second model requires the modeling of a second entity, petroleum or water, which specifies the relational dependence that affects the instance of the reservoir that we consider. Any instance of an anti-rigid role concept has a *relational dependence* on some instance of another concept. It can exist only if the relationship exists. For example, a "student" cannot be a student if there does not exist some school or university in which he/she is

[1] An extensive analysis of the ontological properties of geological entities can be found in (Abel et al. 2015).

[2] We are considering here the time of exploration, not transformations over geological time.

registered. In our second model, an instance of reservoir cannot exist if there does exist a fluid (water or petroleum) inside its voids.

Deciding what is the rigid entity that provides identity to the several roles that an instance can assume is a central task in producing precise and efficient data models. The taxonomic (or hierarchical) structures that are defined, determine the subsumption relations that can be established between the various entities. Entities defined by anti-rigid properties cannot subsume entities (i.e. be the super class of) rigid ones (Guizzardi & Wagner 2005). Let us consider the schema shown in figure 2, which intends to model the variety of reservoirs that are explored in a petroleum company.

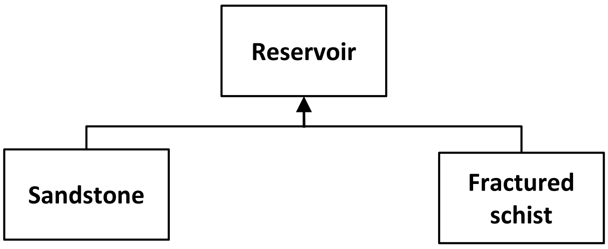

Figure 2 – Wrong use of the subsume relation.

The model shown in figure 2 is wrong because the class *Reservoir* cannot subsume the sub-classes *Sandstone* and *Fractured schist*. According to the schema of figure 2, the reason is that, the extensions (instances) of *Sandstone* and *Fractured schist* should be also extensions of *Reservoir* but this is not right since these rocks do not always constitute reservoirs. According to the design pattern proposed Guizzardi in [Guizzardi & Wagner, 2005] for dealing with such cases, we propose a better model on the schema of Figure 3. In this schema, the entities marked in grey are defined by rigid properties.

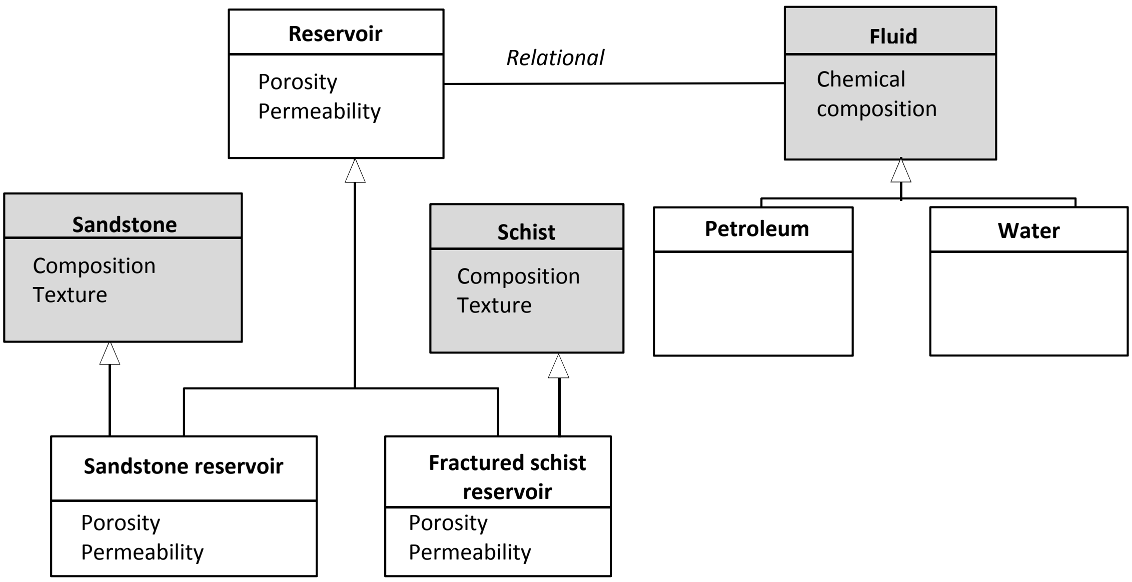

Figure 3 – Conceptual modeling based on ontology properties.

Advantages on ontological analysis

Ontological choices are not only an academic issue related to different modeling options. These choices have practical consequences for model usage and data consultation. In the example, the option of considering the reservoir entity as dependent of the entity fluid allows to create instances of fluid types or occurrences and to associate them to a particular instance of reservoirs. The first modeling option doesn´t allow this usage. Moreover, the model ambiguity can be reduced when the meaning of the represented entities is made explicit. This avoids that the same vocabulary be used to refer to two or more concepts that modelers or users consider being distinct. We additionally claim that providing a common framework based on essential entities allows reducing the number of entities and complexity of the resulting model.
Other ontological metaproperties require a better analysis in conceptual modeling activity. Especially in Petroleum Geology, properties like *identity* and *unity* can help in defining what exactly are the entities of reality that are being modeled in the database and also provide a good support to integrate models in the several scales of analysis (microscopic, well, reservoir, basin scales) into the petroleum chain. These metaproperties will be object of a further discussion.

Acknowledgement. This article is a corrected version of those published on Foundation Journal of the Professional Petroleum Data Management Association, Vol 3, Issue 1, pp.18-19, 2016. We are grateful to Nicola Guarino by the improvement in conceptual issues after the publication.

References

Abel M., Perrin M. & Carbonera J. (2015). Ontological analysis for information integration in geomodeling. *Earth Science Informatics*, 8, 21-36. Springer.

Guarino, N., ed. (1998), Formal Ontology in Information Systems. *Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998*, 3-15. IOS Press,

Guarino, N., ed. (1998), Formal Ontology in Information Systems. Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998. Amsterdam, IOS Press, pp. 3-15.

Guarino N. & Welty C. (2000). "A Formal Ontology of Properties". In: *The ECAI-2000 Workshop on Applications of Ontologies and Problem-Solving Methods.* IOS Press

Guarino N. & Welty C.A. (2004). "An overview of OntoClean". In: Handbook of Ontologies (eds. Staab S & Studer R). pp. 151-171. Springer

Guizzardi G. & Wagner G. (2005). Some applications of a unified foundational ontology in business modeling. *Ontologies and Business Systems Analysis*, 345-367. IGI Global.