UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

RICARDO WERLANG

Ontology-based approach for standard formats integration in reservoir modeling

Thesis presented in partial fulfillment
of the requirements for the degree of
Master of Computer Science

Advisor: Profa. Dra. Mara Abel

Porto Alegre
January 2015

CIP – CATALOGING-IN-PUBLICATION

Werlang, Ricardo

Ontology-based approach for standard formats integration in reservoir modeling / Ricardo Werlang. – Porto Alegre: PPGC da UFRGS, 2015.

93 f.: il.

Thesis (Master) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR–RS, 2015. Advisor: Mara Abel.

1. Geological data integration. 2. Communication standard formats. 3. Conceptual modeling. 4. Ontology. 5. Foundational ontology. 6. Geological objects mapping. I. Abel, Mara. II. Título.

“Never give up on a dream just because of the time it will take to accomplish it.
The time will pass anyway.”
— EARL NIGHTINGALE

ACKNOWLEDGMENT

I would thank my parents, Vasco Antônio Werlang and Natália Thomas Werlang, for their support in all moments of my life. Thank you for always trust my judgment and never doubting my abilities. Thank you for providing me with everything I needed to complete this work. I would thank my sister, Regina Werlang, for her long travels undertaken to provide me very happy moments.

I would thank my girlfriend, Jéssica Mota Paraguassú, by the patience shown during the completion of this project, a period full of turbulences and inconsistencies. Thank you for all happy moments we shared together, that despite the distance, never leave me alone. These moments were crucial to continue the work with renewed energy. I admire you by the person you are and by the strength that you provide to me by the simple fact of being always at my side. I love you very much.

I would thank all BDI group members and the Endeeper's staff that I had the pleasure of meeting: Sandro Rama Fiorini, Joel Luis Carbonera, Ricardo Linck, Douglas Eduardo Rosa, Jose Lozano, Guilherme Schievelbein, Alexandre Lorenzatti, Carlos Santin, Oscar Paesi da Silva, Eduardo Castro, Mauricio Leite Dau, Gabriel Barufi Veras, Vitor Fortes Rey, Luan Fonseca Garcia, Luciano Vargas Flores, Júlia Visnievski Zacouteguy, Luiz Henrique Boff and all others. All are wonderful people who taught me a lot and that continue to surprise me every day. This work takes a bit of each of you.

Finally, I would thank Professor Mara Abel, advisor of this work. I would not be able to list all the reasons why I would like to thank Mara. Anyway, thank you for welcoming me, for sharing your knowledge and experiences with me, for always being available when I really needed and for giving me the opportunity to work with you and your amazing team. Mara, you are a supermom, thank you!

ABSTRACT

The integration of data issued from autonomous and heterogeneous sources is still a significant problem for an important number of applications. In the oil and gas industry, a large amount of data is generated every day from multiple sources such as seismic data, well data, drilling data, transportation data, and marketing data. However, these data are acquired by the application of different techniques and represented in different standards and formats. Thus, these data exist in a structured form in databases, and in semi-structured forms in spreadsheets and documents such as reports and multimedia collections. To deal with this large amount of information, as well as the heterogeneous data formats of the data, the information needs to be standardized and integrated across systems, disciplines and organizational boundaries. As a result, this information integration will enable better decision making within collaborations, once high quality data will be accessible timely.

The petroleum industry depends on the efficient use of these data to the construction of computer models in order to simplify the geological reality and to help understanding it. Such a model, which contains geological objects analyzed by different professionals – geologists, geophysicists and engineers – does not represent the reality itself, but the expert's conceptualization. As a result, the geological objects modeled assume distinct semantic representations and complementary in supporting decision-making. For keeping the original intended meanings, ontologies were used for expliciting the semantic of the models and for integrating the data and files generated in the various stages of the exploration chain.

The major claim of this work is that interoperability among earth models built and manipulated by different professionals and systems can be achieved by making apparent the meaning of the geological objects represented in the models. We show that domain ontologies developed with support of theoretical background of foundational ontologies show to be an adequate tool to clarify the semantic of geology concepts. We exemplify this capability by analyzing the communication standard formats most used in the modeling chain (LAS, WITSML, and RESQML), searching for entities semantically related with the geological concepts described in ontologies for Geosciences. We show how the notions of identity, rigidity, essentiality and unity applied to ontological concepts lead the modeler to more precisely define the geological objects in the model. By making explicit the identity properties of the modeled objects, the modeler who applies data standards can overcome the ambiguities of the geological terminology. In doing that, we clarify which are the relevant objects and properties that can be mapped from one model to another, even when they are represented with different names and formats.

Keywords: Geological data integration. communication standard formats. conceptual modeling. ontology. foundational ontology. geological objects mapping.

Abordagem baseada em ontologias para integração de formatos padrões em modelagem de reservatórios

RESUMO

A integração de dados oriundos de fontes autônomas e heterogêneas ainda é um grande problema para diversas aplicações. Na indústria de petróleo e gás, uma grande quantidade de dados é gerada diariamente a partir de múltiplas fontes, tais como dados sísmicos, dados de poços, dados de perfuração, dados de transporte e dados de marketing. No entanto, estes dados são adquiridos através da aplicação de diferentes técnicas e representados em diferentes formatos e padrões. Assim, estes dados existem de formas estruturadas em banco de dados e de formas semi-estruturadas em planilhas e documentos, tais como relatórios e coleções multimídia. Para lidar com a heterogeneidade dos formatos de dados, a informação precisa ser padronizada e integrada em todos os sistemas, disciplinas e fronteiras organizacionais. Como resultado, este processo de integração permitirá uma melhor tomada de decisão dentro de colaborações, uma vez que dados de alta qualidade poderão ser acessados em tempo hábil.

A indústria do petróleo depende do uso eficiente desses dados para a construção de modelos computacionais, a fim de simplificar a realidade geológica e para ajudar a compreende-la. Tal modelo, que contém objetos geológicos analisados por diferentes profissionais — geólogos, geofísicos e engenheiros — não representa a realidade propriamente dita, mas a conceitualização do especialista. Como resultado, os objetos geológicos modelados assumem representações semânticas distintas e complementares no apoio à tomada de decisões. Para manter os significados pretendidos originalmente, ontologias estão sendo usadas para explicitar a semântica dos modelos e para integrar os dados e arquivos gerados nas etapas da cadeia de exploração.

A principal reivindicação deste trabalho é que a interoperabilidade entre modelos da terra construídos e manipulados por diferentes profissionais e sistemas pode ser alcançada evidenciando o significado dos objetos geológicos representados nos modelos. Nós mostramos que ontologias de domínio desenvolvidas com o apoio de conceitos teórico de ontologias de fundamentação demonstraram ser uma ferramenta adequada para esclarecer a semântica dos conceitos geológicos. Nós exemplificamos essa capacidade através da análise dos formatos de comunicação padrões mais utilizados na cadeia de modelagem (LAS, WITSML e RESQML), em busca de entidades semanticamente relacionadas com os conceitos geológicos descritos em ontologias de Geociências. Mostramos como as noções de identidade, rigidez, essencialidade e unidade, aplicadas a conceitos ontológicos, conduzem o modelador à definir mais precisamente os objetos geológicos no modelo. Ao tornar explícitas as propriedades de identidade dos objetos modelados, o modelador pode superar as ambiguidades da terminologia geológica. Ao fazer isso, explicitamos os objetos e propriedades relevantes que podem ser mapeados a partir de um modelo para outro, mesmo quando eles estão representados em diferentes nomes e formatos.

Palavras-chave: integração de dados geológicos, formatos de comunicação padrões, ontologias, ontologias de fundamentação, mapeamento de objetos geológicos.

LIST OF FIGURES

1.1 A *Fault* being represented differently along the modeling chain. 14

1.2 The many categories of data used by an oil company (from *Data Management Value Study*). 17

1.3 Summary of trends for terms in OnePetro (from *Data Management Value Study*). 18

2.1 The viewpoints of information systems interoperability [Fileto and Medeiros 2003]. 22

2.2 Three dimensions of heterogeneity (adapted from [Benerecetti, Bouquet and Ghidini 2000]). 24

2.3 Data Warehouse related tools and systems. 28

2.4 Three ways for connecting ontologies to information sources (adapted from [Wache et al. 2001]). 32

2.5 GAV and LAV approaches [Buccella, Cechich and Fillottrani 2009]. 33

3.1 UFO fragment that represents the fundamental distinctions between urelement and set, and universal and individual [Guizzardi 2005]. 38

3.2 UFO fragment that represents the distinctions between endurant and perdurant universals and individuals and between substantial and moment universals and individuals [Guizzardi 2005]. 39

3.3 Representation of the relationship between substance universals, quality universals and quality dimensions. In this figure, *i* represents a inherence relation, in the sense that the quality individual *w* is inherent to the substance individual *a*. While *ql* represents the relation between the quality individual *w* and the quale *q*, which represents its value, in the quality dimension [Guizzardi 2005]. 40

3.4 Representation of the taxonomy of UFO's substantial universals [Guizzardi 2005]. 41

4.1 Generation of gas and oil [Hyne 2012]. 44

4.2 Hydrocarbon migration (adapted from:[Perrin and Rainaud 2013]). 45

4.3 Examples of structural traps [Perrin and Rainaud 2013] 46

4.4 Information provided by well bores [Perrin and Rainaud 2013]. 48

4.5 The Earth Modeling Workflow (adapted from [Perrin et al. 2007]). 50

5.1 Historical evolution of software products for earth modeling presented in software packages categories by period of first appearance in the market [Perrin and Rainaud 2013]. 53
5.2 The top level *Basic Geology* ontology [Mastella 2010]. 56
5.3 Geological objects: Erosion surface E, fault F and Sedimentary strata unit U, constituted of volume and boundaries bu and bl [Mastella 2010]. 57
5.4 The concepts related with the LAS, WITSML and RESQML standard formats in Geological File Ontology. 58
5.5 The WITSML data objects relationships [Energistics and SIG 2012]. . . . 59
5.6 The feature/interpretation/representation/properties knowledge hierarchy [Energistics and SIG 2014]. 61

6.1 A *Sedimentary Unit* being represented differently in three models. 63
6.2 Using UFO meta-types to clarify the meaning of geological concepts. . . . 64

7.1 The top level *Basic Geology* ontology modified from [Mastella 2010] to describe the ontological meaning of relationships. 67
7.2 Relationships among the main entities considering an earth modeling scenario. 69
7.3 Possible mappings found between the LAS standard and *Basic Geology* ontology: *Core Data Section*. 71
7.4 Possible mappings found between the LAS standard and *Basic Geology* ontology: *Top Data Section*. 72
7.5 Possible mappings found between WITSML data-objects and *Basic Geology* ontology: *LithologyType*, *MatrixCementType* and *QualifierType*. 75
7.6 Possible mappings found between WITSML data-objects and *Basic Geology* ontology: *MudClass*, *MudSubClass* and *StimFluidSubType*. 75
7.7 Possible mappings found between WITSML data-objects and *Basic Geology* ontology: *StimFluidType*, *WellFluid* and *LithostratigraphyUnit*. 76
7.8 Possible mappings found between RESQML data-objects and *Basic Geology* ontology: *Phase* and *GeneticBoundaryKind*. 77
7.9 Possible mappings found between RESQML data-objects and *Basic Geology* ontology: *TectonicBoundaryKind* and *FluidContact*. 78
7.10 Possible mappings to entities from *Substance* sub-ontology. 80
7.11 Possible mappings to entities from *Geological Unit* sub-ontology.. 81
7.12 Possible mappings to entities from *Geological Boundary* sub-ontology. . . 81
7.13 Possible mappings to entities from *Geological Object* sub-ontology. 82

LIST OF TABLES

7.1 The selected LAS data sections and their descriptions. 70
7.2 The result of the classification realized with LAS types. 73
7.3 The selected WITSML type names and their descriptions. 74
7.4 Ontological analysis of the identified WITSML entities. 76
7.5 The selected RESQML data types and their descriptions. 78
7.6 The result of the classification realized with RESQML types. 80

LIST OF ABBREVIATIONS AND ACRONYMS

ASCII	American Standard Code for Information Interchange
DW	Data Warehouse
E&P	Exploration and Production
EAI	Enterprise-application Integration
EII	Enterprise-information Integration
ETL	Extract-Transform-Load
GAV	Global-as-view
GLAV	Global-local-as-view
GML	Geography Markup Language
IT	Information Technology
LAS	Log ASCII Standard
LAS	Log ASCII Standard
LAV	Local-as-view
LPG	Liquefied Petroleum Gas
OLAP	Online Analytical Processing
OLTP	Online Transaction Processing
PI	Principle of Identity
PRODML	Production Markup Language
PU	Principle of Unity
RESQML	Reservoir Characterization Markup Language
SQL	Structured Query Language
UFO	Unified Foundational Ontology
WITSML	Wellsite Information Transfer Standard Markup Language
XSLT	Extensible Stylesheet Language Transformations

CONTENTS

1 INTRODUCTION 13
1.1 Research motivations 16
1.2 Objectives 18
1.3 Organization of the master thesis 19

2 INFORMATION INTEGRATION 21
2.1 Information systems interoperability 21
2.2 Types of heterogeneity conflicts 23
2.3 The information integration process 25
2.4 System-based information integration 27
2.4.1 Data warehouse 27
2.4.2 Virtual data integration 28
2.4.3 Message mapping 29
2.4.4 Object-to-relational mappers 30
2.4.5 Document management 30
2.4.6 Portal management 30
2.5 Ontology-based integration approaches 31
2.6 Related work in semantic integration in petroleum domain 33

3 CONCEPTUAL MODELING AND ONTOLOGIES 35
3.1 Foundational ontologies 36
3.1.1 Unified Foundational Ontology (UFO) 37

4 PETROLEUM SYSTEM AND RESERVOIR MODELING 43
4.1 Basic concepts 43
4.2 Data derived from E&P activities available for reservoir modeling 47
4.2.1 Seismic data 47
4.2.2 Drilling data 48
4.2.3 Regional geology data 49
4.2.4 Laboratory data 49
4.3 Earth modeling 50

5 CURRENT SOLUTIONS FOR GEOLOGICAL KNOWLEDGE FORMALIZATION 53
5.1 Ontologies developed for Geosciences 54

5.2 Ontologies for earth modeling 55
5.2.1 Description of the Basic Geology ontology 56
5.3 Communication standard formats 57

6 PRODUCING MORE INTEGRABLE MODELS 63
6.1 Proposed methodological approach 65

7 ONTOLOGICAL ANALYSIS OF STANDARD DATA FORMATS 67
7.1 Step 1: Choosing the ontology 67
7.2 Step 2: Analyzing the chosen ontology 68
7.3 Step 3 and 4: Analyzing the standard formats data files and the identified entities 70
7.3.1 LAS: Step 3 70
7.3.2 LAS: Step 4 71
7.3.3 WITSML: Step 3 73
7.3.4 WITSML: Step 4 73
7.3.5 RESQML: Step 3 76
7.3.6 RESQML: Step 4 79
7.4 Step 5: Performing the mapping 80
7.5 Mapping validation and evaluation 82

8 CONCLUSIONS 85
8.1 Future Work 86

REFERENCES 87

1 INTRODUCTION

A large amount of heterogeneous data is generated every day from multiple sources related with petroleum exploration activities, such as 3D seismic interpretation, well bore drilling, reservoir modeling and monitoring and also plant/facility modeling or monitoring capabilities [Mastella 2010]. These data are routinely collected and stored in a structured form in databases, and in semi-structured forms in spreadsheets and documents, such as reports and multimedia collections, of many organizations. As a consequence, these data are complex in nature, often poorly organized and duplicated, and exist in different formats [Ge et al. 2011].

At present, engineers are faced to the challenge of having access to all information about their domain, in order to make well-informed decisions. Moreover, modelers from different disciplines, which must be able to share their diverse views of the world, face problems that make integration difficult to support. Especially for the petroleum exploration activity, where the acquisition, distribution and use of expert knowledge are more critical for the decision-making. When they intend to face the above challenges, geoscientists and modelers must face a difficulty of huge importance: data heterogeneity. To deal with this large amount of information, as well as the heterogeneous data formats, the information needs to be standardized and integrated across systems, disciplines and organizational boundaries. However, integrating these heterogeneous data to capitalize on their information value has been complex and costly [Chum 2007].

The typical information integration solution to provide a uniform interface to a collection of heterogeneous information sources, giving users the illusion that there is a centralized and homogeneous information system, works well for activities in which the sources are static. In the case where the data sources and formats are constantly evolving, the integration needs to be founded in the semantic of the data and the tools used to the construction of computer models need to be dissociated from the data sources and formats. In this context, integration means finding correspondence between entities from different fields, without merging the corresponding instances. The experts from these domains, in particular those from the petroleum industry, need data to remain where they are, and to keep their original format, which can differ completely from one field to the other. What they aim is to be able to have an integrated vision of the data issued from all the different fields [Mastella 2010]. They need a semantic-based integrated vision of the data. The semantic-based integration approach must take into consideration the meaning of the various data for the experts, in order to define mapping rules. In addition, the involved experts need to agree upon a common vocabulary for communicating and also need to describe the meaning of data and data formats within models. Thus, making evident the semantics of the related information and data. Also, correspondence among models should be operated by means of meaningful descriptions, which should be detached from the physical part of the models [Mastella 2010].

In order to take advantage of the value of the information contained in those data presented in different formats, engineers rely on simulations (and hence simulation models) to make important operational decisions on a daily basis. Three problems that are commonly encountered in model based oil field operations are [Soma et al. 2008]: on-demand access to information, integrated view of information, and knowledge management.

The problems of on-demand access and information integration arise because a number of different kinds of simulation models are created and used. Since these models are created by different processes and people, the same information could be represented differently across models. The geological concept named *Fault*, for instance, can be represented differently along the modeling chain (Figure 1.1).

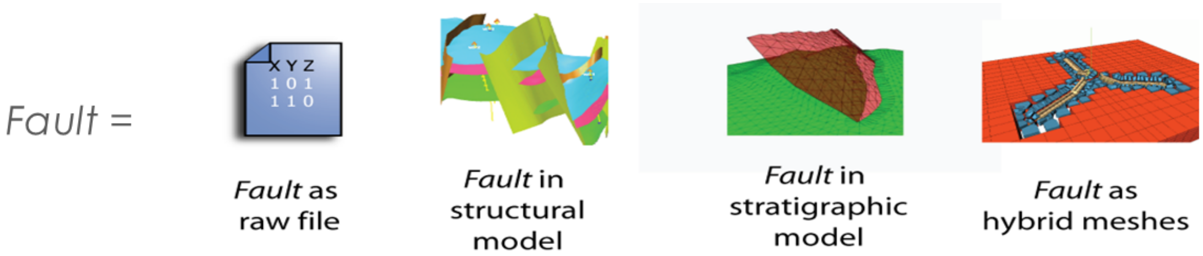


Figure 1.1: A *Fault* being represented differently along the modeling chain.

A unified view of the models and their simulations is desirable for decision making, and thus the necessity for information integration. The problem of knowledge management refers to a systematic way to capture the rationale (knowledge) behind the various analyses performed by an engineer and decisions taken based on the analyses. It is critical to capture this knowledge for auditing, archiving, and training purposes [Soma et al. 2008].

In the case of descriptions related to field outcrops, well core samples and rock thin sections, i.e., geological observations, they will hardly rest on models that would only be expressed by representations based on textual languages. Geologists solve problems by using for a good part their visual knowledge [Lorenzatti et al. 2011]. In these cases, the geologist captures the spatial arrangement and visual aspects of various geological entities for interpreting his/her observations. This way of thinking plays an important role for allowing the geologist building his/her conceptual model [Abel et al. 2004]. Geological observation and earth model building therefore require representations at least partly being based on graphical and pictorial representation languages. Accordingly, visualization tools are important means for allowing geologists to communicate.

In order to deal with the complexity of the world to be interpreted, geologists create mental models involving various geological objects. A *model*, which is an abstraction of a portion of the reality, emphasizes certain aspects of entities that exist in the world and that are relevant according to the modeler point of view [Guizzardi 2005]. *Abstraction* is the mental process that we use when we select some characteristics and properties of a set of objects, and exclude other characteristics that are not relevant [Batini, Ceri and Navathe 1992]. A model is built in the modeler's mind according to some conceptualization or theory. A *conceptualization* designates the internal mental reference that a particular individual abstracts and keeps in mind regarding the world around [Perrin and Rainaud 2013]. A geologist, for instance, creates in his mind abstract objects related to real world objects like geological units or boundaries, faults, sedimentary basins and geological reservoirs. Thus, a *model* designates some individualized part of a conceptualization. However, both conceptualizations and models are abstract entities that only exist in the geologists' mind. In order to be documented, communicated and analyzed, they need to be externalized in specifications, using a *modeling language* that designates the set of rules that defines in which way the representation should be coded/decoded for being understood both by the modeler and the receptor of the model. Thus, *representation* is one of the arbitrary forms in which the modeler expresses the model according to the language that he/she has chosen. The set of shared concepts into the geologist's mind and the correspondent external representations are what we call an ontology.

The term Ontology has its origin in philosophy where it was defined as a particular theory about the nature of being or the kinds of existents [Guizzardi 2005]. In the Computer Science context, the term ontology is used with two different meanings: (1) an artifact that represents a portion of reality according to the theory of existence [Gruber 2003]; (2) a logical theory accounting for the intended meaning of a formal vocabulary utilized for representation [Guarino 1998]. An ontology can be used for information integration and supporting applications in the first meaning, i.e., as an artifact (a model specified in a language and stored in a file), where the ontology can represent a domain. It is important to note that the ontology as an artifact should be built to express the common understanding of a community about the existence of beings. The ontology as a theory provides the support to build models that capture this understanding.

Earth models, that are three- (3D) or four-dimension (4D) representations of data and interpretation concerning subsurface resources, are key tools for identifying and characterizing potential hydrocarbon reservoirs. They are developed by geoscientists who are responsible for evolving a hydrocarbon prospect through various stages of modeling. This chain of activities, which starts with data acquisition and proceeds with several different steps of data analysis and interpretation, can be classified according to the purpose of the model under construction. According to [Perrin and Rainaud 2013], reservoirs can be studied from various points of view: geometry, rock quality and fluids. The study of reservoirs geometry is necessary for identifying possible reservoir sites and evaluating the volume of their envelope. Thus, the geometry is a key factor for determining the location of hydrocarbon traps. The following study is to identify if the located reservoirs have high-porosity rocks that may be potential reservoir rocks and impermeable rocks than serve as efficient covers. Also, to evaluate deposit conditions to determine whether source rocks are present in the sedimentary basin and whether conditions are favorable for the trapping and maturation of organic matter within the rocks. Finally, the study of fluids aims to simulate fluid migration and the evolution of reservoir content for evaluating the quantity of oil or gas that can be extracted. Therefore, the final goal of the modeling process is the building of a reservoir model, which will be used for simulating oil accumulation in the underground. Currently, this final model is connected to the original raw data by a long chain of successive interpretations realized by various groups of experts (e.g., geophysicists, structural geologists, reservoir engineers, etc.) who may have different conceptualizations about the modeled objects.

Experts involved in the modeling process use heterogeneous data management environments that rest on various data representations and encoding conventions for dealing with the same information in different parts of the workflow. Moreover, the data used in one modeling activity, commonly, must be exchanged for other activities, containing the aggregate information already acquired. Clearly, there is a need of a non-conventional semantic integration approach.

In order to facilitate this integration process, intensive development efforts have been made in order to provide highly interactive and user-friendly tools to use these data effectively. Due to such intense development efforts, the oil companies end up using a variety of software, but that only integrate and share data among applications of the same vendor. This scenario, where the software providers have intended to cover as much as they could all the steps of the reservoir characterization workflow, did not last long due the complexity of the reservoir modeling process and the quick technological evolution. Taking advantage of this situation, small software providers have also specialized in some “niches” of excellence, offering products that need to communicate with others. In order to face this interoperability problem, many efforts have been made to create file format standards to transfer of information from one application to another.

1.1 Research motivations

As each category of data is acquired, edited, interpreted and distributed, it moves from one repository to another. Each of these locations makes the data accessible to different groups of potential users. An integration approach to deal with these heterogeneous sources is still a significant problem in the petroleum industry, which is facing the problem of information overload (Figure 1.2). Use these massive volumes of data effectively is becoming a major challenge for this type of industry. At present, the petroleum businesses already consider data as a valuable asset and recognize that it has to be managed as an asset like the other assets in their business, like the wells, the facilities and the platforms. Moreover, all oil companies understand that data is crucial to their operations. This new way of thinking is bringing major investments to the area of integration and data management. Rainaud et al., in [Rainaud et al. 2005], estimated that 43% from the total budget of petroleum is currently dedicated to information integration.

A recent study about data management value[1], commissioned by CDA[2] and performed by Schlumberger[3], found that the way that subsurface information is managed has significant im-

---

[1] http://www.oilandgasuk.co.uk/datamanagementvaluestudy/

[2] http://www.cdal.com/

[3] http://www.slb.com/

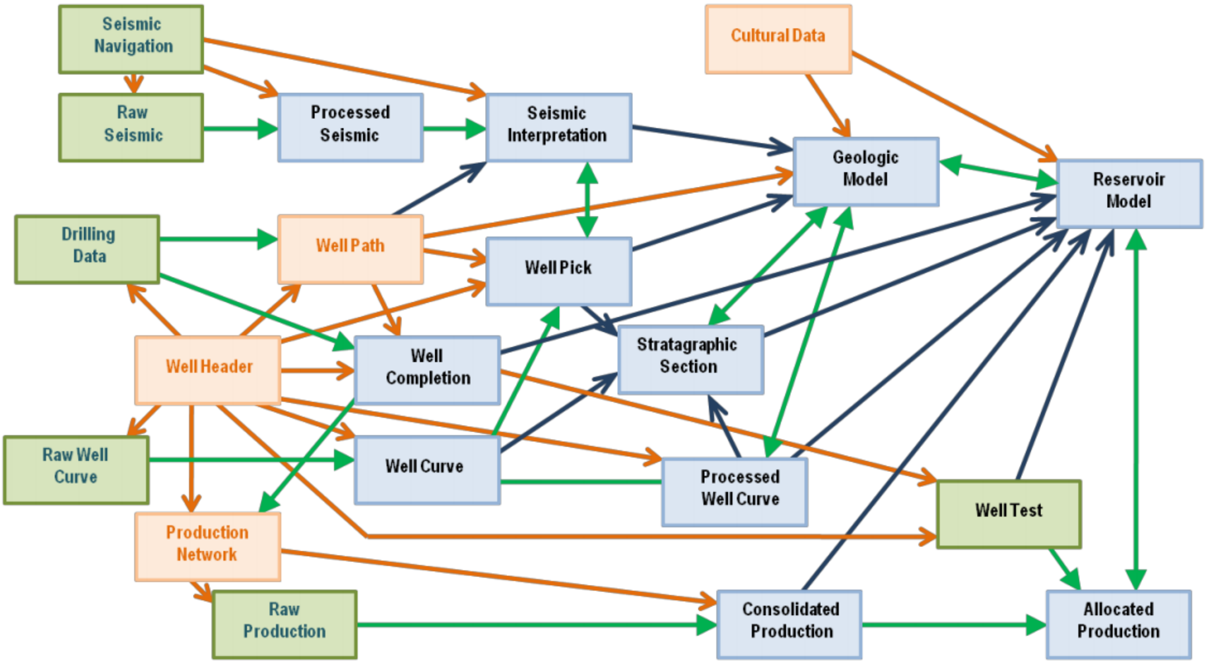


Figure 1.2: The many categories of data used by an oil company (from *Data Management Value Study*).

pact on the overall performance of E&P companies. Participants in the study estimated that "between a quarter and a third of all new value generated by typical E&P companies can be directly attributed to the subsurface data they hold". Some of the study conclusion is that data holds a pre-eminent place alongside people, work processes and IT applications in developing an understanding of the subsurface, which strongly and directly affects business performance. Thus, a review of current practices concerning data quality, access and indexing systems, data preservation, security and data governance would reveal substantial opportunities to increase the total value that E&P companies generate. Furthermore, data management nowadays has become an increasingly "transverse" domain and should integrate the different disciplines: geology; geophysics; reservoir; and IT as well.

Another research motivation is the growing recognition of the importance of data management and use of new technologies (e.g., witsml, ontology, smartfield) in journals and conferences dedicated to E&P business issues. Figure 1.3 summarizes the most active terms used over the last few years in the OnePetro[4] library. The terms listed in the top right hand corner are those that are both recent and growing strongly in usage, showing the importance of information management.

Intense efforts were developed, during the last years, by various organizations (geological surveys, geoscience consortia, oil companies) for issuing codifications and formalizations of geological knowledge. Moreover, the evolution of the technology motivated and enabled the

---

[4] OnePetro.org is a multi-society library that provides access a broad range of technical literature related to the oil and gas exploration and production industry, not least the Society of Petroleum Engineers and World Petroleum Congress. The site indexes more than 85,000 E&P related documents.

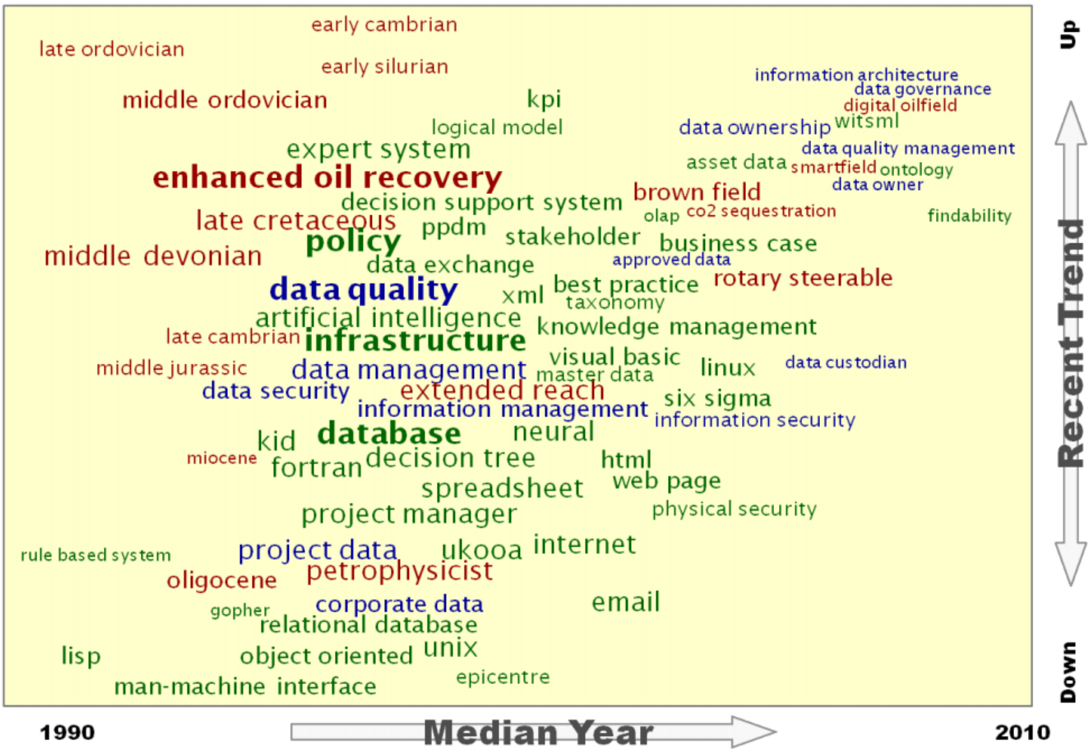


Figure 1.3: Summary of trends for terms in OnePetro (from *Data Management Value Study*).

creation of new file format standards to exchange geological information among the many activities related to reservoir modeling process. However, it is very difficult to some oil company already familiarized with some consolidated standard to start using a new one. The use of ontologies to determine if files represented in different format standards contain semantically related information has recently been used to overcome the problem of semantic heterogeneity, since they allow one to clarify the meaning of the represented concepts according to the intention of the geologists. Ontologies are important because they provide a shared and common understanding of data within a problem/solution domain, and by organizing and sharing enterprise information, as well as managing content and knowledge, they allow “better interoperability and integration of intra- and inter-company information systems” [Chum 2007].

Thus, the oil and gas industry becomes a potentially rich domain for semantic approaches in data modeling. Ontology-driven information integration and delivery leverage rich extensible domain ontologies found in the oil & gas industry, and combine them with industry standard definitions and controlled vocabularies. This results in meaningful metadata that reflects the concepts relevant to the domain.

1.2 Objectives

In this work, we claim that interoperability among earth models built and manipulated by different professionals and systems can be achieved by making apparent the meaning of the geological objects represented in the models. We show that domain ontologies developed with support of theoretical background of foundational ontologies show to be an adequate tool to clarify the semantics of geology concepts. We exemplify this capability by analyzing the communication standard formats most used in the modeling chain (LAS, WITSML and RESQML), searching for entities semantically related with the geological concepts described in ontologies for Geosciences, in order to make explicit the nature and properties of the geological objects found in each format. We aim to identify which entities in the model can be mapped from one application to another. We will restrict our analysis to the exploration steps of the petroleum chain, although we could later extend the study to cover production too.

Therefore, the main objective of this work is:

> *Clarify the semantics of geology concepts presented in communication standard formats most used in the reservoir modeling chain applying a theoretical background of foundational ontologies, in order to identify which entities in the model are able to be mapped from one application to another.*

This objective evokes the exploration of some specific points:

- Survey of available ontologies built for Geosciences;
- survey of the most used communication standard formats in petroleum exploration steps;
- analysis of the communication standard formats found looking for mappable concepts to those found ontologies;
- classification of geology concepts according to foundational ontologies primitives.

1.3 Organization of the master thesis

This work is organized as follows. Chapter 2 provides background for our research in information integration, presenting basic concepts related with information systems interoperability and types of heterogeneity problems that might arise due to heterogeneity of the data. It explains the information integration process and presents the most typical information integration systems. Chapter 3 presents the use of ontologies in the conceptual modeling process and show how the use of properties from foundational ontologies can lead the modeler to more precisely define the geological objects in the model. Also, it presents the UFO, a top-level ontology that supports our work. Chapter 4 provides an overview of the basic concepts related with the petroleum formation and the data derived from the E&P activities. It also presents more details about the earth modeling activity for petroleum exploration. Chapter 5 presents several domain ontologies that were developed in the domain of Geology and petroleum exploration in the last few years, with a special regard to the Basic Geology ontology [Mastella 2010]. This chapter also presents our modified version of the Basic Geology Ontology based on the analysis performed in its sub-ontologies. Chapter 6 describes in more details the analyzed communication standard formats, presenting the results of the analysis performed with LAS, WITSML and RESQML. The final chapter concludes by summarizing the results and contributions of this work and by pointing out possible future work.

2 INFORMATION INTEGRATION

Today, every business, organization, and individual routinely deals with a broad range of data sources. Almost any professional or business task we undertake causes us to integrate information from some subset of those sources. The concept of data or information integration is concerned with unifying data that share some common semantic but are originating from unrelated sources [Calvanese and Giacomo 2005]. This increasing need of dealing with information from multiple data sources is turning the process of information integration being the cornerstone of modern business informatics, and has made the problem of data integration arising as a new research challenge.

The so-called information society demands for complete access to available information, which is often heterogeneous and distributed. In order to establish efficient information sharing, many technical problems have to be solved. First, a suitable information source that might contain data needed for a given task must be located. This problem of finding suitable information sources is addressed in the areas of information retrieval and information filtering [Belkin and Croft 1992]. Once the information sources has been found, each of them have to work together with the system that is querying the information, i.e., the access to the data in the sources has to be provided. The problem of bringing together heterogeneous and distributed computer systems is known as *interoperability problem*.

In this chapter, we will present some results from the literature related with interoperability and, more specifically, information integration. The revision considers some historical approaches for data integration that are not notably evolving in recent years as well as the direction of research nowadays, mainly those based on ontology. Section 2.1 presents basic concepts related with information systems interoperability. Section 2.2 analyzes the types of heterogeneity problems that might arise due to heterogeneity of the data. Section 2.3 explains the process of information integration. Section 2.1 gives a retrospective of the several integration architectural solutions that have been developed over the past decades. Section 2.5 presents a comprehensive view about the alternatives for applying ontologies for data integration. Finally, Section 2.6 presents some works from literature that focus on the various techniques of semantic integration, especially in petroleum domain.

2.1 Information systems interoperability

Information interoperability is the capacity of different information systems, applications and services to communicate, share and interchange data, information and knowledge in an effective and precise way, as well as to integrate with other systems, applications and services in order to deliver new products and services [Fileto and Medeiros 2003].

In order to enable data exchanging and correct interpretation, two systems should have some degree of compatibility. Ideally, cooperative systems should be compliant with computational and application domain standards. However, this level of standardization may be impossible to attain in practice. The main reason is that while the real world is assumed to be unique, its representation depends on the intended purpose. As a result, different applications that share interest in the same real-world phenomena may have different perceptions and therefore require different representations. Also, it is important to note that every representation of reality is user-specific. This, combined with the rate of technology changes, the lack of universally accepted standards, the existence of legacy systems, or just for reasons of autonomy of each information system, has made the process of achieve interoperability a very difficult task. As a result, in many cases, the only way to reach interoperability is by publishing the interfaces, schemata and formats used for information exchange, making their semantics as explicit as possible. Thus, the systems can be properly handled by the cooperative systems.

According to Wache, in [Wache et al. 2001], interoperability has to be provided on a technical and informational level, i.e., the sharing information needs to provide full accessibility to the data and also requires that the accessed data may be processed and interpreted by the remote system. More specifically, Hasselbring, in [Hasselbring 2000], shows that information systems' interoperability must be considered from three viewpoints: application domain, conceptual design and software systems technology. The structure of a set of information systems and their interoperability in each one of these viewpoints are illustrated in Figure 2.1.

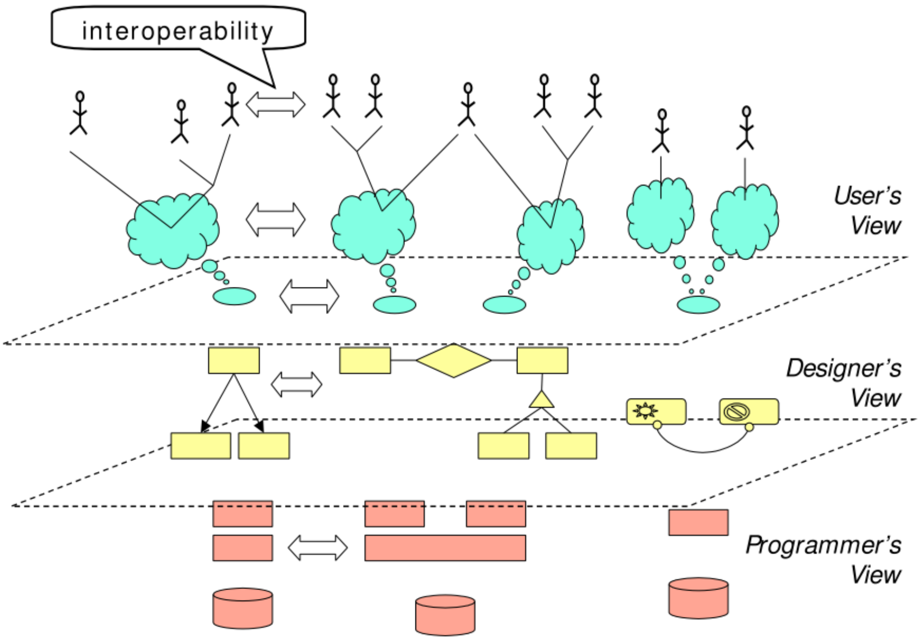


Figure 2.1: The viewpoints of information systems interoperability [Fileto and Medeiros 2003].

The *user's viewpoint* concerns the distinct views and specializations of domain experts. The *designer's viewpoint* refers to requirements modeling and system designs. The *programmer's viewpoint* refers to systems implementation. Interoperability must be achieved in all these viewpoints, i.e., users of a system must understand information coming from another system, the system design must accommodate the “foreign” data, and the computer programs must automate information exchange.

The distribution of information, however, is just one of the problems that must be faced. The fact that the systems are often developed by different agencies with different points of view and vocabularies, leads to problems of heterogeneity. These problems should be found in communication processes between interoperable systems. In these cases, interoperability refers to various interactions between information from different sources, including the task of data integration. Next section presents some problems that might arise due to heterogeneity of the data.

2.2 Types of heterogeneity conflicts

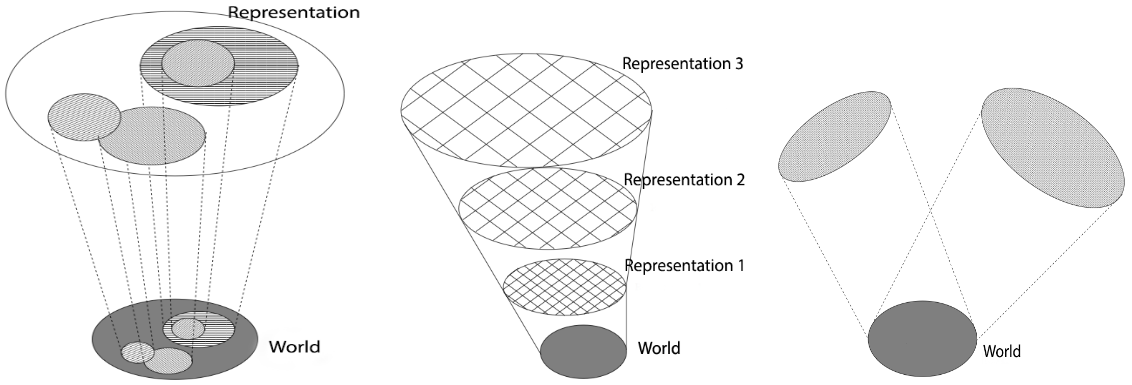
The concept of *heterogeneous data* is used for a long time to classify those data that present differences in their representation or interpretation, although referring to the same reality [Litwin and Abdellatif 1986]. *Data heterogeneity conflicts* are the incompa-tibilities that may occurs among distinct data sets. According to [Sheth and Larson 1990], the most widespread way to characterize data heterogeneity is to separate representation from interpretation concerns. *Representational conflicts* refer to syntactic or structural discrepancies in the portrayal of heterogeneous data. *Semantic conflicts* refer to disagreement about the meaning, interpretation or intended use of the same or related data.

Both representational and semantic conflicts may occurs in any level of abstraction (instance, schema and data model). Thus data heterogeneity conflicts can also be classified according to the following categories [Härder, Sauter and Thomas 1999]:

- *Data conflicts*, that are discrepancies in the representation or interpretation of instantiated data values, which can differ in their measurement unit, precision and spelling;
- *Schema conflicts*, that are differences in schemata due to alternatives to depict the same reality, such as using distinct names for the same entities or modeling attributes as entities and vice-versa;
- *Data versus schema conflicts*, which are disagreements about what is data and metadata. For example, a data value under one schema can be the label of an entity or attribute in another schema;
- *Data model conflicts*, which result from the use of different data models.

These heterogeneity problems and integration conflicts have been the subject of several taxonomies, mainly within the distributed database systems community (e.g., [Kim and Seo 1991] and [Kashyap and Sheth 1996]). Among the three main conflict categories presented (syntactic, structural and semantic), we will focus on the semantic heterogeneity. In a broader view, semantic conflicts are harder to disambiguate, since they occur whenever “two contexts” do not use the same interpretation of the information.

A semantic heterogeneity can vary on account of the dimension of the representations [Benerecetti, Bouquet and Ghidini 2000] (Figure 2.2). When a representation covers a subset of a more comprehensive domain of interest, it is called as a **partial representation**, as demonstrated in Figure 2.2a, where the small circles represent different portions of the same domain of interest (the circle below). Each portion can overlap or be included in other portions. An **approximate representation**, on the other hand, corresponds to the case when the representation abstracts some aspects of a given domain of interest. In Figure 2.2b, the circles above are representations of the world at different levels of approximation (or granularity). Finally, a **perspectival representation** corres-ponds to the case when the representation encodes a spatio-temporal, logical, cognitive or functional point of view on a domain of interest (Figure 2.2c).



(a) Partiality (b) Approximation (c) Perspective

Figure 2.2: Three dimensions of heterogeneity (adapted from [Benerecetti, Bouquet and Ghidini 2000]).

In order to achieve semantic interoperability in a heterogeneous information system, the meaning of the information that is interchanged has to be understood across the systems. Taking the contextual perspective into account, Goh et al., in [Goh et al. 1999], identifies three main causes for semantic heterogeneity:

- *Confounding conflicts* occur when information items seem to have the same meaning, but differ in reality, e.g. due to different temporal contexts;
- *Scaling conflicts* occur when different reference systems are used to measure a value. Examples are different currencies;
- *Naming conflicts* occur when naming schemata of information differ significantly. A frequent phenomenon is the presence of homonyms and synonyms.

Many authors stipulate that the use of ontologies for expressing the semantics of the sources is a possible approach to overcome the problem of semantic heterogeneity. Also, according to [Uschold and Gruninger 1996], that mention interoperability as a key application to ontologies, many ontology-based approaches to information integration have been developed in order to achieve interoperability. In this case, where ontologies are used to describe the heterogeneous data sources, we are talking about ontology-based integration.

In this context, the problem of integrating different sources of information is shifted to integrating different ontologies corresponding to these sources or domains. Thus, the main tasks of this integration process are: ontology merging, ontology mapping, and ontology integration [Klein 2001, Euzenat, Shvaiko et al. 2007]. The process of ontology merging creates a new ontology from two or more existing ontologies with overlapping parts. The second task, ontology mapping, relates similar concepts or relations, according to some metrics, from different sources to each other by an equivalence relation. The mapping process results in a virtual integration. Finally, the process of ontology integration denotes the inclusion of one ontology into another one by usually using bridge axioms.

All these tasks also have to face semantic heterogeneity problems, or in this case, ontological heterogeneity [Visser et al. 1997]. This kind of heterogeneity occurs if two systems make different ontological assumptions about their knowledge of the domain. According to [Chalupsky 2000, Visser et al. 1997, Wiederhold 1994], these mismatches between ontologies can be divided into two main categories: language mismatches and ontology mismatches. Language mismatches occurs when ontologies written in different languages are combined. When the mismatches are find in the same language through the use of abbreviations, acronyms, or punctuations, for instance, they are called linguistic mismatches [Madhavan, Bernstein and Rahm 2001]. Ontology mismatches occurs when ontologies describing overlapping domains are combined. This category describes different representations of a real world domain. Therefore, it does not matter whether the language is the same or different, but whether the concepts in the ontologies are different for a real world entity.

Visser et al., in [Visser et al. 1997], also performed a classification where ontology mismatches are classified according to two sub-processes realized during the creation of an ontology: conceptualizing a domain, and explicating the conceptualization. Conceptualization mismatches may appear when two or more conceptualizations differ in the ontological concepts or in the way these concepts are related. Explication mismatches, on the other hand, are not defined on the conceptualization of the domain but on the way the conceptualization is specified.

2.3 The information integration process

Differently of what really means the information integration process for business, research on information integration has focused individually on particular aspects of integration, such as schema mapping or replication. According to Haas, in [Haas 2006], for business, information integration is really a process, with four major tasks: understanding, standardization, specification and execution.

Understanding the data is the first task in information integration. It is during this task that relevant information may be discovered, such as keys, constraints, data types, and others. The analysis of this information is important to assure quality and to determine statistical properties (e.g., data distributions, frequent values, inconsistent values). The integrator may look for relationships among data elements, such as foreign keys, or redundant columns. Also, for unstructured data, the integrator may use metadata to understand the meaning of the data. Both tools and end users leverage it to find and understand the data to be integrated. It is also produced as the output of analysis, to be exploited by later tasks in the process.

Standardization is the second task, which leverages the work of the previous task. Typically, it determines the best way to represent the integrated information. This includes designing the integrated schema (know as "target"), deciding at the field level the rules on how data is represented, i.e., what the standard representation should be and even defining the terminology and abbreviations to use. In addition, other rules that specify how to cleanse or repair data may be provided. These rules should help troubleshoot issues about how to handle inconsistent or incomplete data and how to identify data that refers to the same objects.

The third task, specification, produces the artifacts that will control the actual execution process. As a result, the techniques and technologies used in this task are related to the execution engine(s) chosen. Mapping tools, for example, which specifies the relationship between source(s) and target(s), can generate a query or other executable artifact (e.g., XSLT) that would produce data in the desired target form. However, as the specification is usually part of actually configuring an integration engine to do the desired integration, determining the execution engine should be thought of as part of specification.

Finally, the fourth task, execution, is where the integration actually happens. According to [Haas 2006], there are three ways in which integration can be accomplished: materialization; federation; and/or indexing. **Materialization** creates and stores the integrated data set through various techniques, such as *Extract/Transform/Load* (ETL) tools, *replication*, and *caching*. The first technique extracts data from one or more data sources, transform them as indicated in the script, and then store the result in another data source. The second one makes and maintains a copy of data, often differentially by reading database log files. And the third technique captures query results for future reuse. In the other hand, **federation** creates a virtual representation of the integrated set, only materializing selected portions as needed. Federation is a form of **mediation**, which refers to an integration technique in which requests are sent to a "mediator" process which does routing and translation of requests. The last way, **indexing** (or **search**), takes a different approach, creating a single index over the data being integrated. Usually, this approach is used for unstructured data, and represents a partial materialization, since typically the index identifies relevant documents, which will be fetched dynamically according the user's request.

It is important to note that all these four tasks, when implemented, are interdependent. Also, existing tools often support several of these tasks (or at least part of them). In practice, these tasks may be overlapped, since it is not necessary to have a complete understanding before starting to standardize. Likewise, a particular integration may not require all of the subtasks for any task. Especially in really simple cases, some tasks may seem to vanish altogether. Therefore, the integration process is iterative and never-ending, since changes are constant. There is always another source to deal with, a new application with new requirements, an update to some schema, or just new data that requires analysis.

2.4 System-based information integration

Even after decades of research, integrating information remains an omnipresent and extremely expensive challenge. Large enterprises spend a great deal of time and money on information integration. In the expenses are included the software purchases and also integration activities, that cover any form of information reuse, such as moving data from one application's database to another, translating messages for business-to-business e-commerce, and providing access to structured data and documents via a Web portal. In relation to software, they are becoming increasingly frequent, making the process of information integration much easier [Bernstein and Haas 2008].

There are many architectural approaches that can be used to solve problems related to information integration. Also, many software vendors have offered numerous tools to reduce the effort and the cost of integration and to improve its quality. Moreover, due to the complexity of the process of information integration, and also its multifaceted task, many of these tools are highly specialized. As a result, finding the most suitable tool for a particular integration problem requires a general understanding of the available tools, which can be confusing. Consequently, choosing the most appropriate tool is becoming increasingly more challenging. Furthermore, if the choice is not suitable, instead to facilitate, the chosen tool may increase the time required to provide information integration for a new application [Haas 2006].

In this Section, we will summarize some integration approaches, along with the general types of products used. Our main goal is to provide a brief explanation of the main features and uses of each integration approach more used by software vendors. Thereby, assisting the professionals involved in the integration process to choose the most appropriate tool for your needs.

2.4.1 Data warehouse

A *Data Warehouse* (DW), by definition, is a database that consolidates data from multiple sources [Chaudhuri and Dayal 1997]. A DW system stores copies of data from various sources. Moreover, these data could be updated regularly. As a result, these data could be overlapping and may have inconsistent information, requiring a previous data cleaning to reconcile such differences in a DW. Furthermore, each data source integrated by the DW may have a database schema (that is, data representation) that differs from the warehouse schema. Thus, each data source must be reshaped into the common warehouse schema.

The use of ETL tools, that address this problem [Kimball and Caserta 2004], is helping by simplifying the programming of scripts to transform the data to suit the warehouse schema. An ETL tool typically includes a repertoire of cleansing operations (such as detection of approximate duplicates) and reshaping operations (such as Structured Query Language [SQL]-style operations to select, join, and sort data). The tool may also include scheduling functions to control periodic loading or refreshing of the data warehouse.

Some applications require customized ETL tools to produce a DW that holds the master copy of critical enterprise reference data, such as information about customers or products. In this case, we are talking about customized ETL tools for master data management, where the mater data is first integrated from multiple sources and then itself becomes the definitive source of that data for the enterprise. Sometimes, theses master data-management tools include domain-specific functionality, like standardization and cleansing functions to validate and correct customer or vendor information (e.g., postal codes based in the address).

A DW system significant differs of a database system, especially in relation to data processing. A database system uses an online transaction processing, or **OLTP**, which is application oriented, has its emphasis on efficiency through a fast and secure processing, and its operations do not alter the structure of the data. On the other hand, a DW system uses an online analytical processing, or **OLAP**, which is a decision-making oriented process, has it emphasis in analytical flexibility, and its operations alter the structure of the data. OLAP is the DW basic activity. It does not help having a well-defined DW unless we have analytical tools that allow queries to decision-making purposes. The Figure 2.3 illustrates the uses of the tools explained.

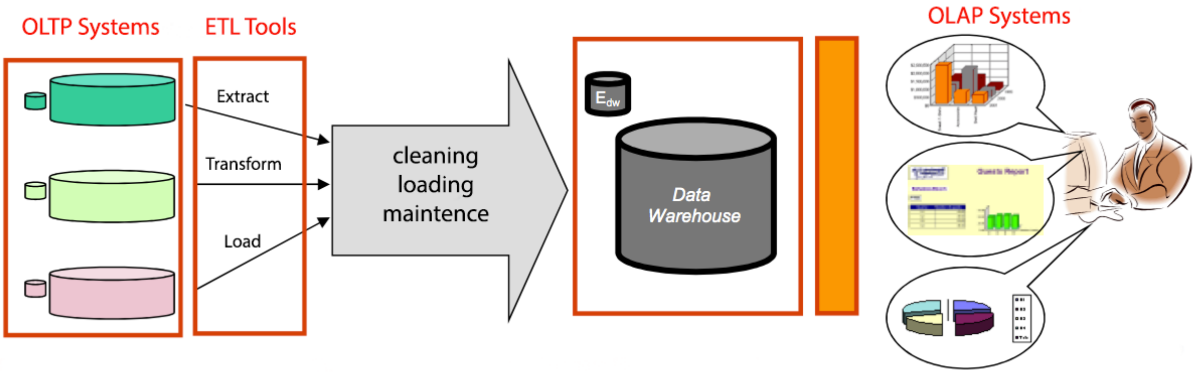


Figure 2.3: Data Warehouse related tools and systems.

2.4.2 Virtual data integration

Differently of a warehouse integration, which materializes the integrated data, a *Virtual Data Integration* gives the illusion that data sources have been integrated without materializing the integrated view. A virtual data integration system offers to the users a mediated schema, where they can perform queries. The implementation is often called a query mediator [Wiederhold 1992] or an enterprise-information integration (EII) system [Halevy et al. 2005, Morgenthal and Kernochan 2005]. It translates the user's queries on the data sources and integrates the result of those queries so that it appears to have come from a single integrated database. This technology is more recent and currently less popular than data warehousing.

An EII system might be use to integrate data from heterogeneous databases that cover related subject matter, but that may use different database systems and structure the data using different schemata. To handle with this heterogeneity in an EII system, a designer needs to create a mediated schema that covers the desired subject matter in the data sources and maps the data source schemata to the new mediated schema. However, data cleansing and reshaping problems also appear in the EII context, but the solutions are different in EII systems because data must be transformed as part of query processing rather than via the periodic batch process associated with loading a data warehouse.

As described in [Bernstein and Haas 2008], depending on the types of data sources to be integrated, EII products vary. Some products, for example, focus on integrating SQL databases, some on integrating Web services, and some on integrating bioinformatics databases.

2.4.3 Message mapping

In order to integrate applications developed independently, a special software can be required, i.e., a computer software that provides services to software applications beyond those available from the operating system (a middleware). Middleware makes it easier for software developers to perform communication and input/output, so they can focus on the specific purpose of their application. Specifically, a *message-oriented middleware* helps integrate independently developed applications by moving messages between them.

The resulting product can be of three types: an *enterprise-application integration* (EAI) system [Alonso et al. 2004]; an *enterprise service bus*, and a *workflow system*. The first is when the messages pass through a broker. The second is when the broker is avoided through all application's use of the same protocol, such Web services. And the third is when the focus is on defining and controlling the order in which each application is invoked (as part of a multistep service). Another form of information integration, a message-translation service, is also needed beyond the protocol-translation and flow-control services provided by these products.

Usually, a message-mapping tool is used to translate messages from one application to another. Such a tool offers a graphical interface to define translation functions, which are then compiled into a program to perform the message translation. Similar mapping tools are used to help relate the schemata of the source databases to the target schema for ETL and EII and to generate the programs needed for data translation.

2.4.4 Object-to-relational mappers

The majority of application programs today are written in an object-oriented language, but the data accessed by them is usually stored in a relational database. There are many approaches for mapping applications to databases that requires integration of the relational and application schemata. In all these approaches, differences in schema constructs can make the mapping even more complicated.

In order to simplify this problem, an *Object-to-Relational Mapper* can be used to offer a high-level language in which to define mappings [Melnik, Adya and Bernstein 2007]. The resulting mappings can be compiled into programs that will translate the queries and updated over the object-oriented interface into queries and updates on the relational databases.

2.4.5 Document management

The vast majority of companies have to deal daily with a large amount of information. This information, usually, is relevant to critical business functions (e.g., product designs, marketing plans, pricing, and development schedules). Furthermore, to promote collaboration and avoid duplicated work in a large organization, this information needs to be integrated and published. However, much of this information is contained in documents, such as text files, spreadsheets, and slide shows that contain interrelated information. In this case of integration, we are talking about a *Document Management* approach, that may simply involve making the documents available on a single Web page (such as a portal, that will be presented in Section 2.4.6) or in a content-management system, possibly augmented with per-document annotations, i.e., with document's metadata (author and status, for example). This integration process also may mean the combination of the information from these documents into a new document, such as a financial analysis.

Using a document management approach, whether or not they are collected in one store, they can be indexed to enable keyword search across the enterprise. It could be useful, in some applications, to extract structured information from documents, such as customer name and address from email messages received by the customer-support team. This ability of extracting structured information may also allow business to integrate unstructured documents with preexisting structured data.

2.4.6 Portal management

A *Portal* is a Web site built to integrate related information by presenting them side-by-side, i.e., on the same screen. For example, the home page of a financial service Web site typically presents market prices, business news, and analyses of recent trends. The person viewing it does the actual integration of the information.

The design of a portal requires a mixture of content management and user-interaction technology. The first one allows the users of the portal to deal with documents and databases. The second is important to present the information to the users in a useful and attractive ways. These technologies could be packaged together into a product for portal design [Firestone 2003]. However, usually they are selected separately, based on the required functionality of the portal and the preferences and experience of the developers who assemble it.

2.5 Ontology-based integration approaches

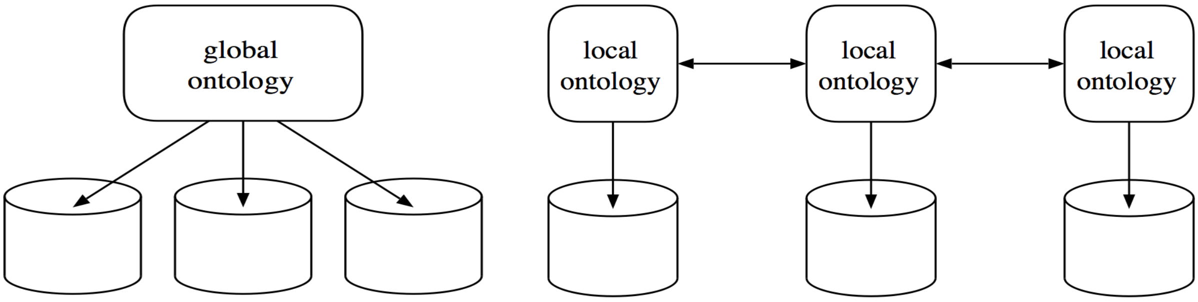
Traditionally, ontologies are applied in information integration as a supportive resource for some specific aspects of semantic integration such as schema matching. In contrast, in the concept of ontology-based integration, ontologies are utilized to overcome different types of semantic heterogeneity on data and schema level, as explained previously.

There are, however, different ways of how ontologies can be employed. Wache et al., in [Wache et al. 2001], provides an overview of the three different conceptual design alternatives for applying ontologies for data integration: single ontology approach, multiple ontology approach, and hybrid ontology approach (as presented in Figure 2.4).

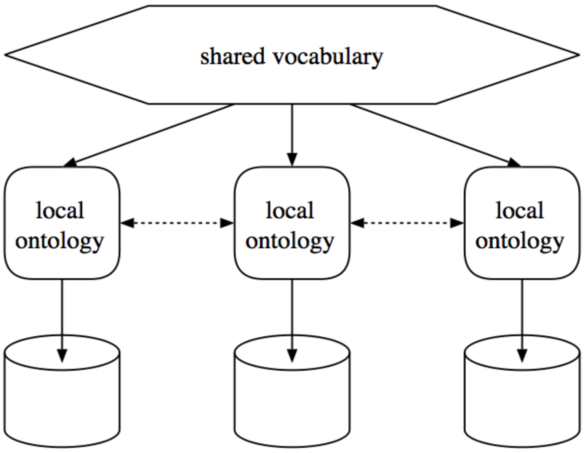
Usually, ontology-based integration is a top-down process, on which, first, a context-specific ontology is designed, which take the role of a global schema, to then relate (i.e., map) the heterogeneous datasets to this global ontology. The application of subsumption or inference algorithms can then discover further semantic relationships between the data from the different systems. This *single ontology approach*, represented in Figure 2.4a, has as main advantage the quick development. However, its common problem is the request of managing a global integrated ontology, which involves administration, maintenance, consistency and efficiency problems that are very hard to solve.

Alternatively, the local datasets can be mapped to their own local ontologies, which are then harmonized in a further ontology alignment step. Using this *multiple ontologies approach*, represented in Figure 2.4b, the addition of a new source will be easier than in the single approach, since it demands the built of only one ontology. However, in practice, it is difficult to compare different source ontologies without a common vocabulary, and inter-ontology mappings need to be defined.

A *hybrid ontology approach*, represented in Figure 2.4c, would combine the use of local ontologies and a global ontology. This approach was developed to overcome the drawbacks of the other approaches. In the hybrid approach, each information source is described by its own ontology, which in turn is built using basic primitives described in a global shared vocabulary. When the ontologies need to be developed from scratch, this approach brings advantages. However, the shared vocabulary can be a problem if it increases rapidly. Also, the reuse of existing ontologies can be a problem in this approach, since they need to be rewritten to refer to the shared vocabulary.



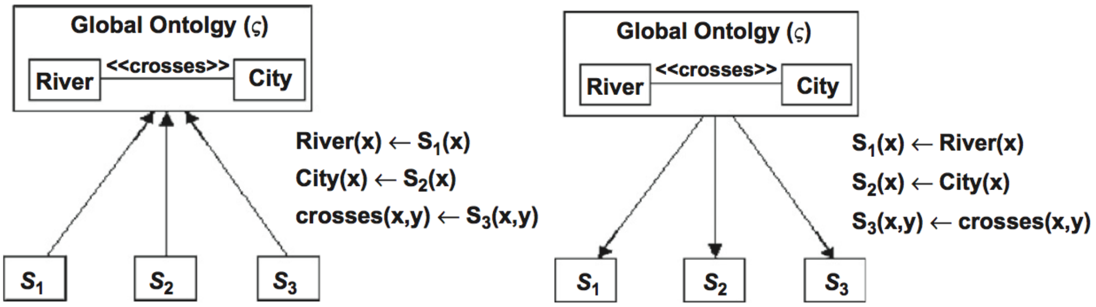
(a) Single ontology approach (b) Multiple ontology approach



(c) Hybrid ontology approach

Figure 2.4: Three ways for connecting ontologies to information sources (adapted from [Wache et al. 2001]).

Based on this hybrid approach, two new approaches for specifying mappings in an integrated system have emerged [Calvanese, Giacomo and Lenzerini 2002, Calvanese and Giacomo 2005]: global-as-view (GAV), and local-as-view (LAV) (Figure 2.5). The GAV approach is based on the definition of global concepts as views over the sources, i.e., each concept of the global view is mapped to a query over the sources. Figure 2.5a presents how the mappings are defined over the sources (S1-S3) giving a global ontology (G). The LAV approach is based on the definition of the sources as views over the global view, i.e., the source's information is described in terms of this global view, as presented in Figure 2.5b. According to advantages and disadvantages of each approach, that have to be considered when an integration process is initiated [Calì et al. 2003], both approaches have their ups and downs. With respect to restrictions over the sources, the LAS approach takes advantage, but in the GAV approach it is easier to define restrictions over the global view. With respect to scalability, the LAV approach allows that a new source is added in the integrated system without requiring changes in the global as view, whereas the GAV approach requires changes over the global view when a new source is added. Considering query processing, LAV may demand reasoning mechanisms in order to answer them, whereas in GAV conventional mechanisms can be implemented.



(a) Global-as-view approach (b) Local-as-view approach

Figure 2.5: GAV and LAV approaches [Buccella, Cechich and Fillottrani 2009].

In order to take advantage of both approaches, a new approach has emerged [Friedman et al. 1999, Fagin et al. 2005]: global-local-as-view (GLAV). This approach can be considered as a generalization of both GAV and LAV [Calì 2003] and consists of association views over the global schema to views over the sources, i.e. mappings can be specified in both directions.

2.6 Related work in semantic integration in petroleum domain

There are many works in the literature that focus on the various techniques of semantic integration. Among them, we can highlight the reviews of database schema matching approaches developed by [Rahm and Bernstein 2001] and [Doan and Halevy 2005]; the surveys of ontology-based approaches for information integration developed by [Wache et al. 2001] and [Noy 2004]; and the work carried on by [Kalfoglou and Schorlemmer 2003] that focus on the current state of the art in ontology matching.

Besides these works, Gagnon, in [Gagnon 2007], proposes an ontology-based information integration with a local to global ontology mapping as an approach to the integration of heterogeneous data sources.

Soma, in [Soma et al. 2008], proposes the use of ontologies or the information schemata that model various elements from the domain, and a knowledge base, which is a central repository of the instance information, to address problems commonly encountered in model based oil field operations (as explained in Chapter 1). This work is part of the Integrated Asset Management (IAM) project at the Chevron-funded Center for Interactive Smart Oilfield Technologies at the University of Southern California, Los Angeles[1]. The current focus of the IAM project is on enabling model-driven reservoir management.

Mastella, in [Mastella 2010], describes an architecture for ontology-based integration of engineering models. Her approach, using a hybrid structure of local and global ontologies, uses *annotations* to connect the local data sources representations (*local views*) to their ontological

---

[1] Cisoft IAM Project, http://pgroup.usc.edu/wiki/Smart_Oilfield

meaning (*local ontologies*). The mapping between the global and the local ontologies are called *local-to-global* alignments.

Calvanese, in [Calvanese et al. 2007] describes Mastro-I, a data integration management system designed in order to maintain data complexity within reasonable bounds. It relies on the IBM product Infosphere Federation Server 3 to access source databases. In other work [Calvanese et al. 2011], the same group describes a database integration case using Mastro-I, in which five different data models were used, including XML-based and relational databases. The integration was made in two steps: first the different data models were combined using the InfoSphere Federation Server; then the Mastro-I system was used to map those entities into concepts, thus achieving data integration. In this architecture, there are two layers of heterogeneity solving: first, all relational data is mapped at the Federation Server, and then mapped into DL concepts, where integration is actually achieved.

Ge, in [Ge et al. 2011], presents a system framework for ontology-based knowledge integration and sharing in petroleum exploration domain conceived to minimize the complexity of heterogeneous data, and enhances power of knowledge integration and information sharing among different operational units. They designed a framework divided into two parts: ontology-based knowledge integration, which includes ontology mapping and ontology merging, and ontology-based knowledge sharing that includes concept searching and concept selecting. The concept searching consists of concept matching and purpose identification, and concept selecting consists of concept expansion and knowledge selection.

Buccella et al., in [Buccella, Cechich and Fillottrani 2009], presents a survey of the current approaches of ontology-driven geographic information integration. In other work [Buccella et al. 2011], the same group proposes a novel system (called GeoMergeP) to integrate geographic sources by formalizing their information as normalized ontologies. Their integral merging process — including structural, syntactic and semantic aspects — assists users in finding the more suitable correspondences.

3 CONCEPTUAL MODELING AND ONTOLOGIES

Conceptual modeling is concerned with identifying, analyzing and describing the concepts of a domain, as well as their semantic relationships. This is done with the help of a modeling language that is based on a small set of basic meta-concepts (forming a metamodel). On the other hand, ontological modeling is concerned with capturing the relevant entities of a domain in an ontology of that domain using an ontology specification language that is based on a set of domain-independent ontological categories (forming an upper level ontology) [Guizzardi et al. 2004].

In reservoir analysis, the modeling process is carried out by the geologist, who tries to capture and represent parts of the geological reality of the subsurface. In order to cope with this task, the geologist creates mental models of his comprehension of the subsurface geology. A mental model abstracts the relevant aspects of the objects, omitting those considered irrelevant for the task in hands. The *conceptualization* represented in an earth model is an abstract entity that only exists in the geologist's mind. Consequently, what is finally represented in the model is the geologist's idea about the reality and not the reality itself. In order for a model to be understood by a community of professionals along with the modeling chain, the set of concepts in the geologist's mind and the corresponding external representation need to share the same respective meanings accepted by the community. These shared concepts and the agreed representations are what we call an *ontology*.

The term "ontology" has its origin in philosophy, where it is understood as a particular system of categories accounting for a certain vision of the world. As such, this system is independent of the language used to describe it [Guarino 1998]. In the context of Computer Science, the term ontology usually designates an artifact that refers to some shared conceptualization and to its external representation in a computer processing language [Gruber 1992]. However, Nicholas Guarino [Guarino 1998] reinforces that an ontology is a logical theory accounting for the *intended meaning* of a formal vocabulary utilized for representation. Indeed, ontology is the theory that helps us in keeping the correspondence between the geological reality and the models produced from this reality.

Gruber, in [Gruber 2008], tells that the AI community came to use the term ontology in the 1980's to refer to both a theory of a modeled world and a component of knowledge systems. And then, in the early 1990's, ontologies were identified as a key component for creating interoperability standards. At that moment, a globally accepted definition of an ontology in computer science was offered by [Studer, Benjamins and Fensel 1998], which states that an ontology is a formal explicit specification of a shared conceptualization.

3.1 Foundational ontologies

Foundational ontologies are meta-models that orient the way in which some conceptualization should be identified and modeled in a formal representation to build an artifact representing the domain ontology. Domain ontologies refer to the concepts and meanings that a group of people needs to share when they communicate for solving problems in some restricted domain. The object of foundational ontologies is the study of *universals*, which are the intended meaning or the abstraction of the main properties that characterize a set of *individuals* recognized in the real world. Universals are equivalent to classes or objects in modeling languages, while individuals are their instances. Ontological analysis also provides an important distinction among the entities in terms of their behavior with refers to time: *endurants* and *perdurants*. Endurants correspond to individuals wholly present whenever they are present, i.e., they *are in time* (e.g., rock, reservoir and petroleum). Perdurants correspond to individuals composed of temporal parts, they *happen in time* in the sense that they extend in time accumulating temporal parts (e.g., deposition, oil migration or earthquake). Endurants have a relation of participation in perdurant occurrences, and these are commonly responsible for the creation of those. Also, foundational ontologies deal with ontological properties drawn from Philosophy (e.g., identity, dependence, unity), i.e., with formal aspects of universal and individuals irrespective of their particular nature. The properties applied in this work are essentiality, rigidity, identity, and unity as proposed by Guarino in [Guarino and Welty 2009].

A property of an entity is *essential* to that entity if it must hold for it. It is a definitional property that explains what causes an individual be an instance of a specific universal. For example, it is essential for a mineral being solid and having a crystalline structure at a normal temperature. Some essential properties are *rigid*, meaning that they are essential to all their possible instances. If a mineral ceases of being solid, for instance, it will stop being a mineral, since being solid is a rigid property for a mineral that helps in identifying and defining it. Some properties are *anti-rigid*. For example, a reservoir can stop being considered a reservoir because it was fully depleted or became non-economic, but the concrete entity that was before a reservoir (the rock unit) is still there. Rigid properties are important because they identify the objects that are present in all geological models and can be mapped from one model to another.

Identity refers to the issue of being able to recognize individual entities in the world as being the same (or different), and *unity* refers to the issue of being able to recognize all the parts that form an individual entity. Both ontological properties are crucial for geological interpretation and are used by geologists to interpret stratigraphic and structural correlations. Identity involves, for instance, the rigid properties of a rock unit that need to be considered for deciding, whether a body of rock corresponds or not to some geological unit although these two entities were possibly described in two different places (a wellbore and an outcrop, for example). Unity refers to the problem of describing the parts and boundaries of an object, so that we can decide what is part of the object, what is not, and under what conditions the object is a whole. For example, water is a concept whose instances were not wholes, except if they are limited by an instance of some other object, such as a cup, a bottle or a lake. In the same sense, rocks have no unity, since a rock can only be individualized by a core, a sample or a geological unit.

3.1.1 Unified Foundational Ontology (UFO)

Guizzardi, in [Guizzardi 2005], states that foundational ontologies, also called *upper level ontologies*, are meta-ontologies that describe a set of ontological categories of high-level abstraction, domain independent, and constitute a general grounding for multiple ontologies more elaborate and specific to particular domains. In this sense, foundational ontologies can be used to describe the categories used to build conceptual models of lower level, such as domain-specific ontologies, making them a suitable reference for the development of languages. Thus, foundational ontologies act as guides for modeling decision making in a conceptual modeling process, explaining and justifying the meaning of the models, increasing the understandability and reusability of their potential [Guizzardi 2005].

Among the available foundational ontologies highlights the UFO (Unified Foundational Ontology), who proposed the unification of concepts addressed by other foundational ontologies, such as OntoClean/DOLCE and GFO/GOL, in order to offer solutions to unsolved problems in these other ontologies [Guizzardi and Wagner 2010]. According to [Guizzardi, Falbo and Guizzardi 2008], UFO has been applied successfully to evaluate, re-design and integration of models of conceptual modeling languages and also to provide real-world semantics to their modeling constructs.

Figure 3.1 presents the initial distinctions of UFO, starting with its most generic concept, *Thing*, and its distinction between two fundamental entities: *Urelement* and *Set*. A *Urelement*, which is an entity that is not a set, is divided between the categories of individuals and universals. *Individuals*, also called *Particulars*, are entities that exist in reality, as a person, a table, etc. On the other hand, *Universals* are standard features instantiable by different individuals, so that they can be understood as high-level abstractions that characterize different classes of individuals. In this sense, Person is a universal that describes the common characteristics of the different individuals (particular persons) of this type, for instance, name and fingerprint.

As previously mentioned, UFO universals are specialize in *Endurant Universal* and *Perdurant Universal*. This distinction can be understood in terms of the behavior of their individuals with refers to time. Endurant Universals are those whose individuals are wholly present whenever they are present, i.e., individuals that preserve their identity over the passage of time. Perdurant Universals are those whose individuals are composed of temporal parts. The first part, related to Endurant Universals, corresponds to the core fragment of UFO, called UFO-A and described in [Guizzardi 2005]. The second part, related to Perdurant Universals, corresponds to the UFO fragment called UFO-B, which increases the scope of the UFO-A systematizing concepts and temporal relationships between objects and events. UFO-B was first described

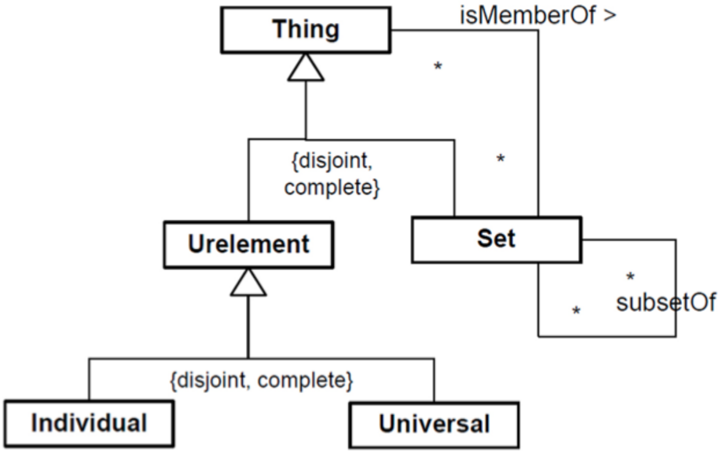


Figure 3.1: UFO fragment that represents the fundamental distinctions between urelement and set, and universal and individual [Guizzardi 2005].

in [Guizzardi, Falbo and Guizzardi 2008] and recently formalized in [Guizzardi et al. 2013]. A third part of UFO is called UFO-C, which was developed over UFO-A and UFO-B and constitutes an ontology of social notions and intentional agents. The UFO-C was described in [Guizzardi et al. 2007], [Guizzardi and Guizzardi 2011] and [Souza 2006]. In this work, we emphasize the concepts of UFO-A and UFO-B .

Figure 3.2 presents the main distinctions between Endurant Universals, which are divided into two categories: *Substantial Universals* and *Moment Universals*. Substantial Universals are universals whose individuals are existentially independent, have spatio-temporal properties and are founded in matter (e.g., a rock). Moment Universals are universals whose individuals are existentially dependent, so that they can only exist in other individuals and that, therefore, are said inherent in these individuals (for example, a deposition). According to UFO, the relationship between a Substance Universal and a Moment Universal is called participation.

Moreover, moment universals are divided into *Intrinsic Moments* and *Relators*. Intrinsic moments represent individuals that depend on just one individual. For example, the intrinsic moment Color, where each of its individuals depend uniquely to a single individual (e.g., the color of an apple). On the other hand, relators represent individuals that depend on many individuals (e.g., a job, which involves an employer and an employee) and materialize the notion of material relation in conceptual models.

Relations, which are entities that glue together other entities, are divided into two categories: *formal relations* and *material relations*. Formal relations occur directly between two or more entities directly, i.e., do not require the intervention of another individual. Formal relations include relations of the UFO's mathematics superstructure, such as *existential dependence*, *part of*, *subset of*, *instantiation*, etc., and the relations of comparison, established in the domain, such as *greater than*, *older than*, among others. Formal relations of comparison are completely founded in intrinsic moments being compared. Moreover, the material relations have their own

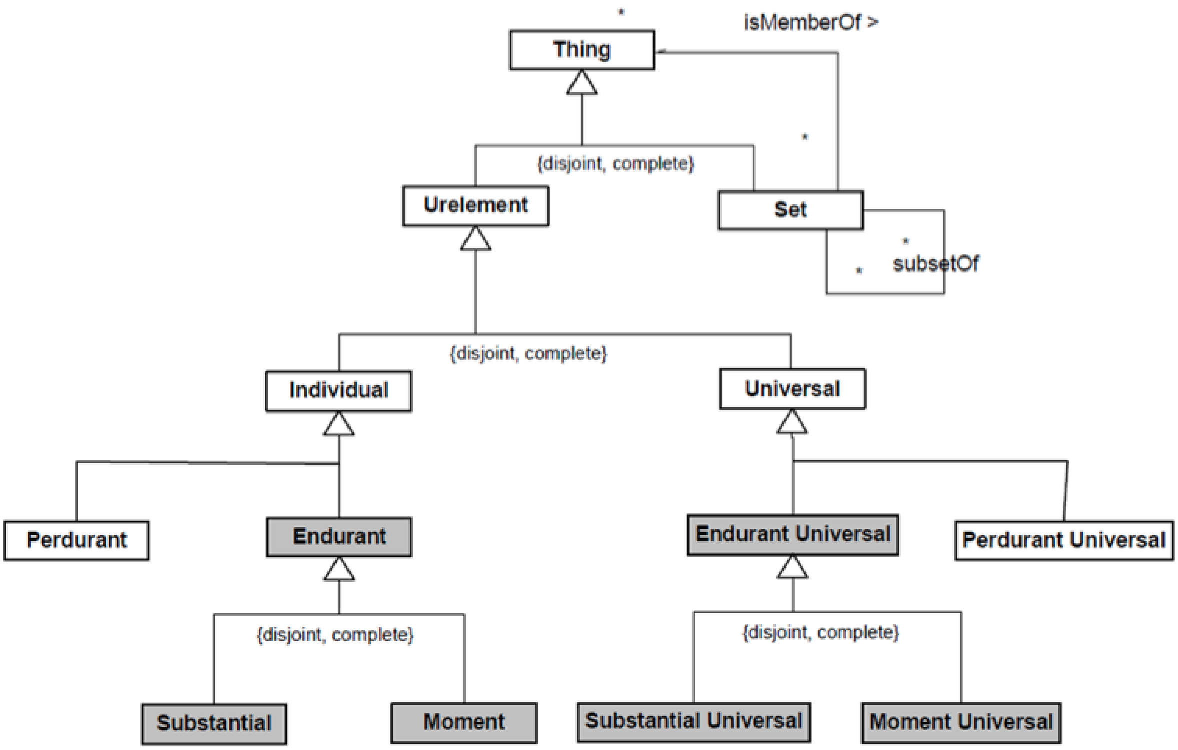


Figure 3.2: UFO fragment that represents the distinctions between endurant and perdurant universals and individuals and between substantial and moment universals and individuals [Guizzardi 2005].

structure material. For example, the relations *work at*, *be enrolled in*, etc. In this case, to make a material relation to happen, such as the relation *be treated in* between John and a medical unit, there must be another entity, called treatment, to mediate John and medical unit. Those entities that mediate other individuals are called relational moments.

Some Intrinsic Moments Universals are *Quality Universals*, which represent properties in conceptual models. A quality universal characterize other universals and is related to *Quality Structure*, which represent a set of all values that a quality can assume. A quality structure can be a *Quality Domain* or a *Quality Dimension*, such that a quality domain can be composed of various quality dimensions. Thus, the property *Weight*, for example, is associated with a one-dimensional structure comprising the non-negative part of the real numbers line, so that this quality structure constitutes a quality dimension. On the other hand, the universal Color is associated with a multidimensional structure, which includes brightness, contrast and saturation, so that this structure is a quality domain. The perception or conception of a moment can be represented as a point in a quality structure. This point is called *quale*. The entities comprising Quality Structure and Quale are, along with sets, numbers and propositions, examples of *Abstract Particulars*. Intrinsic Moments Universals that are not associated with quality structures are called *Modes*. Modes represent individuals who are existentially depend of another individual, but can be conceptualized in terms of multiple quality dimensions separable, as occurs with substantial individuals. Thus, modes may have other intrinsic moments inherent in them (e.g., symptoms, desires, thoughts and skills). Figure 3.3 summarizes the relationship between a substantial, one of its qualities and the associated quale.

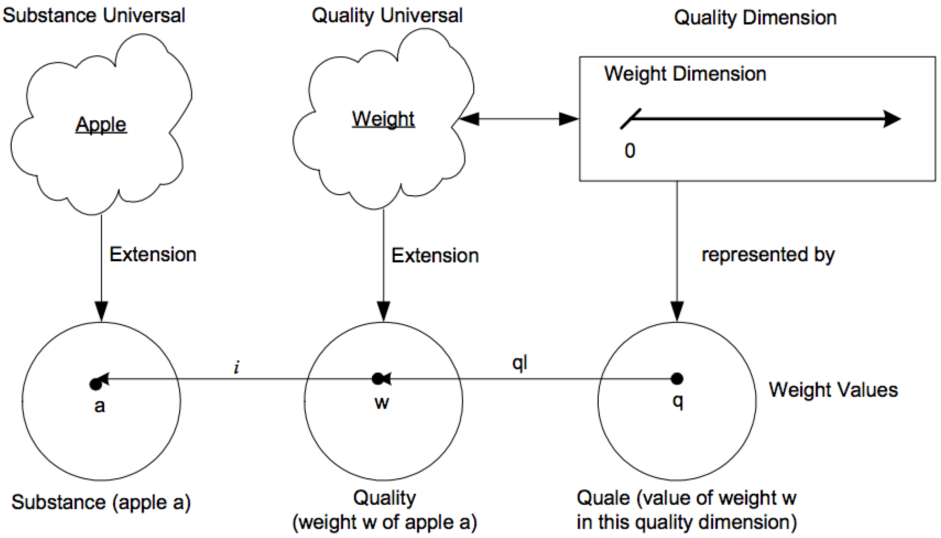


Figure 3.3: Representation of the relationship between substance universals, quality universals and quality dimensions. In this figure, *i* represents a inherence relation, in the sense that the quality individual *w* is inherent to the substance individual *a*. While *ql* represents the relation between the quality individual *w* and the quale *q*, which represents its value, in the quality dimension [Guizzardi 2005].

Substantial Universals are divided into two main specializations: *Sortal Universals* and *Mixin Universals*. Sortal Universals are Substantial Universals that provide a principle of identity (PI), which allows us to tell when two instances of a Universal are the same (as a puppy and the same adult dog), and a principle of unity (PU), which allows us to identify all the parts that make up an entire object, supporting the individualization of individuals of a universal. In this context, the concept Rock, for example, provides a PI, but only a sample of a rock provides a PI and PU, since it is not possible to count a rock or define its beginning and end. Mixin Universals, in its turn, are defined by UFO as dispersive universals that generalize various universals with different identity principles.

Using as a criterion the rigidity property, UFO establishes two types of Universals: *Rigid Sortal Universals* and *Anti-Rigid Sortal universals*. A rigid sortal universal is the one where its individuals are the same in all possible worlds (in the modal sense), i.e., its individuals do not cease to be them without ceasing to exist. For example, a Person is a rigid sortal because only ceases to be a Person when ceasing to exist. On the other hand, anti-rigid sortal universals are those that are not needed for any of the individuals. For example, a Student is an anti-rigid sortal because a person can no longer be a Student and still exist.

As presented in Figure 3.4, UFO provides three fundamental types of rigid sortal universals: *Kind*, *Collective* and *Quantity*. Kinds represent universal whose individuals are functional complexes, such as Person, Car, etc. Collectives represent universal whose instances are collections of functional complexes that have a uniform structure, such as Forest, Deck, etc. Quantities refer to maximally self-connected portions of matters, such as Rock, Wood, Water, etc. *Subkind* are rigid sortals that carry the identity criterion for its instances, offered by Substantial Universals and inherited in the hierarchy. For example, Man and Woman are subtypes of the universal Person.

Among the anti-rigid sortal universals, UFO provides two types of universals: *Roles* and *Phases*. Roles are relationally dependent universals, i.e., an individual performs a Role when it is related to a foreign entity or when participating in events. For example, a person performs a Role of student when related to an educational institution. Phases are relationally independent universals that define different stages of a universal. Phases define disjoint partitions of a set, so that one individual may be in only one of these phases in a given world. The individual of a certain universal can go through various stages throughout their existence because of the occurrence of intrinsic changes without losing its identity. For example, Child, Adolescent and Adult are different phases of Person.

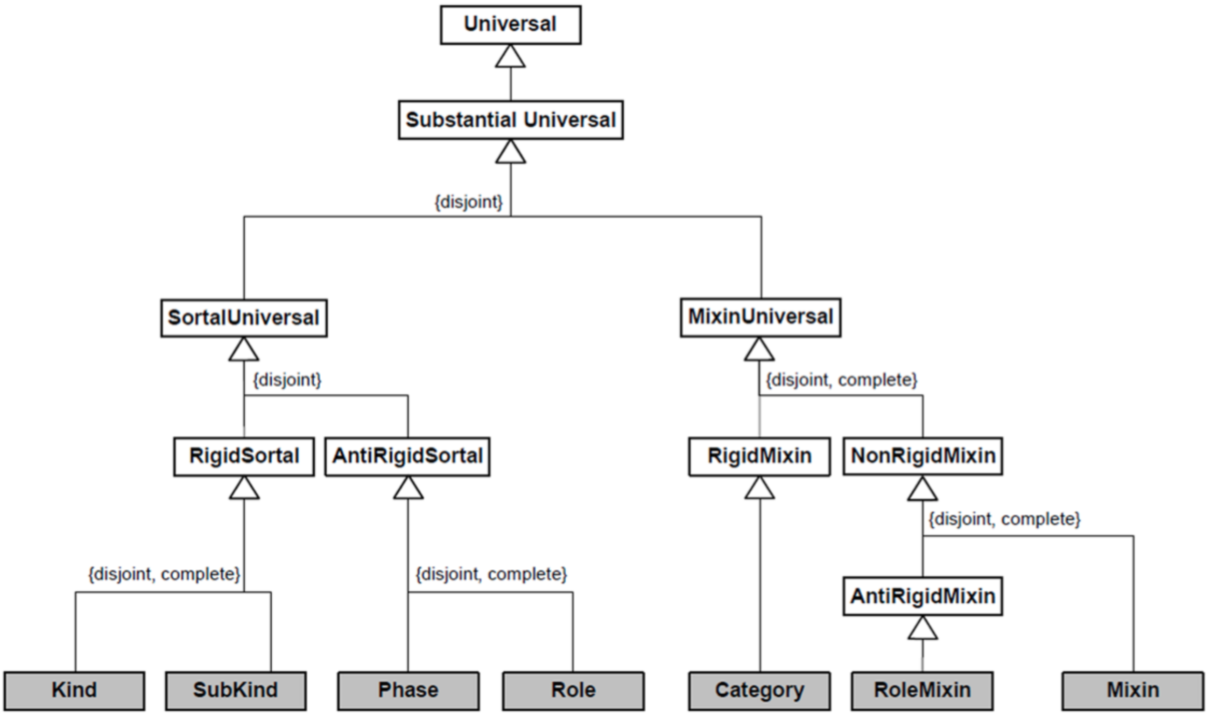


Figure 3.4: Representation of the taxonomy of UFO's substantial universals [Guizzardi 2005].

The Mixin Universals, also presented in Figure 3.4, may be understood as generalizations of distinct sortal universals, so that they are distinguished according to their rigidity. UFO distinguishes them in three types: *Category*, *Role Mixin* and *Mixin*. The Categories are universals that abstract an essential feature of many disjoint universals' individuals, being itself a rigid Universal. For example, a Relational Entity, which abstracts an essential feature shared by Person and Artificial Agent. Role Mixins are universals that abstract features that are accidental to individuals of various disjoint types, with different identity criteria. Thus, they can be understood as generalizations of multiple distinct roles. For example, Customer is a role mixin, which generalizes Personal Client (played by Person) and Corporate Client (played by Organization). Mixin are universal that abstract features that are essential to some of its instances, but accidental for others. An example of a mixin is Seatable Object that abstracts properties that are essential for Chair and Bench, but accidental to Table.

Finally, UFO proposes four types of *parthood relations* with different semantics, based on the types of entities that they relate: *componentOf*, *memberOf*, *subCollectionOf*, and *subQuantityOf*. Each parthood relation can be established only between individuals of specific UFO metatypes, which respect certain ontological constraints embodied in the UFO. [Guizzardi 2005] defines these relationships as follows:

- *ComponentOf* is the relation between a component and a functional complex, where the component preserves its individuality. For example, an engine is a *ComponentOf* a car, and a mineral is a *ComponentOf* a rock.
- *SubQuantityOf* describes the relation of constitution, in which a component is some amount of matter that cannot be individualized. For example, potassium is a *SubQuantityOf* feldspar, alcohol is a *SubQuantityOf* wine.
- *SubCollectionOf* describes the relation between two collections, a division of a collection, and the collection itself. For example, the northern part of the forest is a *SubCollectionOf* a forest, and the set of deposits produced by the transgressive cycle is a *SubCollectionOf* a stratigraphic sequence.
- *MemberOf* describes the relation of some individual to a collection of individuals. As an example, a tree is a *MemberOf* a forest, and a depositional unit is a *MemberOf* a stratigraphic sequence.

Regarding UFO-B is an ontology of particulars whose fundamental concept is *Event Universal*, which constitutes possible transformations of one situation of reality to another, i.e., events can change the state of affairs of a state (pre-state) to other (post-state). A *Situation*, in this sense, is an Endurant Universal that represents the state of affairs. Events can be of two types: *Atomic Event* or *Complex Event*. Atomic events are those that can not be decomposed into other events, such as an explosion. Complex events are composed of other events, such as a football match and a war. Events are ontologically dependent entities, since they depend on participants to occur. For example, considering the event *e*: the attack of Caesar by Brutus [Guizzardi 2005]. In this event there are the participation of Caesar, Brutus and the knife used in the attack. So, *e* is composed of the individual participation of each of these entities. Each *participation* is existentially dependent on a single substantial and may be itself an Atomic Event or a Complex Event.

4 PETROLEUM SYSTEM AND RESERVOIR MODELING

The oil and gas industry can be divided into three major sectors: upstream, midstream and downstream. The first one, which is also commonly known as the exploration and production (E&P) sector, finds and produces crude oil and natural gas. The midstream industry processes, stores, markets and transports commodities such as crude or refined petroleum products. Pipelines and other transport systems can be used to move crude oil from production sites to refineries and deliver the various refined products to downstream distributors. Finally, the downstream industry refers to the refining of petroleum crude oil and the processing and purifying of raw natural gas, as well as the marketing and distribution of products derived from crude oil and natural gas. The downstream sector provides to final consumers thousands of products such as gasoline, kerosene, jet fuel, diesel oil, heating oil, fuel oils, lubricants, waxes, asphalt, natural gas, and liquefied petroleum gas (LPG) as well as hundreds of petrochemicals [Feijo 2010].

The E&P sector involves activities in which acquisition, distribution and use of expert knowledge are more critical for the decision-making. In order to overcome this problem, the petroleum industry depends on derived interpretations, developed scenarios and taken decisions from a massive volume of data generated on computer models related to several process. Among these process, we can mention 3D seismic interpretation, well bore drilling, reservoir modeling and monitoring and also plant/facility modeling or monitoring capabilities.

In this chapter, we will present more details about the earth modeling activity for petroleum exploration. The operations related with this activity generated the core data analyzed in this work. However, before going into details about the earth modeling process and its activities, we will present an overview of the basic concepts related with the petroleum formation and the data derived from the E&P activities. The comprehension of these concepts is important for a better understanding of the data used in our approach. Thus, Section 4.1 presents the basic concepts related with generation, migration and accumulation of petroleum. Then, Section 4.2 presents more details about the data derived from E&P activities available for reservoir modeling. Finally, Section 4.3 presents an overview of the earth modeling activity for petroleum exploration.

4.1 Basic concepts

As previously mentioned, the main concern of all oil companies is to find commercial hydrocarbons deposits that are present in the subsurface as solids (solid products contained in oil shales), liquids (oil), or gas (natural gas). Both crude oil and natural gas are mixtures of molecules formed by carbon and hydrogen atoms. Among the different types of crude oil and natural gases, some are more valuable than others. Heavy crude oils, for instance, are very think and viscous, making it very difficult or impossible to produce, whereas light crude oils are very fluid and relatively easy to produce.

All hydrocarbon products owe their origin to organic matter preserved in sedimentary rocks [Biju-Duval 2002]. A *sedimentary rock* is a rock composed of sediments, and a *source rock* is a rock that forms gas or oil. These sediments are relatively simple materials such as sands deposited along beaches, mud on the sea bottom, and beds of seashells. As these sediments are deposited, together with organic matter (dead plants and animals), both are mixed.

The portion of the Earth's surface in with sediments transported by water have accumulated over a significant portion of the geological time is called a sedimentary basin. The most common site of sedimentary basins is in shallow marine environments around emerged continental areas. Also, but less frequently, a sedimentary basin can be found inside continental areas in lacustrine environments. As a result, the most part of the deposited sediments came from the seas, which have also contributed in the formation of the many sedimentary layers that composes a sedimentary basin. Mainly by the fact that during the vast expanse of geological time, sea levels has not been constant. Many times in the past, the seas have risen to cover the land and then fallen to expose the land [Hyne 2012].

Even in the presence of a mixture of inorganic and organic sediments in a source rock, the process of generation of oil and gas depends of many factors. In the subsurface, the most important factor in turning organic matter into oil is the temperature. The temperature for the formation of oil varies between 150°F (65°C), that occurs at a depth of about 7000 ft (2130 m) below the surface, and 300°F (150°C) at about 18,000ft (5500 m) (Figure 4.1). This zone where the oil is generated is called the *oil window* (HYNE, 2012). Beside these conditions, the reactions that change organic matter into oil are complex and take a long time.

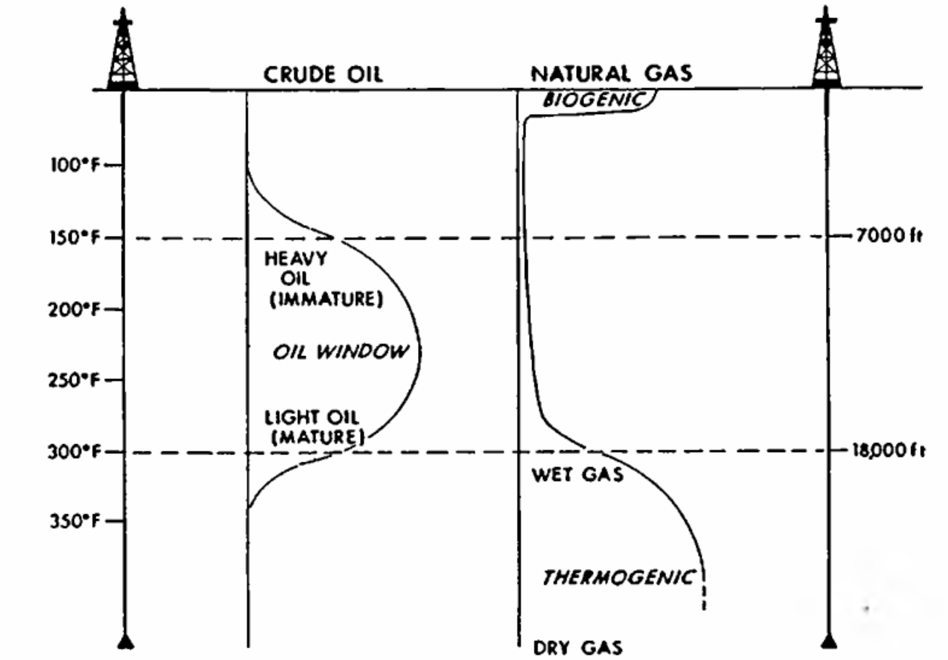


Figure 4.1: Generation of gas and oil [Hyne 2012].

After oil and gas have been generated, they are firstly trapped in the pores of host sedimentary sources rocks. However, due the stress caused by a large increase in volume after the generation of a liquid (crude oil) or gas (natural gas) from a solid (organic matter), the source rock is fractured. As a result, due to their low density and the pressure gradient to which they are subjected, the oil and gas rise outside their host rocks through fractures in the subsurface rocks, as shown in Figure 4.2. The vertical and lateral flow of the gas and oil from source rock is called *migration*. The rising oil and gas may continue until they reach the ground surface or are stopped by an impermeable obstacle (the seal), which will accumulate these liquid or gaseous products in the trap. According to [Hyne 2012], among all the gas and oil generated in a sedimentary rock basin, on the average, only 10% is trapped. The rest of the gas and oil either did not get out of the source rock, was lost during migration, or seeped into the earth's surface.

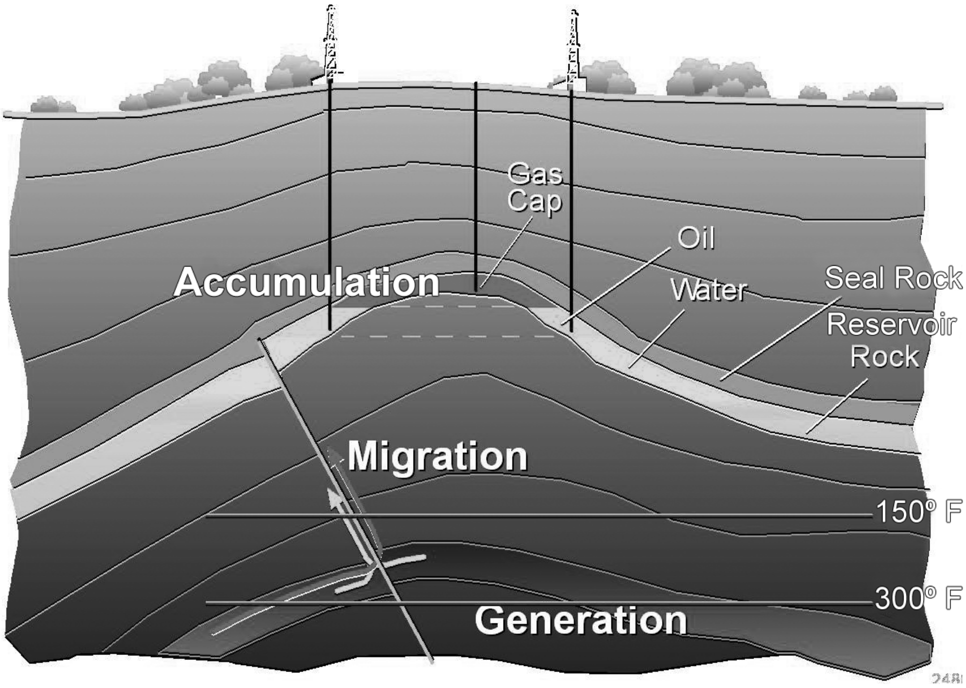


Figure 4.2: Hydrocarbon migration (adapted from:[Perrin and Rainaud 2013]).

The rock that can both store and transmit fluids is called *reservoir rock*. There are two important properties that a reservoir rock must have: porosity and permeability. *Porosity* is the percent volume of the rock that is not occupied by solids. *Permeability* is a measure of the ease with which a fluid can flow through a rock. Therefore, the greater the permeability of a rock, the easier it is for the fluids to flow through the rock. In the oil or gas reservoir, the oil or gas always shares the pore spaces with water, but the relative amount of the fluid sharing the pores of the reservoir will vary from reservoir to reservoir and is called *saturation*.

Once the gas and oil migrates into the trap, they will find the water that also occurs in the pores of the subsurface rocks, causing a natural separation process of the fluids according to their density. The lightest is the gas and goes to the top of the trap to form the free gas cap. The oil goes to the middle to form the oil reserve. The salt water, the heaviest, goes to the bottom. This process may results in a hydrocarbon deposit, which is a portion of a sedimentary basin in which oil or gas has accumulated and are present in quantities that allow them to be profitably extracted.

Therefore, the generation of oil and gas deposit depends of the migration of liquid or gaseous matter upward through sediment and encounter a structural trap that has a seal consisting of a concave volume of impermeable matter. Two common sedimentary rocks that can be seals are shale and salt. Figure 4.3 provides examples of structural traps. Identifying the format of reservoirs and the structure of petroleum traps are the main concern of the earth modeling activities.

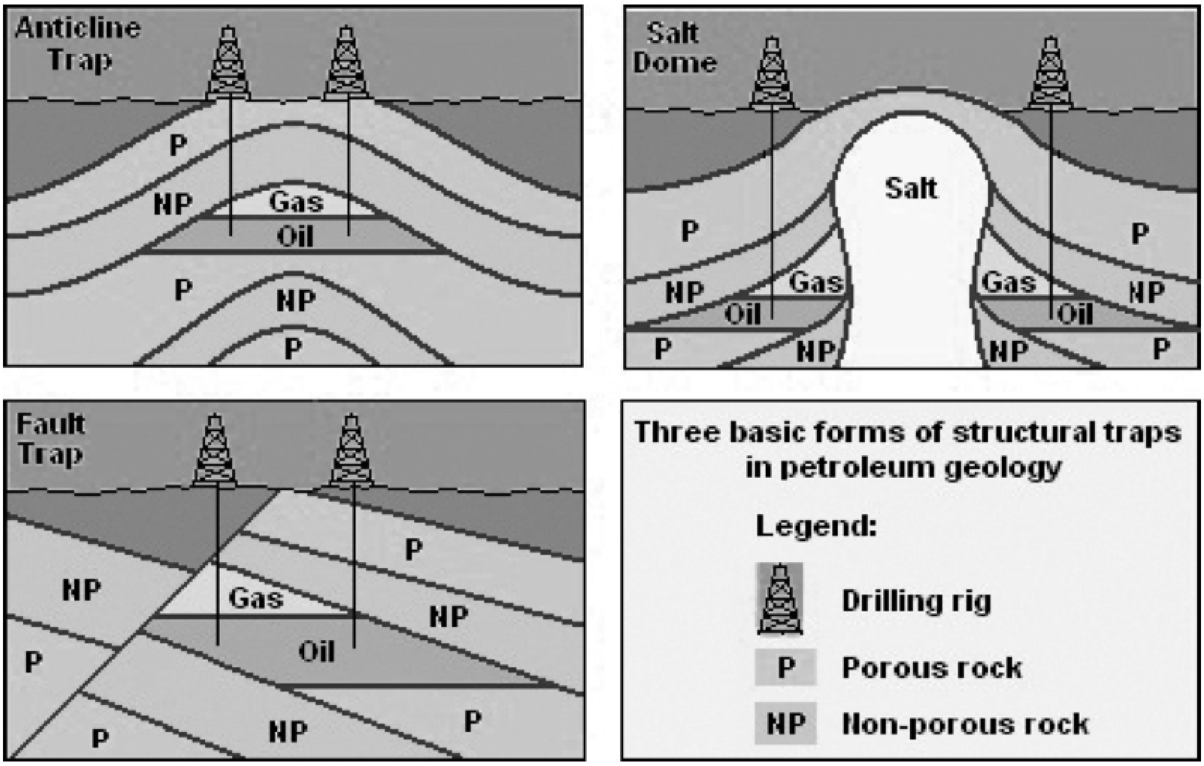


Figure 4.3: Examples of structural traps [Perrin and Rainaud 2013]

Thus, in order to have a commercial deposit of gas or oil, three geological conditions must have been met. First, there must be a source rock in the subsurface that generated the gas or oil at some time in the geological past. Second, there must be a subsurface reservoir rock to hold the gas or oil. Third, there must be a trap on the reservoir rock to accumulate the gas or oil into commercial quantities. Then, to find a hydrocarbon deposits we must first identify the basins with have sediments with significant quantities of organic matter and where favorable conditions existed for the transformation of this organic matter into hydrocarbons. Then, we must identify, although on a smaller scale, favorable structures where hydrocarbons were trapped in significant quantities and now constitute potential economically viable reservoirs.

4.2 Data derived from E&P activities available for reservoir modeling

Petroleum industry depends on commercial deposit of gas or oil to continue active in the market. To accomplish this difficult task of finding good reservoirs, many professionals are involved, each of them responsible for a specific task. A *geologist* is a scientist who studies the earth by examining rocks and interpreting their history. A *petroleum geologist* specializes in the exploration and development of petroleum reservoirs. An *exploration geologist* searches for new gas and oil fields. A *development geologist* directs the drilling of wells to exploit a field. A *petroleum geochemist* uses chemistry to explore and develop petroleum reservoirs.

These professionals are involved in the analysis of reservoirs, which can be investigated either by direct exploration of their rock content (*geological approach*) or indirectly by evaluating the spatial distribution of one or more rock-related physical properties (*geophysical approach*). The investigation process may be limited to the earth's surface, while using only geological mapping or remote sensing, or extended underground through the use of seismic analysis or drilling.

4.2.1 Seismic data

Originally, the seismic data was recorded by analog in the field on a sheet of paper. It was noisy and not very accurate. According to [Hyne 2012], the greatest improvement in petroleum exploration in the last several decades have involved new seismic acquisition techniques and computer processing of digital seismic data.

The generation and recording of seismic data involves different techniques based on the analysis of elastic waves generated in the earth by artificial means. It is then used to produce a 2D time cross-section, interpreted by scientists to define the composition, fluid content, extent and geometry of rocks in the subsurface (*Schlumberger Oilfield Glossary, 2009*[1]).

The interpretation of time cross-sections is very complex due to the fact that they represent wave travel times and not actual depths. Moreover, the time/depth conversion demands a computationally intensive process, which requires careful planning because of the succession of modeling operations (modeling chain) involved.

At present, new seismic acquisition techniques and computer processing of digital seismic data allowed the generation of models in three (3-D view) and four (4-D view) dimensions of the subsurface. Thus, seismic data are the only source of information in a full 3D volume. However, given that seismic information is made up of physical signals related to geological objects indirectly, seismic data provide only an indirect information about the geology. According to [Perrin and Rainaud 2013], seismic horizons cannot be thoughtlessly assimilated to geological horizons, and there is no seismic signal that directly corresponds to a fault. Also, the extraction

---

[1] http://www.glossary.oilfield.slb.com/

of surfaces to be modeled is very difficult when realized from raw seismic data. As a result, these raw data must be interpreted by geologists and geophysicists, since it involves both geological skill (identify structures of interest and surfaces relevant for modeling in seismic cubes or cross-sections) and image processing skill (follow the relevant seismic traces on a seismic cross-section or across 3D seismic data).

4.2.2 Drilling data

Together with seismic data, drilling is the other major source of information used for reservoir studies and earth modeling. Drilling data provides well information, which can be used to characterize rocks (lithology, petrophysical properties) and calibrate other data. The types of information provided by drilling can be classified into three main categories: rock samples, well logs and miscellaneous information (Figure 4.4).

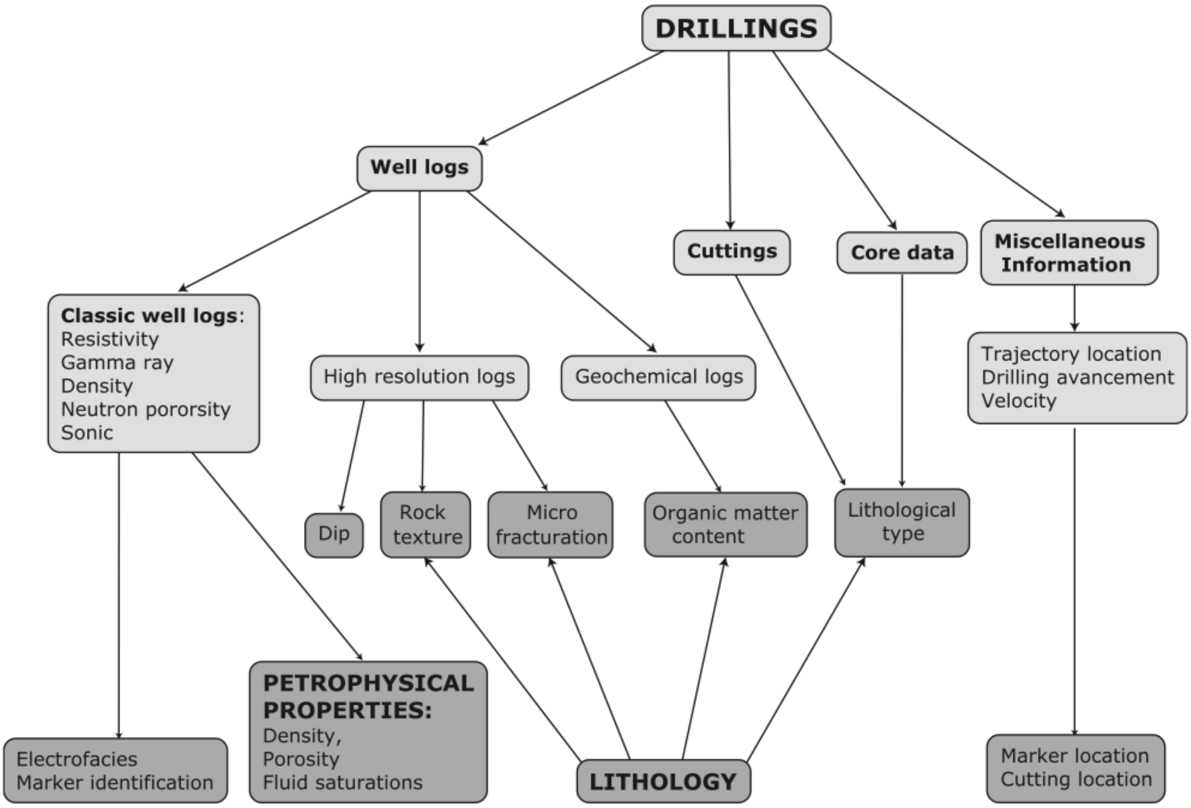


Figure 4.4: Information provided by well bores [Perrin and Rainaud 2013].

Among the rock samples that can be extracted from drilling operations, there are: core sample, cuttings and sidewall samples. The first one, are rock cylinders extracted during drilling. However, core sampling is a cumbersome, expensive, and time-consuming operation. As a result, core samples are obtained for just a few drilling and on limited depth intervals only. The second one, cuttings, are small rock pieces cut by the drilling tool and brought up to the surface with the mud flow that is used for cooling and lubricating the drill. Cuttings can be used to provide information about the various rocks encountered by the drill during its progress. However, for a cutting to be considered a representative wallside sample, it is necessary determine its original position along the well trajectory. The last one, sidewall samples, are small plugs cut from the sides of the boreholes.

The second main category, well logs, are records of various parameters measured along the well trajectory during or after drilling operations. There are three types of log: conventional well logs, that provide information about lithology and reservoir quality; high resolution logs, that provide complementary information concerning lithology (texture, microfracturation) and the orientation of stratifications (dip); and geochemical logs, which help quantify the proportion of organic matter present in the rocks.

Finally, the third main category, miscellaneous information, consists of measurements of parameters such as trajectory location, drilling advancement rate, and mud flow velocity, which are used to substantiate the data recorded along the well trajectory.

These drilling data (logs, cuttings and core data) is usually used on correlation activities between electric signatures and actual lithologies.

4.2.3 Regional geology data

Another information source used for reservoir studies and earth modeling is the data related to the geology of the prospected area. These data are acquired by geologists in their studies of the field and are essential for geological interpretation. This information consists in texts or maps, most often recorded in the form of paper documents, such as research papers, doctoral theses, public or corporate reports.

4.2.4 Laboratory data

The regional geology data are usually combined with laboratory data, also very important for reservoir studies and earth modeling. The laboratory data includes: the petrography of rock thin sections, which provide detailed petrologic information and allow rock petrofacies to be defined; the results of tests on rock samples to determine properties such as porosity, permeability, transmissivity, mechanical strength, and so on; petrophysical properties can be static (porosity) or dynamic, that is, they depend on some external condition (transmissivity, for example, depends on the differential pressure applied to both sides of a given rock sample); and geochemical data resulting from chemical analysis of rocks or individual minerals. Depending on their nature, the data are likely to be available as numerical data, image or graphic data, or textual information.

4.3 Earth modeling

In the modern petroleum industry, earth models are key tools for identifying and characterizing potential hydrocarbon reservoirs. Earth models are three- (3D) or four-dimension (4D) representations of data and interpretation concerning subsurface resources developed by geoscientists who are responsible for evolving a hydrocarbon prospect through various stages of modeling. Their final goal is the building of a reservoir model, which will be used for simulating oil accumulation in the underground. Currently, the building process involves a long chain of activities, which starts with data acquisition and proceeds with several different steps of data analysis and interpretation. This chain of activities is known as the *Earth Modeling Workflow* or *Reservoir Modeling Workflow* [Mastella 2010]. Figure 4.5 illustrates the most important steps of this workflow.

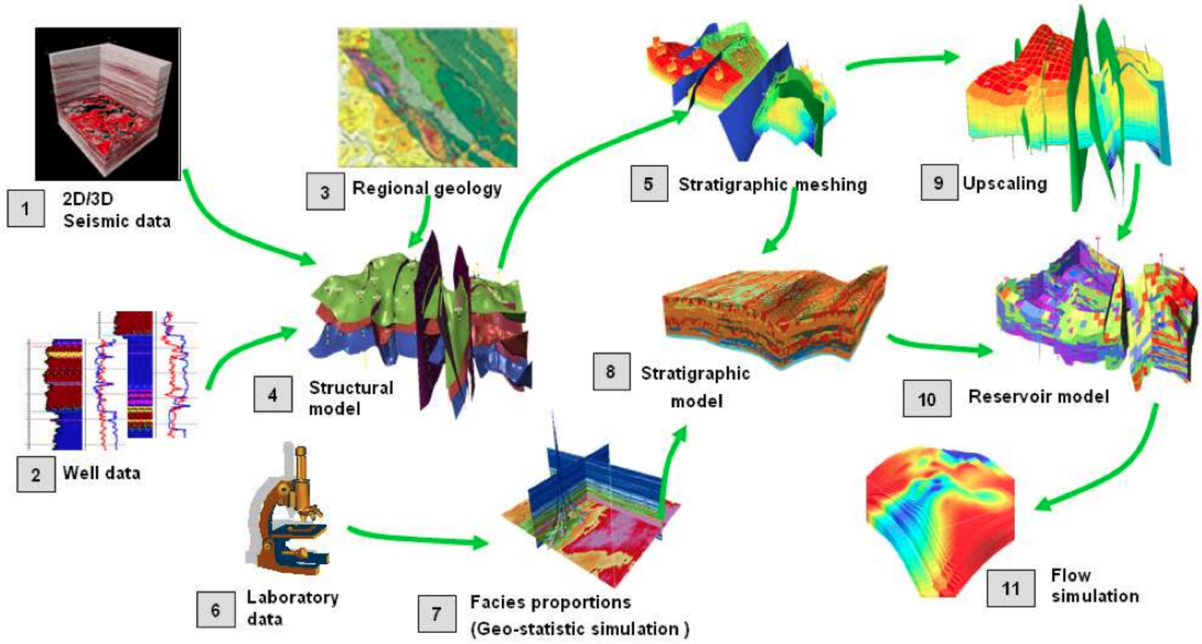


Figure 4.5: The Earth Modeling Workflow (adapted from [Perrin et al. 2007]).

The first modeling activity is the definition of the spatial 3D area of interest, called *prospect*. The available data concerning the prospect are derived from E&P activities (explained in Section 4.2) such as seismic reflection (Figure 4.5(1)) and well boring drilling (Figure 4.5(2)). Also, geoscientists take into account former studies about the prospect, such as documents concerning regional geology, geological maps or cross-sections (Figure 4.5(3)). All these data must go through a rigid quality control, in order to be validated, taking into account uncertainty values, and to verify their consistency. Moreover, the data are used by different professionals in various interpretation tasks, which fall into three broad types: *selection*, *association* and *data modification*. An overview of the different interpretation task involved in the reservoir modeling workflow is given in [Perrin and Rainaud 2013].

The next task is the structural interpretation, which is a crucial step in the workflow because it is here that the structural model (Figure 4.5(4)) is built. The geoscientists use all available data issued from regional geology studies and computer modeling tools (called geomodelers) to carry on this task. The surfaces identified by seismic interpretation are loaded in these tools, allowing the geoscientists specifying spatial and chronological relationships between the identified objects. The topology of the object assemblage is of paramount importance since it strictly depends from geological interpretation [Perrin 1998]. The structural model is the “skeleton” on which other earth models will be built, consisting in an assemblage of geological surfaces that mark the boundaries of individual geological blocks.

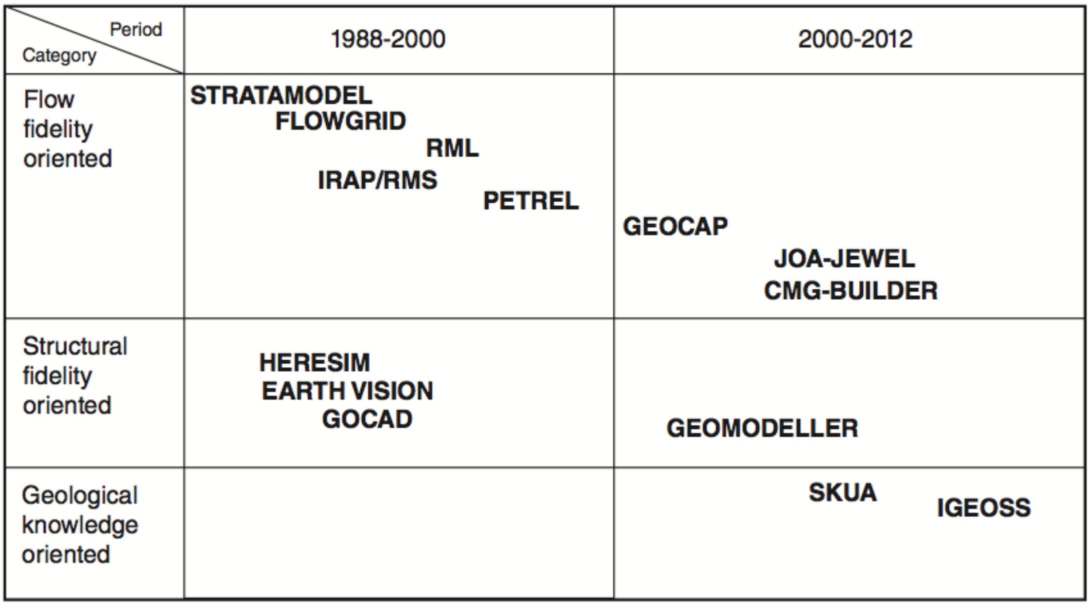
Then, inside each of these blocks is build a stratigraphic mesh (Figure 4.5(5)) in an activity called stratigraphic modeling. Each cell of each mesh is affected by the petrophysical properties acquired from isolated points corresponding to samplings and to laboratory studies (Figure 4.5(6)). Using geostatistic simulation, these properties are then propagated to the whole volume (Figure 4.5(7)). The resulting model, where the stratigraphic mesh cells are filled with rock properties, is called stratigraphic model (Figure 4.5(8)).

This stratigraphic model mesh has its geometry transformed, in order to obtain a coarser reservoir mesh, and its property values upscaled (Figure 4.5(9)). The result of these tasks is a reservoir model (Figure 4.5(10)), which provides a complete set of continuous reservoir parameters (i.e. porosity, permeability, water saturation) for each cell of the 3D grid. Finally, this model can be used by reservoir engineers to compute realistic hydrocarbon fluid migration simulation (Figure 4.5(11)). In this simulation, the amount of exploitable hydrocarbon reserves in the prospect can be estimate, as well the quality of these reserves (heavy or light oil, gas etc.).

5 CURRENT SOLUTIONS FOR GEOLOGICAL KNOWLEDGE FORMALIZATION

The creation of a common work platform has always been a major concern for software providers to the oil industry. However, from the user point of view, it is still very difficult to transfer data from one application platform to another. The main reason is the complex history of reservoir modeling market, characterized by extensions or redevelopments of products, software acquisitions and company merging. In this market, the major software providers have intended to cover as much as they could all the steps of the reservoir characterization workflow. However, considering the complexity of the reservoir modeling process and the quick technological evolutions, it was a very ambitious scope. In order to supply the deficiency of some "niches" of excellence, some small software providers began to offer more specialized products, that started to replace some products designed by the major software providers and that needed to be plugged into major earth modeling workflows.

Perrin et al., in [Perrin and Rainaud 2013], presents the historical evolution of software products for earth modeling (represented in Figure 5.1), that began to emerge in the middle of 80's with the goal of representing geological surfaces and the spatial repartition of rock petrophysical properties. From these early times on, many software began to appear on the market according to two broad orientations: towards fluid flow simulation; and towards high-resolution representations of geology. During this period, geophysicists, geologists, and reservoir engineers were working in strictly separated departments with few communication among them. This situation has prompted a third orientation of software development, which intends to integrate the various kinds of knowledge required for earth modeling in order to produce "*Knowledge based models*".



Period / Category	1988-2000	2000-2012
Flow fidelity oriented	STRATAMODEL FLOWGRID RML IRAP/RMS PETREL	GEOCAP JOA-JEWEL CMG-BUILDER
Structural fidelity oriented	HERESIM EARTH VISION GOCAD	GEOMODELLER
Geological knowledge oriented		SKUA IGEOSS

Figure 5.1: Historical evolution of software products for earth modeling presented in software packages categories by period of first appearance in the market [Perrin and Rainaud 2013].

However, the construction of this type of model requires that the oil companies agree on a common way of capturing knowledge about geological objects. According to the knowledge engineering community, the type of knowledge about the categories of objects that exist in a domain, and about the manner in which these objects are organized, is called *static knowledge* [Mastella 2010]. Ontologies are shown to be the best approach to make explicit the static knowledge that expresses the common understanding about earth sciences domains.

In this chapter, we will present some of the ontologies formerly developed for Geosciences, in Section 5.1, and for 3D earth modeling, in Section 5.2. Among them, we can highlight the *Basic Geology* ontology, proposed initially in [Mastella 2010] and keeps evolving. Finally, in the Section 5.3, we will describe the most used communication standard formats: LAS, WITSML and RESQML. This review about the current ontologies and standard formats will be useful to understand how the main geological concepts, described in the ontologies, are stored in the standard formats' files, enabling the creation of mappings between them, as presented later in this work.

5.1 Ontologies developed for Geosciences

Intense efforts were developed, during the last years, by various organizations for issuing codifications and formalizations of geological knowledge. According to [Mastella 2010], these efforts can be classified in various categories according to the specific domains or activities that they address: geological surveys; specific geoscience domains; and petroleum industry.

Geological surveys are national or regional institutions, which are notably in charge of issuing geological maps. Their main goal is exchanging the information contained in field or laboratory observations and linking it with the objects that they intend to represent on a geological map. Between the geological surveys, a few models stand out. The NADM model (*North American Geologic Map Data Model* [Richard 2006])[1] and the derived GeoSciML[2] model are designed as ontologies for developing interoperable geologic map-centered databases. The GeoSciML formalization is based in addition on the normative GML (*Geography Markup Language*) for the representation of geographic features and geometry. The GEON (*Geosciences Network*)[3] project is interested for its part in the problem of integrating geologic maps, whose source files contain geologic age or rock type information in the tables with different schemata and vocabularies.

For specific geoscience domains, many knowledge models were defined as well as specialized domain ontologies and upper level ontologies. Some of these models are to be found in [Sinha 2006]. In addition to those, stand out the ontologies proposed in: [Abel 2001], for petrographic description of reservoir rocks; [Cox and Richard 2005] and [Perrin et al. 2011]

---

[1] http://ngmdb.usgs.gov/www-nadm/

[2] http://www.geosciml.org/

[3] http://www.geongrid.org/

for geological time; [Lorenzatti et al. 2009], for modeling of sedimentary structures and textual features of rocks; and [Carbonera 2012], for sedimentary stratigraphy. Some of them are proposed to be upper-level ontologies, which define high-abstract objects in some domains, intended to support the organization and knowledge interchanging is some area of knowledge. That is the case of the ontologies proposed by the project SWEET (*Semantic Web for Earth and Environmental Terminology*), which includes several thousand terms, spanning a broad extent of concepts from Earth system sciences and related concepts [Raskin and Pan 2005].

Among the oil companies projects, stand out the IPP (*Integrated Information Platform*, [Omdal 2006, Sandsmark and Mehta 2004] project, comprising one of the largest ontologies ever developed for an industrial field for formalizing the terminology used in petroleum production. The project addresses many domains, such as subsea production equipment, seismic, drilling and logging, reservoir evaluation, but does not include earth sciences. Parts of the ontology are based on the ISO 15926 standard, for oil and gas production life-cycle data, which considerably differs from the oil and gas exploration life-cycle data, but they also include concepts issued from other terminologies.

5.2 Ontologies for earth modeling

Ontologies for 3D earth modeling differ from those ontologies developed for geosciences or for petroleum production industry due the fact that they were built specifically for representing knowledge about 3D geological modeling, i.e., for describing the objects that are manipulated within earth modeling workflows. These ontologies began to be developed from 2001 on, when the IFP[4]/ENSMP[5] team for Geo-modeling developed a new knowledge-driven paradigm for reservoir studies based on the belief that geo-model building should not be directly dependent from data (data-driven) but rather from geoscientists' interpretations (knowledge-driven) [Rainaud et al. 2005]. In 2005, the IFP/ENSMP team issued the first version of a *Geo-ontology*. This ontology was described in [Perrin et al. 2005].

After the issued of this first *Geo-ontology*, stood out the *Basic Geology* ontology, defined in [Mastella 2010], that describes and interconnects geological entities considered in reservoir modeling. This ontology is divided into sub-ontologies, which provide more detail to the main top-level concepts, and into other domain ontologies, which represent fields that are independent of the *Basic Geology* ontology, but whose concepts are used by its concepts. The domain ontologies linked to the *Basic Geology* ontology are: *GeoLocation ontology*, an ontology of geographical terms; ontologies for the disciplines of *Palaeogeography*, *Lithology* and *Hydrogeology*; and ontologies for defining and managing geological ages: *Geological Time* and *Geological Dating* ontologies (both formalized in [Mastella 2010]). The Figure 5.2 presents the top-level part of the *Basic Geology* ontology.

---

[4] http://www.ifpenergiesnouvelles.com/

[5] http://www.mines-paristech.fr/

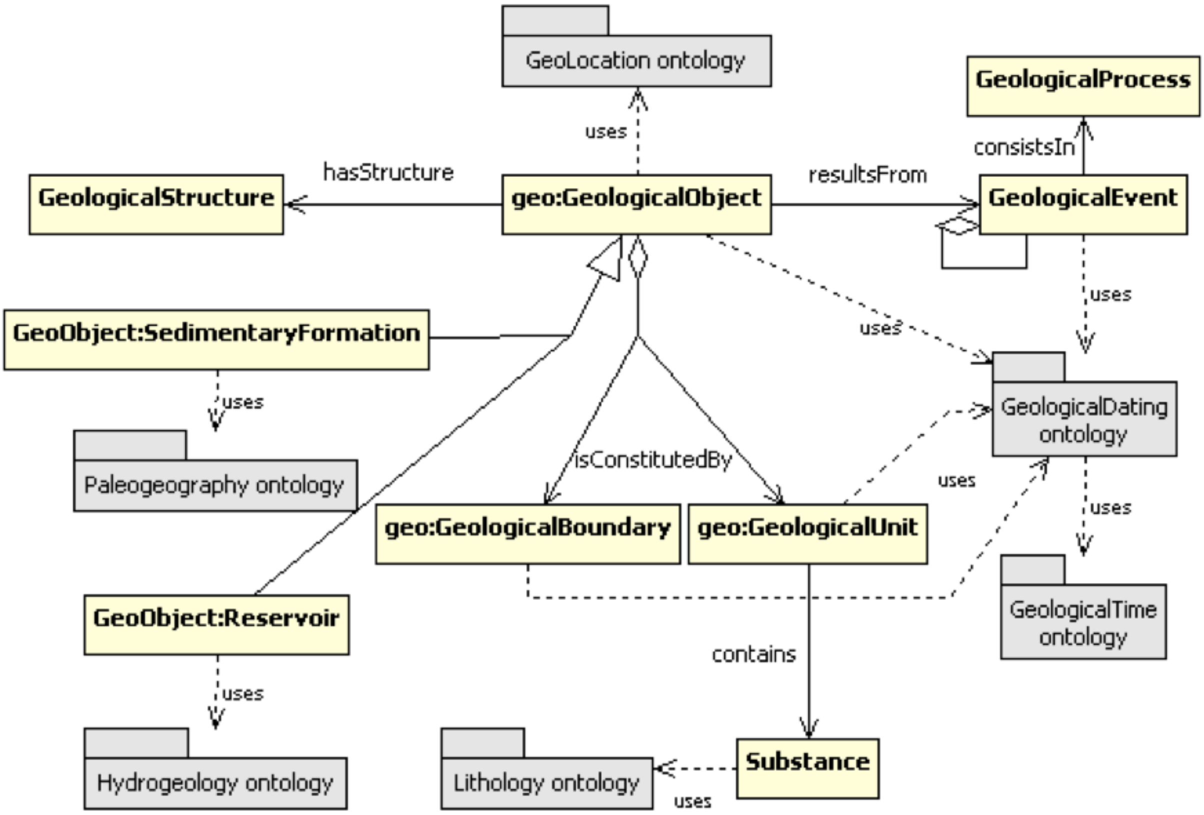


Figure 5.2: The top level *Basic Geology* ontology [Mastella 2010].

The original definition of the ontology can be found in the site of the E-WOK HUB (Environmental Web Ontology Knowledge Hub) [6] project. The RDFS/OWL version of the ontologies related to this project can be downloaded from this same site[7].

In the Section 5.2.1 we will describe some details about the *Basic Geology* ontology in order to clarify the objects related to the concept *GeologicalObject*, which will be used further as a case study to exemplify the methodology proposed in our work.

5.2.1 Description of the Basic Geology ontology

The *Geological Object* ontology summarizes a diversified amount of endurant geological objects that can be simple or complex (examples among many other are: a stratified sedimentary unit, a reef, a diapir, a fault network, etc.). Complex geological objects can be made of a various number of atomic geological objects. The *Basic Geology* ontology was proposed with the important role of providing high level, general geological objects that helps in linking further knowledge and data representations.

There are two kinds of elementary geological objects [Mastella 2010]:

---

[6] http://www.inria.fr/sophia/edelweiss/projects/ewok/

[7] http://www-sop.inria.fr/edelweiss/projects/ewok/ontologyview/ontologies.html/

- 2D objects, corresponding to **Geological Boundaries**, such as the erosion surface *E*, the fault *F* and the upper and lower boundaries *bu* and *bl* on Figure 5.3;
- 3D objects, which are **Geological Units**, such as the sedimentary unit *U* limited by the boundaries *bu* and *bl* on Figure 5.3. A *Geological Unit* is a volume of continuous geological matter limited by one or several *Geological Boundaries*.

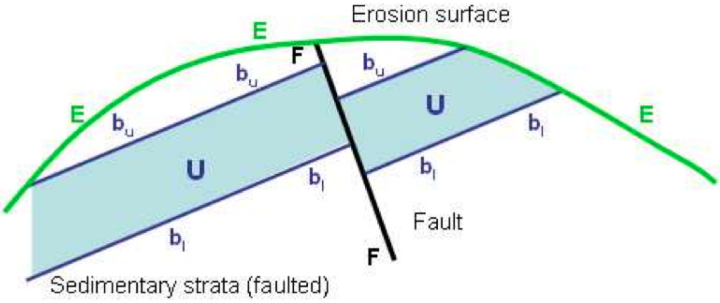


Figure 5.3: Geological objects: Erosion surface E, fault F and Sedimentary strata unit U, constituted of volume and boundaries bu and bl [Mastella 2010].

5.3 Communication standard formats

A huge amount of data formats and standards are being used for data exchange in reservoir characterization models. Usually, these data are stored in different file formats and represented in different formats.

The *Geological File* ontology defines the file formats used for storing information related to geological data. Basically, every data are stored in textual documents or classified as one-dimensional, two-dimensional, or three-dimensional documents, according to the way in which the data are stored (one, two or three dimensional arrays). Among those formats, the LAS and the WITSML standards store geophysical log data in one-dimensional documents. The RESQML standard stores its data in three-dimensional documents, divided by three model types: 'Reservoir', 'Structural' and 'Stratigraphic'. Figure 5.4 represents a part of the *Geological File* ontology with the concepts related with the LAS, WITSML and RESQML standard formats.

PRODML (PRODuction-ML) is an industry standard that supports data exchange representing the flow of fluids from the point they enter into the wellbore to the point of custody transfer, together with production operation workflows, in a vendor-neutral, open format. As well as WITSML and RESQML, it was proposed by the *Energistics Consortium*[8] with the intention of covering the whole exploration and production chain of petroleum data exchanging. The PRODML includes standardized objects for: DTS Measurements; Fluid Analyses; Fluid Samples; Flow Networks; Production Operations Reports; Production Reports; Historical Data;

---

[8] http://www.energistics.org/

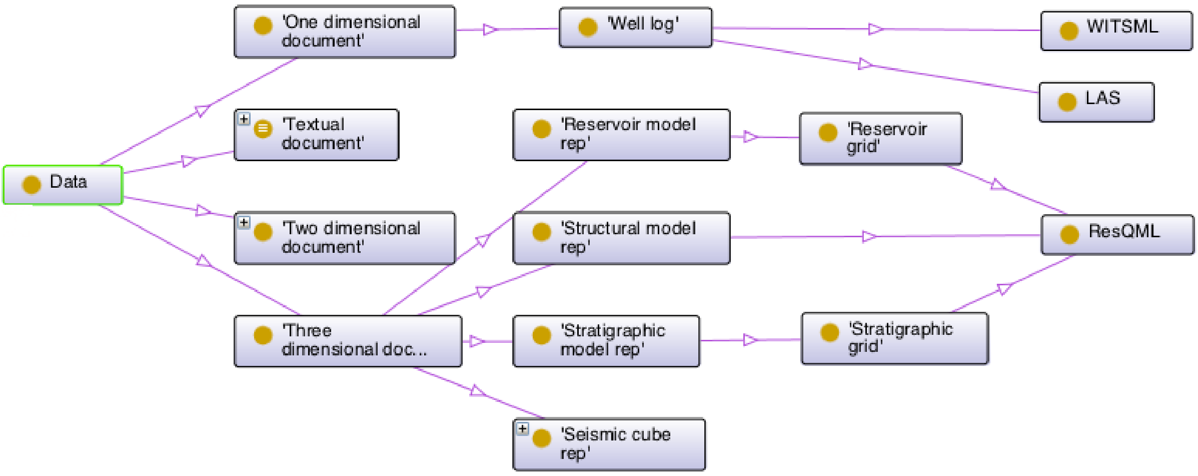


Figure 5.4: The concepts related with the LAS, WITSML and RESQML standard formats in Geological File Ontology.

Well Tests; and Wireline Formation Tests. However, none of these objects are directed related with geological objects. For this reason, this standard has no geological entities or correspondence with the *Geological File* ontology and, therefore, will not be analyzed in our study.

The LAS (*Log ASCII Standard*) format began with the aim of a simple format for exchanging well log data. The worldwide acceptance of LAS proved the need for such a format. The *Canadian Well Logging Society* [9] introduced the LAS standard, in 1989, to standardize the organization of digital log curve information for personal computer users. Version 1.2 was the first version to be used in industry and was followed, in September 1992, by version 2.0, in order to address some inconsistencies. A more versatile version, LAS 3.0, was released in 1999. However, LAS 2.0 remains the dominant product until now. LAS 3.0 clarifies several of the poorly defined specifications of LAS 2.0 and provides expanded data storage capabilities, but has seen limited implementation. In this work, we will focus on LAS version 3.0.

LAS 3.0 files are divided into logical sections. Each section begins with a title line, which is marked with a tilde (~) at the beginning of the line. Sections contain lines where data is described and/or stored. There are several types of sections and several types of lines within sections. The LAS 3.0 standard defines which combinations of sections must exist in LAS files, and in which order. For example, the *~Version* and *~Well* section must exist in that order in any file of the version 3.0 LAS. As in LAS version 2.0, only one well is described within a single file.

As we saw before in the *Geological File* ontology, the LAS standard stores “Well Log” data in one-dimensional ASCII documents, meaning that the data are stored in one-dimensional arrays. However, when representing information through arrays, LAS version 3.0 allows that these documents have data stored as one, two or three-dimensional arrays, something that was

---

[9] http://www.cwls.org/

not possible until this version. The data are usually indexed to depth or time, but may be presented as discrete measurements if required. Also, data are grouped by type into related sections, as they relate to the well where the data was acquired. Types include depth and time indexing logging, core, inclinometry, drilling, formation tops, test data, user defined types, etc. In the Chapter 7 we will present an ontological analysis of these types.

On the other hand, the WITSML (Wellsite Information Transfer Standard Markup Language)[10] is a standard used for sending well site information in an XML document format developed to promote the right-time, seamless flow of well data between operators and service companies, as well as regulatory agencies, in order to speed and enhance decision-making and reporting.

A WITSML document consists in one or more complete WITSML data-objects that correspond to a logical representation and organization of the data items associated with the major components and operations involved in well drilling - such as well, wellbore or log - represented as an XML document, which is essentially a text string. Thus, each WITSML data-object is defined by an XML schema and its own document. Each schema defines a set of data that can be transmitted within a single XML document and represents a cohesive subset (e.g. well, wellbore, rig, etc.) of an overall logical schema related to a single domain (well). Data object schemas contain attributes, elements, and included component sub-schemas. Figure 5.5 represents the WITSML data objects relationships.

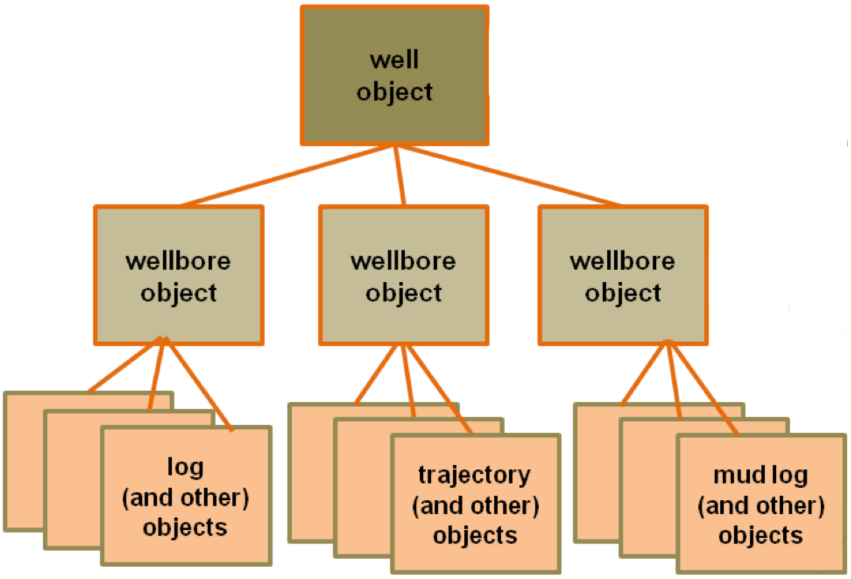


Figure 5.5: The WITSML data objects relationships [Energistics and SIG 2012].

Component schemata are XML schemata, but these schemata do not represent complete data objects and do not contain global elements. A component schema may be used by more than one data object schema. All component schemata are prefixed with (cs_). Each component schema file generally defines one type that has the same name as the file name.

---

[10] http://www.energistics.org/drilling-completions-interventions/witsml-standards

Finally, the RESQML is an industry initiative to provide open, non-proprietary data exchange standards for reservoir characterization, earth and reservoir models proposed by *Energistics's Special Interest Groups* (SIG)[11] that includes most of 3D-modeling software providers besides most of large petroleum companies and it getting fast acceptance. RESQML is an XML-based data exchange standard that helps addressing the data-incompatibility and data-integrity challenges faced by professionals in petroleum industry when using the multiple software technologies required along the entire subsurface workflow, for analysis, interpretation, modeling, and simulation.

The release available for the public is RESQML V2.0 [Endres et al. 2013, Deny et al. 2013], which evolves from V1.1 by incorporating more semantic to the representation model. The key goal of this version is to provide a mechanism for transferring relationship information (between data-objects, such as faults, horizons and grids), while continuing to expand the fundamental data types within the standard, for example, unstructured simulation grids.

There are significant differences between the two RESQML versions. Among them, we can highlight that in the RESQML V2.0, the concepts feature, interpretation and representation (introduced by RESQML V1.1 for individual structural components such as horizons and faults) are applied to more individual elements, like *FluidBoundaries*, *Geobodies*, *StratigraphicUnit*, and *FluidFlowUnit*. Also, the RESQML V2.0 defines some new concepts [Endres et al. 2013]: organization, organization interpretation, and organization representation. The objective is to gather the relationships between individual features. Thus, V2.0 organization structures allow a more flexible association of the model elements as well as partial model descriptions. The main intention on RESQML V2.0 is to capture the steps of interpretation between each step of the modeling process, preserving the author and the evidences of interpretation. This capability allows that some misunderstanding of features along of the modeling process can be tracked and corrected without affecting the whole result.

In RESQML V1.1, a RESQML document consisted of one XML file and one HDF5 file, associated together by standard naming conventions. V1.1 used what was *essentially* a hierarchical data model, such that all semantic information was assembled in only one XML instance consisting of one XML file, with hierarchical "implicit" XML containment to associate the data-objects together. The optional HDF5 file was used for better processing efficiency of large arrays of data. Both XML and HDF5 are still used in RESQML V2.0, but now multiple XML and HDF5 files are used.

RESQML has also moved from a hierarchical data model to an object-relationship data model to organize its data-objects. This change allowed the components of the earth model to be represented as separate data-objects and the relationships among them to be more accurately represented as parent-child, with one-to-many or many-to-may relationships.

Thus, V2.0 is a significant redesign of RESQML V1.1, including more and richer data-objects, a clearly defined knowledge hierarchy, and the ability to create and transfer complete

---

[11] www.energistics.org/reservoir/resqml-sig

or partial models [Energistics and SIG 2014].

A RESQML data-object, which is now stored in a separate XML file, contains the semantics associated with the data model (or modeling data) in the context of RESQML. V2.0 introduced a new design that supports the transfer of abstract subsurface features, human interpretations of those features, the data representations of those interpretations, and the properties indexed onto those representations, which results in a well-defined knowledge hierarchy of feature/interpretation/representation/properties (informally referred to as “FIRP”) (Figure 5.6). Each of these terms is a type of RESQML data-object. V2.0 also includes the relationships between them, which allows a more precise classification.

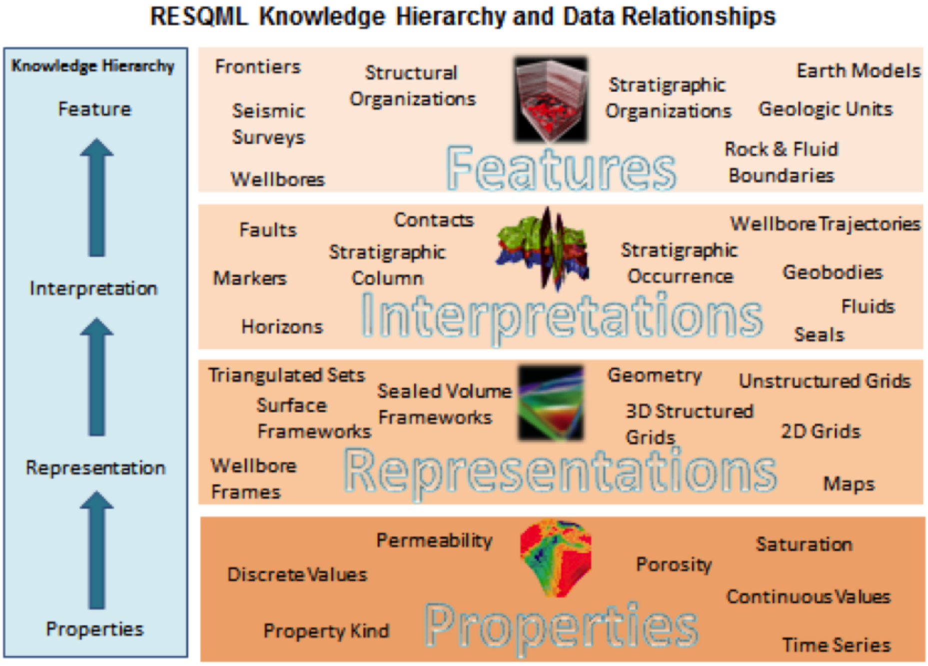


Figure 5.6: The feature/interpretation/representation/properties knowledge hierarchy [Energistics and SIG 2014].

Feature can be defined as something that has physical existence at some point during the exploration, development, production, or abandonment of a reservoir. *Features* are divided into two categories: *geological features*, which are objects that exist a priori, in the natural world (e.g., a boundary), and *technical features* that are objects that exist by the action of humans (e.g., a well).

Interpretation can be explained as a single consistent description of a feature. An interpretation is subjective and very strongly tied to the intellectual activity of the project team members.

Representation is a digital description of a feature or an interpretation and contains the *topology* and *geometry* of a structural feature. *Topology* defines how to associate nodes and other “indexable elements” to represent points, lines, surfaces or volumes (like structured and unstructured grids). *Geometry* is the spatial location of each selected indexable element, mainly nodes. This information may be provided as numerical arrays stored in HDF5 datasets, or specified implicitly.

A *property* can be attached to any indexable element of any representation. *Properties* refer to semantic variables (for example, porosity, permeability, etc.) and the corresponding data values, which are recorded in arrays, which may be stored in HDF5 datasets.

For each of these data-objects, independently of its level, each instances is uniquely identified with a UUID and metadata (a citation data-object). In this work, we will analyze only those instances related to the data-object *features* (e.g., boundary, rock, stratigraphic unit, etc.).

6 PRODUCING MORE INTEGRABLE MODELS

In order to produce more integrable models, we first need to deal with the heterogeneity in geological information. The construction of a final reservoir model requires the integration of data sets across different disciplines related to earth sciences, such as geophysics, geology, petrology and petrophysics. This multi-disciplinary problem can be seen as a problem of perspectival heterogeneity, as explained in Section 2.2 of Chapter 2, in which various models represent various points of view in relation to the same domain of interest.

The integration of these heterogeneous and multi-disciplinary data allows the emergence of new knowledge, which is essential for timely and correctly decision-making. However, integration in this context means finding correspondence between entities without looking for the identity of the instances themselves. In that way, we can reach integration by identifying entities (that exist in the real world) being represented differently among various models and data sources and by offering an integrated vision of those entities through a mapping between them and a common vocabulary of a particular community. We propose the use of ontologies to solve this issue.

At this point, the first challenge that must be faced is the problem of reaching interoperability. We claim that interoperability can be reached by making apparent the meaning of the geological concepts represented in the models. Figure 6.1 illustrates a common situation that can happen in a modeling scenario: a geological concept being represented differently among models. In the illustrated example, the geological concept named *Sedimentary Unit* is represented in three different ways. In order to reach interoperability among these three representations, we need to determine how each of them represents a *Sedimentary Unit* and the exactly places that the data and interpretations referring this concept are stored. Thus, this concept could be used to anchor these three models.

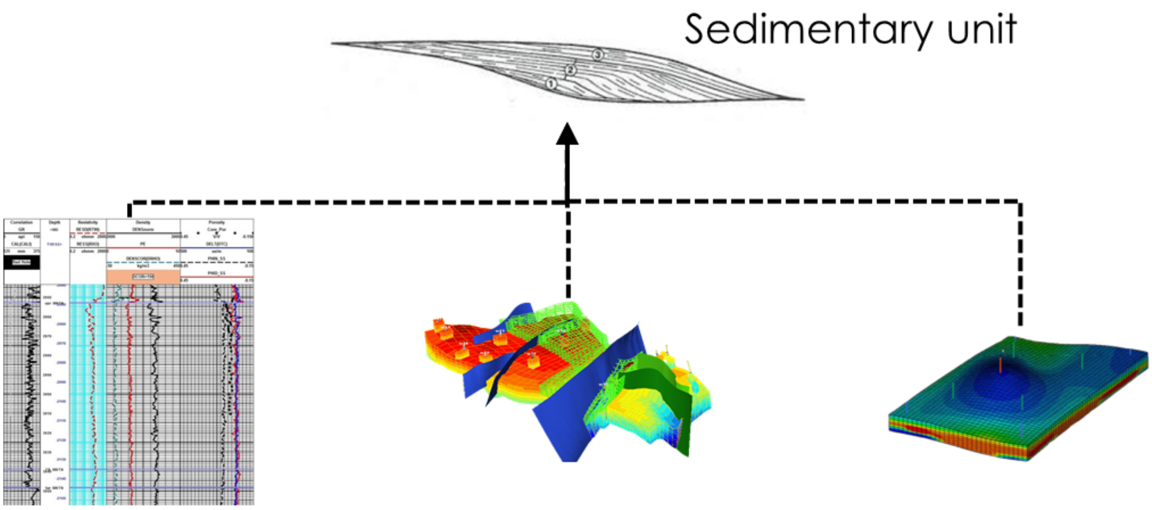


Figure 6.1: A *Sedimentary Unit* being represented differently in three models.

However, this solution results in a new question: can any geological concept be used to anchor the models? The answer for this question is no. The geological concepts are very susceptible to modeling errors, since each professional involved in a modeling process can have different conceptualizations about entities of the world. Furthermore, the available representational languages applied by the experts, usually cannot distinguish some entities, since the languages applied for the representations do not keep a fair syntactical and ontological capability for representing world entities. What is needed, therefore, is a conceptual tool that allows one to explicit differences between entities of the world that representational languages usually consider as being of same type, and to decide which objects exist in the world reality. Then, in order to increase the semantic of each geological object, we applied the concepts of foundational ontology. The use of UFO meta-types, described in Chapter 3, allowed us to explicit the differences between the many entities of the world directly involved with reservoir modeling, solving the lack of expressivity of representational languages.

Thus, using the previously example, Figure 6.2 presents the classification of a *Sedimentary Unit* according to UFO meta-types. In this example, we classified a *Sedimentary Unit* as a *Kind*. Moreover, we considered that the representations of the classified concept are instances described by the data and interpretations stored in interchange data formats, such as LAS, WITSML and RESQML.

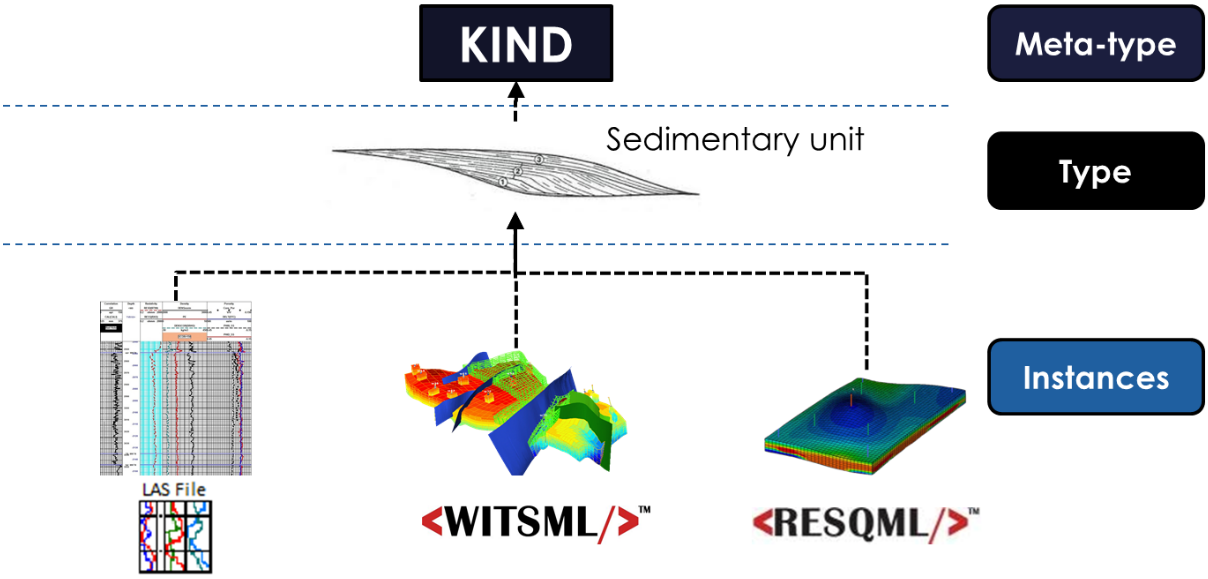


Figure 6.2: Using UFO meta-types to clarify the meaning of geological concepts.

Thus, the second challenge that must be faced is to determine which concepts can be used to anchor the models. We claim that only those concepts that preserve the identity through time (rigid concepts) and that are countable (whole concepts) can be used. The reason is that rigid concepts can be found in all models and can be used to support the interoperability among systems.

We will show that the use of ontological tools can help the modeler in creating more integrable models. More specifically, referring to the UFO methodology, exposed in Section 3.1 of Chapter 3, we will show that the use of ontological concept notions of identity, rigidity, essentiality and unity allows modelers to distinguish which concepts exist in world reality and which of them are distinguished from each other.

6.1 Proposed methodological approach

In this section, we will describe the methodology defined and applied to analyze the several standards being used by petroleum industry nowadays (the result of our ontological analysis will be presented in the Chapter 7). Our approach aims to explicit the semantic of geologic concepts presented in these standards, assisting in solving problems of semantic heterogeneity.

The first step (1) is to identify an ontology that covers the entities represented by the data. The chosen ontology should describe and interconnect all entities considered in the integration process. In order to accomplish this task, it will be necessary to perform a survey into the literature to discover if the available ontologies have all considered concepts. Alternatively, the vocabulary of various ontologies can be used. In the case that no available ontology satisfies the requirements, a new one should be built.

The second step (2) is to analyze the chosen ontology in order to identify the rigid and anti-rigid types, the definitional relationships (which generate the relational dependencies) and the essential properties of the concepts. In the case that a foundational ontology has been chosen, this step could be very simple. However, when a domain ontology has been chosen, a lot of effort could be necessary. Moreover, this step requires the knowledge of an expert in the domain. It is an essential task in order to identify the entities that exist in the real word.

The third step (3) is to analyze the datasets in order to identify the entities that can be mapped to the ontology's rigid types (identified in the step (2)) through their names and essential properties.

In this approach, as a first basic indication, we assumed that two individuals are the same when they have the same name. However, considering geological models, for instance, this statement is not always true, since we can find many cases where a same name is applied to distinct ontological objects (polysemy). The concept *Rock*, for instance, can refer to a *Rock Sample*, to a *Rock Unit* and so forth.

In order to define if two modeled concepts are the same, independent of their names, we proposed the identification of their rigid properties, allowing structural matching. Thus, we considered that two entities are the same if (1) they are described by the same set of attributes, (2) each one of these attributes is instantiated to the same values and (3) each entity are related with the same objects in reality (if they have concrete existence). This task will be as easier as the language applied for the representation keeps a fair syntactical and ontological capability for representing world entities. Guarino, in [Guarino 1998], describes this capability as the ontological commitment of the language with the entities of the representation.

The forth step (4) is to analyze the identified entities under the view of their ontological properties: identity (supplies (O) or carries (I) identity); rigidity (R); relational dependence (D); and unity (U). For each of these properties, it is necessary to analyze if the instances of the entities hold (+), not hold (-), or can hold or not ( ) the property without affecting their existence. This analysis will increase the semantic of each entity found in heterogeneous datasets, resolving the lack of expressivity of representational languages. This is a very critical step since it will allow the mapping of entities from different datasets to the same ontology.

Finally, the last step (5) is to perform the mapping between the entities identified in step (3) and classified in step (4) with the chosen ontology and its concepts analyzed in step (2).

Thus, applying all these five steps correctly, data and information integration can be achieved, since integration in this context means finding correspondence between entities without merging the corresponding instances and by identifying the entities of the real world that the data describe [Guizzardi 2005].

7 ONTOLOGICAL ANALYSIS OF STANDARD DATA FORMATS

In order to exemplify the proposed methodological approach, we will use as a case study the communication standard formats that are going to become standard in the modeling chain: LAS, WITSML, and RESQML. Despite being an application domain particularly difficult, this case of study is interesting in the context of information integration because it describes and interconnects concepts that must match with different geological models considered in reservoir modeling. The result of the performed analysis is also available in [Werlang et al. 2014].

7.1 Step 1: Choosing the ontology

We started our analysis by searching and studying the current solutions for geological knowledge formalization, described in Chapter 5, looking for ontologies that cover the geological concepts that exist in real word and are represented in different models along the petroleum chain and could have their data stored in the data standard formats analyzed in this case study.

As a result of our first analysis, we ended up choosing the *Basic Geology* ontology. Thus, instead of developing the ontology from scratch, we reused it, once its structure was already set up by experts. Figure 5.4 presents our modified version of the chosen ontology adherent to ontological principles of modeling.

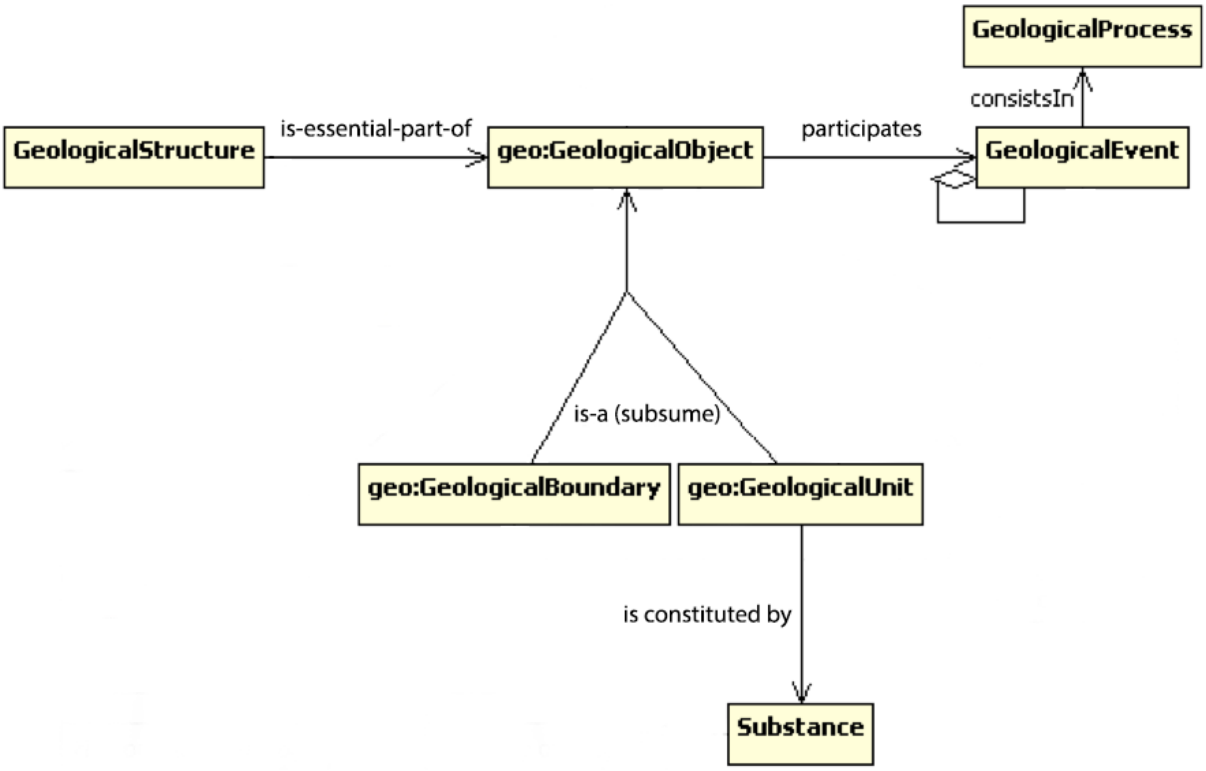


Figure 7.1: The top level *Basic Geology* ontology modified from [Mastella 2010] to describe the ontological meaning of relationships.

7.2 Step 2: Analyzing the chosen ontology

Then, following the proposed methodological approach, we analyzed the main concepts of the chosen ontology in order to identify their rigid types and their definitional relationships. The analyzed concepts were: *Substance*, *Geological Boundaries*, *Geological Units*, *Geological Structures* and *Geological Event*. We started our analysis by the main concepts related with our case study: *Geological Boundaries*, *Geological Units* and *Geological Structures*. All of them are rigid objects that preserve the unity property. This means that their instances in reality cannot stop being instances of these objects or they would disappear. Along with the concept *Substance*, they are the most important concepts for data. We will detail the ontological properties and relationships of these concepts in the way geologists usually conceive them.

Substances, whose nature is detailed in the *Substance* ontology (such as *Rocks*), fill *geological units*, making possible geological units to exist in space. It is equivalent to the meta-type *amount of matter* as described in [Guarino and Welty 2009]. A geological unit provides unity to an instance of rock, but not identity. The identity of a rock is given by its internal properties, such as composition, texture, fabric. A sub-class of a substance needs to preserve the same identity properties. A taxonomy of the concept rock would include carbonate and siliciclastic rocks, but rock cannot subsume the concept sill (a horizontal intrusion of rock) as a sub-class of rock, since the identity of sill is not provided by a lithology, but by some lithological unit with a particular format and position.

The relationship between a geological unit and a rock is called constitution. When we say that a geological unit has some property like granulometry, we are meaning that the substance that constitutes a lithological unit has this property, but not the unit itself. When we assert that a lake is salty, we indeed refer to the property of the water that is inside the lake. Taking care of preserving the independence and the specific properties of two concepts collocated in space such as rock and geological unit will allow us to integrate them within other applications in a much easier way.

Geological boundaries, in contrast with geological units, have no internal constitution. However, they are still rigid concepts that are an inseparable part of a geological unit (according to the classification of [Guizzardi 2005]). *Geological structures* are also rigid objects and inseparable parts of geological units, whose identity is defined by geometric internal and external properties.

These few objects - boundaries, structures, units and substance - are omnipresent in earth models and are specialized, extended and derived to represent all the different aspects that some model aims to emphasize. Most of problems in earth model integration are related to the problem that substance and rigid objects are collocated in the space (such as the sill and rock that constitute the sill) but still have proper independent ontological identity. The misuse of rock and its several types of portions (sample, core, unit, etc.) can led to wrong geological interpretations inside the information systems. It is important to stress that only rigid objects and those objects that they subsume (their subclasses) have instances, which means that any instance manipulated by some software application can be mapped to these few objects and further integrated. That statement is the basis for our approach. Figure 7.2 illustrates the explained relationships.

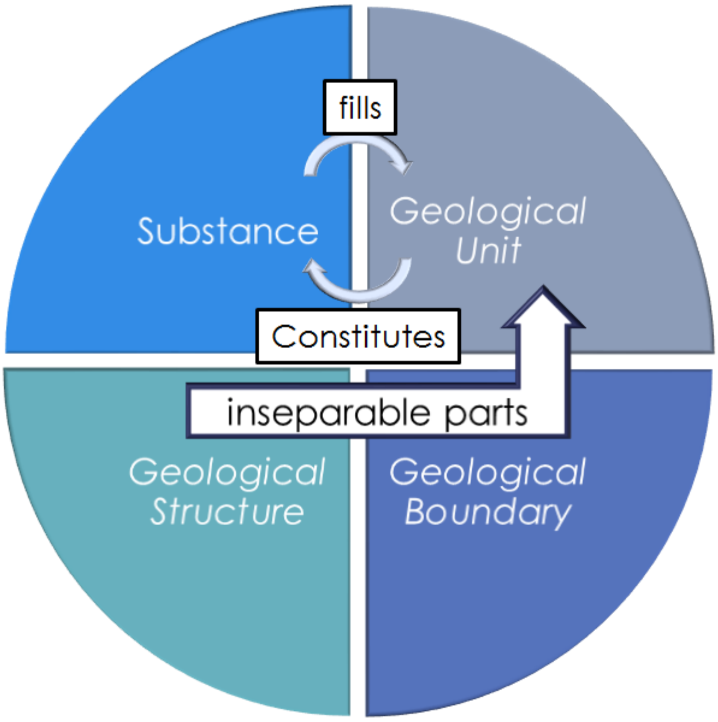


Figure 7.2: Relationships among the main entities considering an earth modeling scenario.

Other concepts that will be further analyzed in this work have an anti-rigid identity, which means that they can preserve their existence when they lose their identity. Such concepts are existentially dependent of another concept to exist. A reservoir[1], a source-rock and a trap are examples of roles played by rock units and geological structures, without configuring a specialization of rock unit. Any instance of reservoir or a source rock is necessarily an instance of a *geological unit*, and an instance of a trap is necessarily an instance of *geological structure*.

The *Basic Geology* ontology does not describe concepts to represent the different bodies of rocks created by human actions, such as cores and samples. We assume that a concept in a distinct domain ontology related to well development would subsumes the concepts *Core* and *Sample* utilized in our classification. We named it *SampleOfRock* to be mentioned in this work and it represents all partial exposition of a rock, spatially delimited by human action, such as core, samples and thin-section. Also, we consider that the concept *Rock* is a specialization of *Substance*.

Besides the endurant universals described above, the *Basic Geology* ontology also describe the perdurant concept *Geological Event*, that instantiates individuals that are composed by tem-

---

[1] We consider in this work that the existence of a reservoir is related to a rock unit with presence of oil, not only to a rock with intrinsic properties like porosity and permeability. The relational dependence between rock unit and petroleum allows us to classify reservoir as an anti-rigid object. If we would consider a reservoir as a rock with some amount of porosity and permeability, independent of having petroleum or not, it would be a rigid concept.

poral parts and have no physical representations in reality. A geological event may consist of a single *Geological Process* (e.g. the deposition of a sedimentary unit) or be composed of multiple geological processes (e.g., a graben filled with syntectonic sediments deposited while the faults limiting the graben were in movement). The various specializations of geological process, corresponding to creation, destruction or transformation of geological matter, are detailed in the *Geological Process* sub-ontology.

The *Basic Geology* ontology is complemented by universals that have no identity and are existentially dependent of other objects. They describe the properties that characterize the concepts and are usually of special interest in reservoir characterization. Attribute domains are associated to conceptual spaces and integrated through rigid objects. However, we will not discuss here how to deal with interoperability in the case of attributes.

7.3 Step 3 and 4: Analyzing the standard formats data files and the identified entities

In this section, we will present the result of the analysis performed with the standard formats data files. This section contemplates the steps 3 and 4 from the proposed methodology. Each data standard format has its own subsections: LAS (7.3.1 and 7.3.2) WITSML (7.3.3 and 7.3.4) and RESQML (7.3.5 and 7.3.6).

In our analysis, we considered only those objects that exist a priori in the natural world. Although others data-objects may have some data related to those analyzed objects and have substantial information to the modeling process, they are not within the scope of this work.

7.3.1 LAS: Step 3

We analyzed each data section set, including all parameter and definition data described in the *LAS 3.0 File Structure Specifications* [Heslop et al. 2000], searching for possible mappings with the chosen ontology. Between the analyzed data sections, we identified two data sections that store data with possible mappings with the chosen ontology. A brief description of the identified data sections is listed in the Table 7.1. Figures 7.3 and 7.4 provide more details about the possible mappings found.

Core Data Section	It includes information about the core, such as *core source*, *core type*, *primary formation cored*, *core oil and water saturation and volume*.	Substance, Geological Unit.
Top Data Section	It includes information about the formation tops, such as *formation top source* and *formation name*.	Geological Unit, Geological Object.

Table 7.1: The selected LAS data sections and their descriptions.

The *Core Data Section* contains data sections that store data referring to attributes of *Oil* and *Water* that can be mapped to concepts from *Substance* sub-ontology. This data set also has parameters with possible mapping, like *Core Source* (mapped to *Mineralogy* sub-ontology) and *Core Type* (mapped to *Lithology* sub-ontology). However, both concepts represent the substance *Rock*, described in the *Substance* sub-ontology. Another parameter of the core data section is the *Primary Formation Cored*, which can be mapped to *Geological Unit* sub-ontology.

The *Top Data Section* has the parameter *Formation Top Source* that refers to a *Lithostratigraphic Unit* described in the *Geological Unit* sub-ontology. This concept is defined by data like *Formation Name*, which is an attribute of a *Sedimentary Unit* from *Geological Object* sub-ontology.

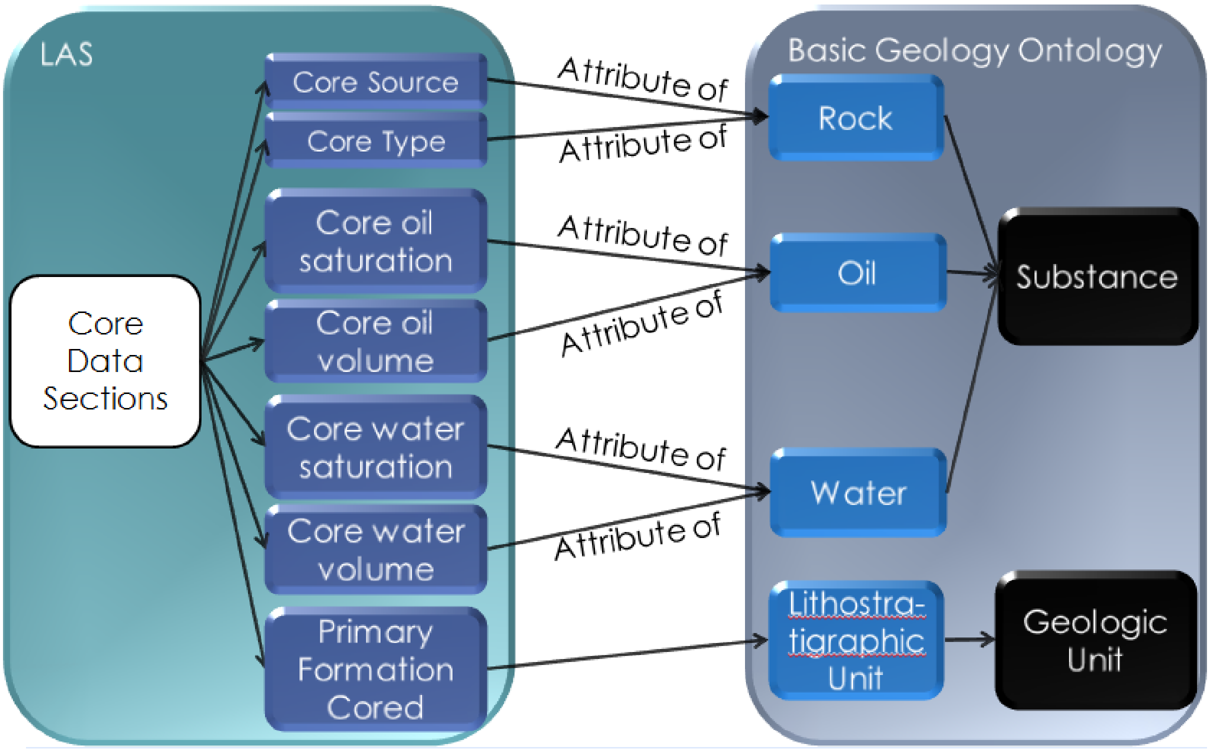


Figure 7.3: Possible mappings found between the LAS standard and *Basic Geology* ontology: *Core Data Section*.

7.3.2 LAS: Step 4

Our next analysis shows the classification of the geological concepts found in the selected data set sections, according to the properties of identity, rigidity, essentiality and unity. The analyzed concepts were: *Lithological Unit*, *Formation*, *Core*, *Oil*, *Water* and *Rock*.

A *Lithological Unit* is a body of rock that is sufficiently distinctive and continuous for being mapped. The concept *Lithological Unit* offers the principle of identity for its instances, since geologists are able to distinguish one instance from another by observing their visual

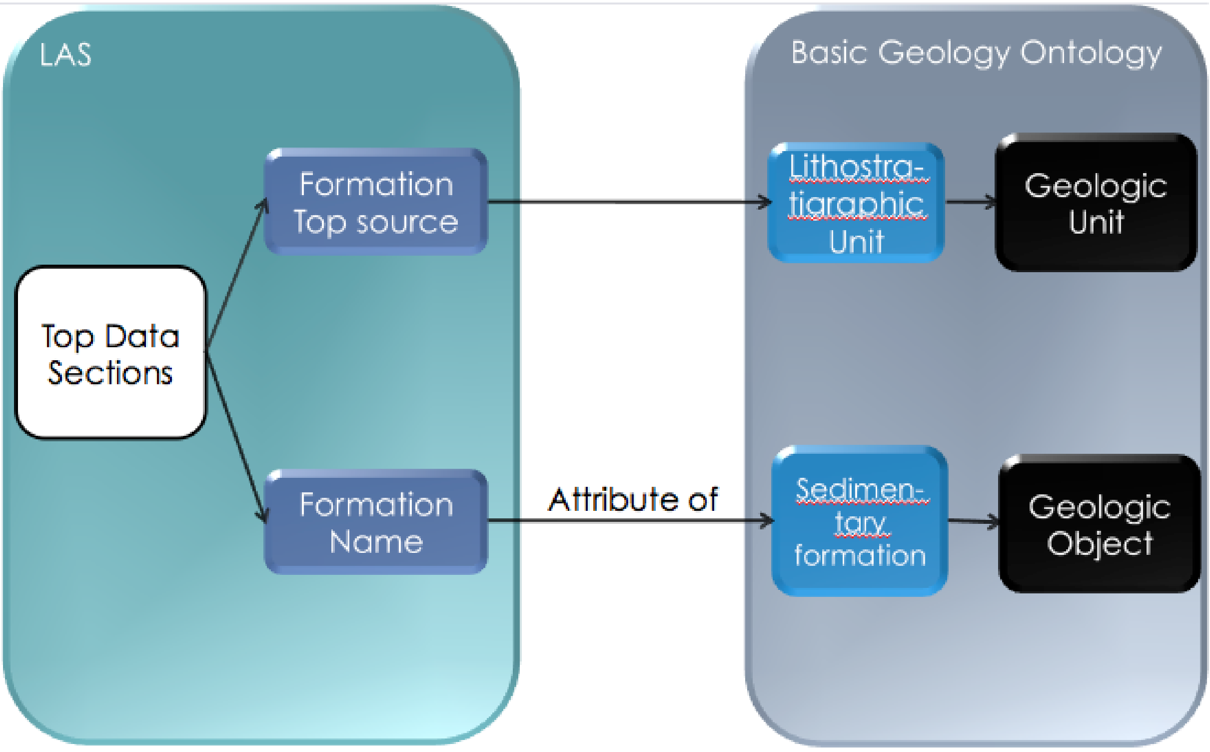


Figure 7.4: Possible mappings found between the LAS standard and *Basic Geology* ontology: *Top Data Section*.

characteristics. Also, this concept provides the principle of unity for their instances. This principle involves the observation of discontinuities in the visual characteristics of some body of rock, allowing the geologist to distinguish different lithological units. Furthermore, this concept is rigid, since its instances cannot fail to be so, unless it ceases to exist.

If the age and stratigraphic position of a *Lithological Unit* can be determined, this unit is referred as a *Formation* and it is given a name in order to support stratigraphic correlation. Even if the identity of the concept *Formation* is defined by rigid properties (like the lithological property age), there are no instances of the concept *Formation* that are not instances of *Lithological Unit*. Actually, a *Formation* does not exist until some *Lithological Unit* is identified as such by geologists. This shows an existential dependence that is not acceptable for a rigid concept. Consequently, *Formation* is an anti-rigid concept and plays a “role” for some *Lithological Unit*. So, in terms of integration, a *Formation* should be mapped to a *Lithological Unit* in other models.

A *Core* is a cylindrical sample extracted from a *Lithological Unit* that is being drilled. It is composed by one or more types of rocks. It is a specialization of *SampleOfRock* and it has a relationship of constitution with *Rock*. It is a rigid concept that preserves identity and unity.

Among the analyzed concepts, the concepts *Oil* and *Water* do not have unity, since they can only be individualized by other concepts (like a barrel of oil, a bottle of water, or a sample of rock). They are rigid concepts: they cannot stop being *Water* and *Oil*, without stopping to exist. But they are both uncountable objects, whose instances need to be individuated through a container, for instance, a reservoir. *Water* and *Oil* are interesting for modeling in relation of their internal properties. When they are presented in some application, rigid and not unified objects preserve a relation of container or location with some rigid and whole entity (e.g., oil is present within a reservoir). So, rigid and not unified objects are identified by their internal properties rather than by their spatial characteristics.

Less intuitively, the concept of *Rock* (specialization of *Substance*) in Geology is similar in ontological terms to *Oil* and *Water*. *Rock* is a rigid concept with no unity property. It is individuated by some other entities, typically a *Lithological Unit*, a *Core*, a *Sample* or a *Layer* in an outcrop, with whom rocks have the ontological relation of constitution.

The result of this analysis is presented in the Table 7.2.

Type	Ontological Identity	Rigidity	Relational Dependence	Unity	Ontological Category
Oil	O+I+	R+	D-	U-	Quantity
Water	O+I+	R+	D-	U-	Quantity
Rock	O+I+	R+	D-	U-	Quantity
Lithological Unit	O+I+	R+	D-	U+	Kind
SampleOfRock	O+I+	R+	D-	U+	Kind
Formation	O-I+	R-	D+	U-	Role

Table 7.2: The result of the classification realized with LAS types.

7.3.3 WITSML: Step 3

The WITSML standard enumerates 113 types with valid values for those elements. Analyzing these enumerated types, we identified 9 types whose values can be mapped to concepts from the Basic Geology ontology. A brief description of the identified type names is listed in the Table 7.3. Figures 7.5, 7.6 and 7.7 provide more details about the possible mapping between the presented ontology and the involved WITSML data objects.

7.3.4 WITSML: Step 4

As we did with LAS types, we classified the geological concepts found in the selected WITSML data objects according to the presented properties from foundational ontologies. The identified concepts were: *Oil*, *Water*, *Gas*, *Rock*, *Cuttings Samples*, *Formation*, *Core* and *Mud*. The concepts *Oil*, *Water*, *Rock*, *Formation* and *Core* were already analyzed in the LAS analysis and were already classified in Section 7.3.2 (result presented in the Table 7.2). Thus, the concepts that remain to be classified are: *Gas*, *Cutting Samples* and *Mud*.

The type *Gas* has no unity like the concepts *Rock*, *Oil* and *Water*, since all these concepts can only be individualized by other concepts. We considered Gas as a specialization of *Substance*.

Type Name	Description	Ontology Concept
LithologyType	It enumerates lithology values, such as *basalt* and *sand*, as well mineral names, such as *dolomite*, *clay* and *feldspar*.	Substance, Mineralogy
LithostratigraphyUnit	It specifies the unit of lithostratigraphy: *formation*, *member* and *bed*.	Geological Unit
MatrixCementType	It enumerates lithology matrix/cement descriptions, such as *dolomite* and *calcite*.	Mineralogy
MudClass	It defines the class of a drilling fluid, such as *water based*, *oil based* and *pneumatic* (gas based).	Substance
MudSubClass	It defines mud subtype at event occurrences, such as *brackish water*, *diesel oil-based* and *natural gas*.	Substance
QualifierType	It enumerates values that represent the type of qualifier in lithology, such as *dolomite*, *calcite*, *clay* and *feldspar*.	Mineralogy
StimFluidSubtype	It enumerates fluid sub types, such as *fresh water*, *oil* and *carbon dioxide*.	Substance
StimFluidType	It enumerates the fluid used for some stage of the stimulation job, such as *gas*, *oil-based* and *water-based*.	Substance
WellFluid	It enumerates the type of fluid being produced from or injected into a well facility, such as *gas*, *oil* and *water*.	Substance

Table 7.3: The selected WITSML type names and their descriptions.

Cuttings Samples are samples of rocks that were extracted from the wellbore during the drilling process. So, this concept offers the identity and unity principles for its instances. Also, it is a rigid concept, since its instances cannot fail to be so, unless it ceases to exist. We classify it as a specialization of the proposed concept *SampleOfRock*.

Mud is a term that is generally synonymous with drilling fluid and that encompasses most fluids used in hydrocarbon drilling operations, especially fluids that contain significant amounts of suspended solids (cuttings samples or samples of rocks), emulsified water or oil. Mud has

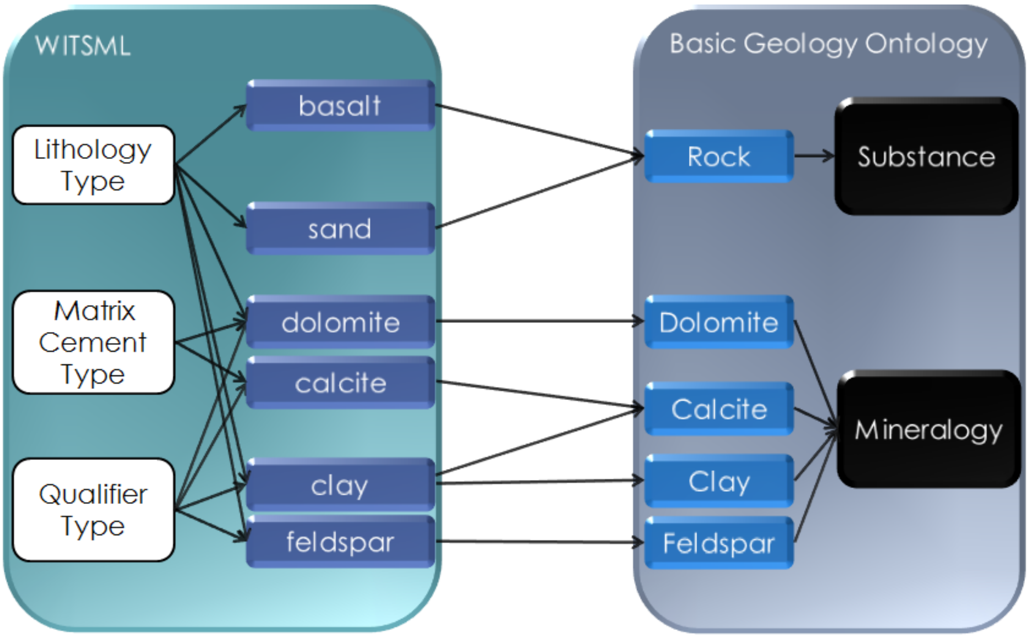


Figure 7.5: Possible mappings found between WITSML data-objects and *Basic Geology* ontology: *LithologyType*, *MatrixCementType* and *QualifierType*.

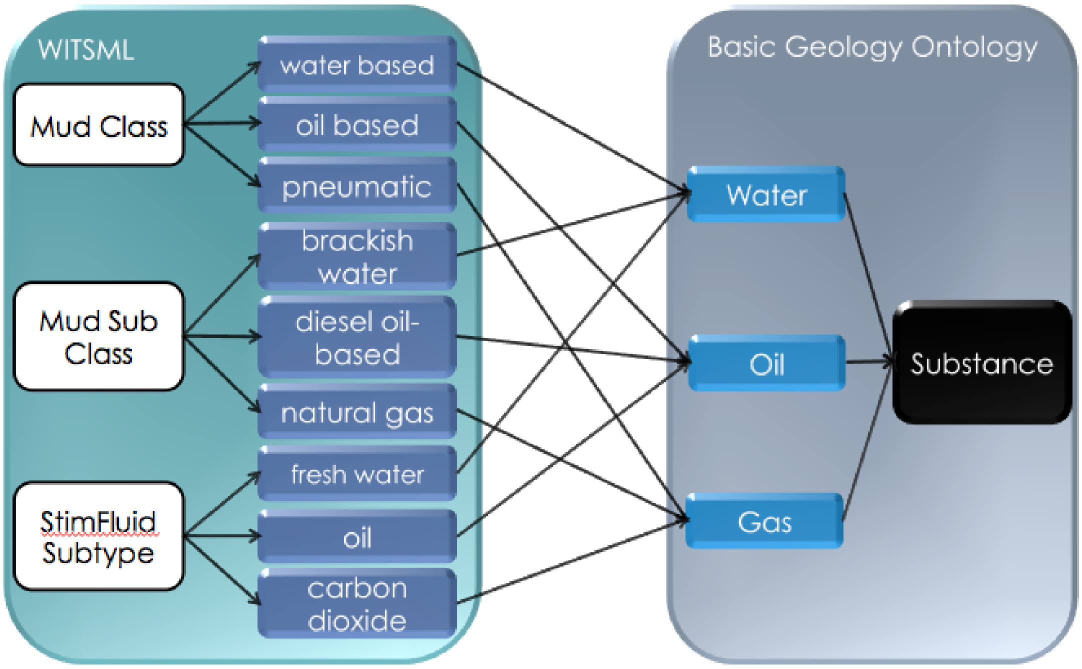


Figure 7.6: Possible mappings found between WITSML data-objects and *Basic Geology* ontology: *MudClass*, *MudSubClass* and *StimFluidSubType*.

not unity and includes all types of water-base, oil-base and synthetic-base drilling fluids, which means that its instances are the instances of several other rigid concepts which have their own identity. Mud corresponds to the ontological type category (rigid concepts that have not proper identity and group other instances using some property of interest).

The result of this analysis is presented in the Table 7.4.

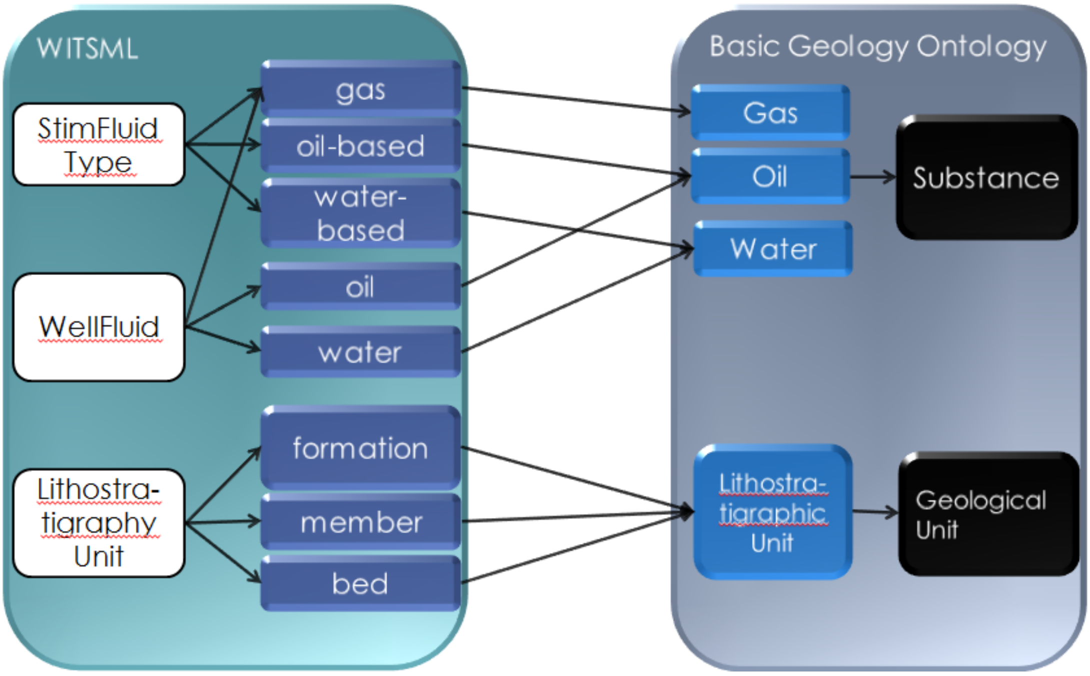


Figure 7.7: Possible mappings found between WITSML data-objects and *Basic Geology* ontology: *StimFluidType*, *WellFluid* and *LithostratigraphyUnit*.

Type	Ontological Identity	Rigidity	Relational Dependence	Unity	Ontological Category
Gas	O+I+	R+	D-	U-	Quantity
Mud	O-I-	R	D-	U-	Mixin

Table 7.4: Ontological analysis of the identified WITSML entities.

7.3.5 RESQML: Step 3

We analyzed the data types referenced by elements and attributes in RESQML data-object, as we did with WITSML data types. We started our analysis by the main classes of data-objects (features, interpretations, representations and properties) followed by the data-objects within these classes. For this, we used the "official" source for available data-objects: RESQML V2.0 schema (XSD) files[2].

After analyzed all available data-objects, we identified that only those related to *features* are really relevant to our work, since this main class groups together the data-objects that have physical existence at some point during the exploration, development, production or abandonment of a reservoir. Between them, we focus our analysis on those that exist in the subsurface and are directly involved with the modeling process. We considered both those objects that exist a priori in the natural world (geological features such as fault and horizon) as well those objects

---

[2] Available for download at http://www.energistics.org/reservoir/resqml-standards/current-standards

that exist by the action of humans (technical features such as well and wellbore). Although others data-objects may have some data related to those analyzed objects and have substantial information to the modeling process, they flee the scope of this work.

Moreover, it is important to note that the RESQML V2.0 already allows a full integration with the WITSML data-objects. Some of them were already supported by V1.1, such as well, wellbore, trajectory, formation marker and well log. In the V2.0, it is also possible to integrate WITSML rig-issued well data, such as trajectory and associated raw measurements, and well analysis results stored in other data formats, including PPDM standards. The full integration of RESQML V2.0 and WITSML V1.4.1.1 was developed by a Energistics?s SIG sub-group that focused on using the WITSML data model to abstract a simplified WITSML well consisting of three fundamental objects: its trajectory, either in depth and/or in time; a *WellboreFrame*, which is used to reference the WITSML log data; and its extension referring to WITSML marker data. These objects refer WITSML but now exist in the RESQML data model, which allows them to access existing RESQML capabilities [Deny et al. 2013]. Thus, some RESQML data-objects (such as well and wellbore) are related to WITSML concepts and were already classified in Section 7.3.4 (result presented in the Table 7.4).

We identified four simple data types (from the file "Geological.xsd") that enumerate geological objects with possible mapping with the chosen ontology. A brief description of the identified data types is listed in the Table 7.5. Figures 7.8 and Figure 7.9 provide more details about the possible mapping between the *Basic Geology* ontology and the involved RESQML data-objects.

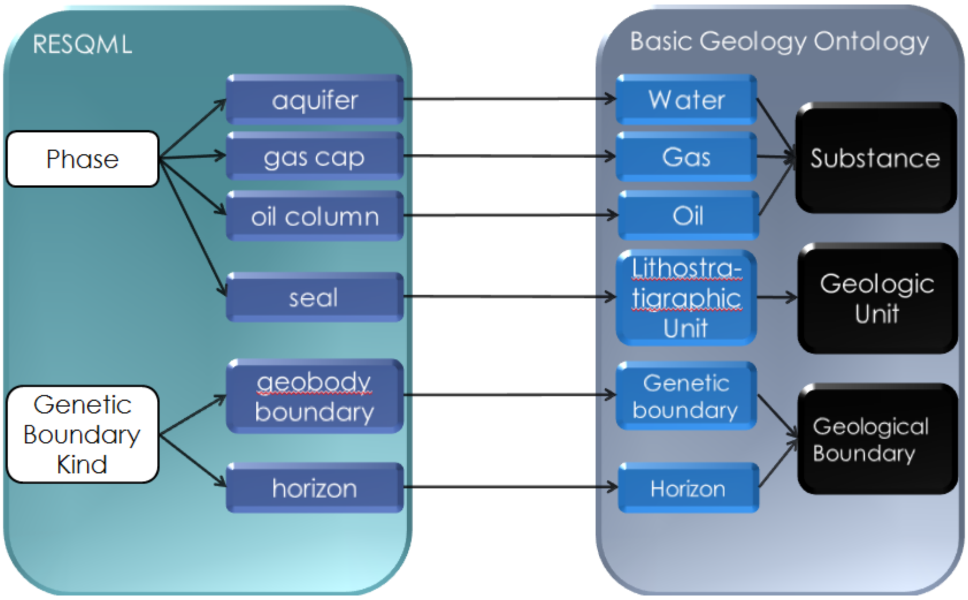


Figure 7.8: Possible mappings found between RESQML data-objects and *Basic Geology* ontology: *Phase* and *GeneticBoundaryKind*.

Type Name	Description	Ontology Concept
Phase	It describes and enumerates the possible rock fluid phase units in a hydrostatic column (either gases or liquids): *aquifer*, *gas cap*, *oil column* and *seal*.	Substance, Geological Boundary
GeneticBoundaryKind	It describes and enumerates the types of genetic boundary feature: *geobody boundary* and *horizon*.	Geological Boundary
TectonicBoundaryKind	It describes and enumerates the types of tectonic boundaries: *fault* and *fracture*.	Geological Boundary, Geological Object
FluidContact	It describes and enumerates values used to indicate a specific type of fluid boundary feature: free water contact, *gas oil contact*, *gas water contact*, *seal* and *water oil contact*.	Geological Boundary

Table 7.5: The selected RESQML data types and their descriptions.

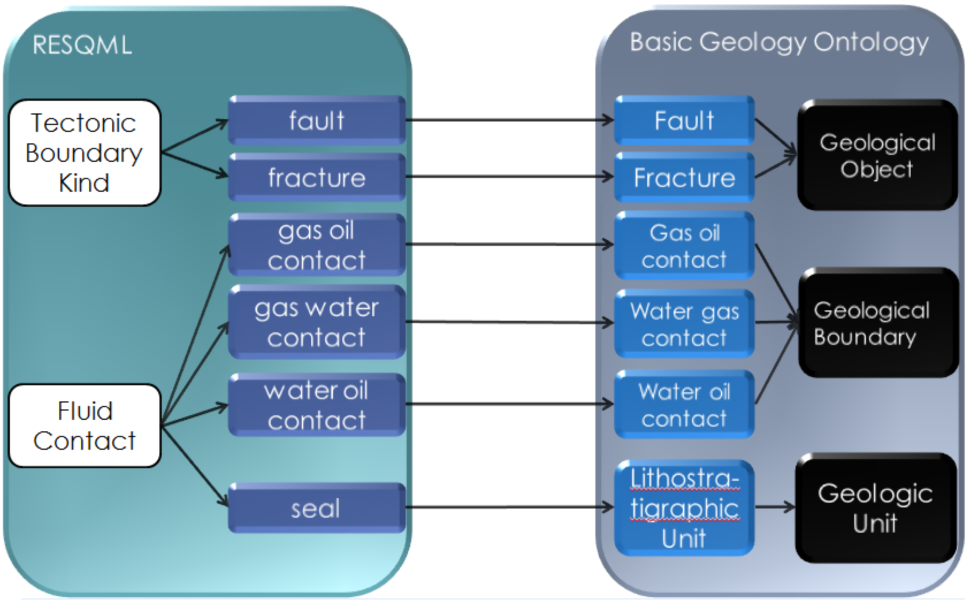


Figure 7.9: Possible mappings found between RESQML data-objects and *Basic Geology* ontology: *TectonicBoundaryKind* and *FluidContact*.

7.3.6 RESQML: Step 4

Between the identified geological objects, we still have to analyze five of them: *seal*, *horizon*, *geobody boundary*, *fault* and *fracture*. The others were already analyzed previously. However, before performing the ontological analysis of these concepts (according to the properties of identity, rigidity, essentiality and unity), it is important to understand the following meanings:

- *Fault*: is a break within geological material across which there exists an observable displacement (corresponding to the offset of segments or points that were once continuous or adjacent). As an object, a fault can be approximated as a very thin tabular volume possibly made of brittle rock. Within earth models, faults are frequently represented as mere surfaces or as volumes of zero thickness. Depending on the relative direction of displacement between the rocks and fault blocks, on either side of the fault, the fault movement is described as normal, reverse or strike-slip. Thus, we can assert that a fault separates different geological units that are not parts of the faults but only neighboring it. However, even when an observed geological discontinuity does not separate different geological units, it may still be a fault, since a fault can locally separate rock entities that belong to one thick unit.
- *Horizon*: according to the definition given in the glossary of the book “Shared Earth Modeling” [Perrin and Rainaud 2013], horizon is “a term used for designating either a unit boundary or a remarkable bed of small thickness. It can also correspond to a sedimentary boundary”. A horizon is a discontinuity in the rock volume whose main direction is basically horizontal. A horizon may be a surface where seismic waves are reflected and thus correspond to a seismic horizon. In this work, we adopt only the second meaning, and then a horizon has not substance.

Therefore, it is possible (and essential for geomodeling of geological mapping) to consider faults and horizons as objects, because things that can be observed can materialize both concepts. Then, we can say that they offer the principle of identity and unity for their instances. Also, both concepts are rigid, since their instances will remain what they are along the whole existence. A fault or a horizon may change its name or its age but there is practically no chance that it changes its nature. A fault will always remain a fault; a horizon will always remain a horizon. The same classification is applied to *geobody boundary* and *fracture*.

Finally, the geological object *seal* is an impermeable volume, which provides a seal that prevents the continuation of the upward vertical movement of the oil and gas. We can say that being a seal is a role played by some rock units under specific conditions. It is possible to individuate a seal and counting it. Moreover, it is necessary to include a rigid type — *lithological unit* — that guarantees the identity of the instances of seal. The result of this analysis is presented in the Table 7.6.

Type	Ontological Identity	Rigidity	Relational Dependence	Unity	Ontological Category
Fault	O+I+	R+	D-	U+	Kind
Horizon	O+I+	R+	D-	U+	Kind
Geobody boundary	O+I+	R+	D-	U+	Kind
Fracture	O+I+	R+	D-	U+	Kind
Seal	O-I+	R-	D+	U-	Role

Table 7.6: The result of the classification realized with RESQML types.

7.4 Step 5: Performing the mapping

Finally, our last step was performing a mapping between the identified entities from the analyzed standard formats and from the *Basic Geology* ontology. We illustrate the result of this mapping process, emphasizing the identified data-objects from the analyzed standards and the mapped entities from the chosen ontology.

Thus, Figure 7.10 provides the mappings to entities from *Substance* sub-ontology. In this case, all analyzed standard data formats have data sets that enumerate concepts mapped to entities from the *Substance* sub-ontology.

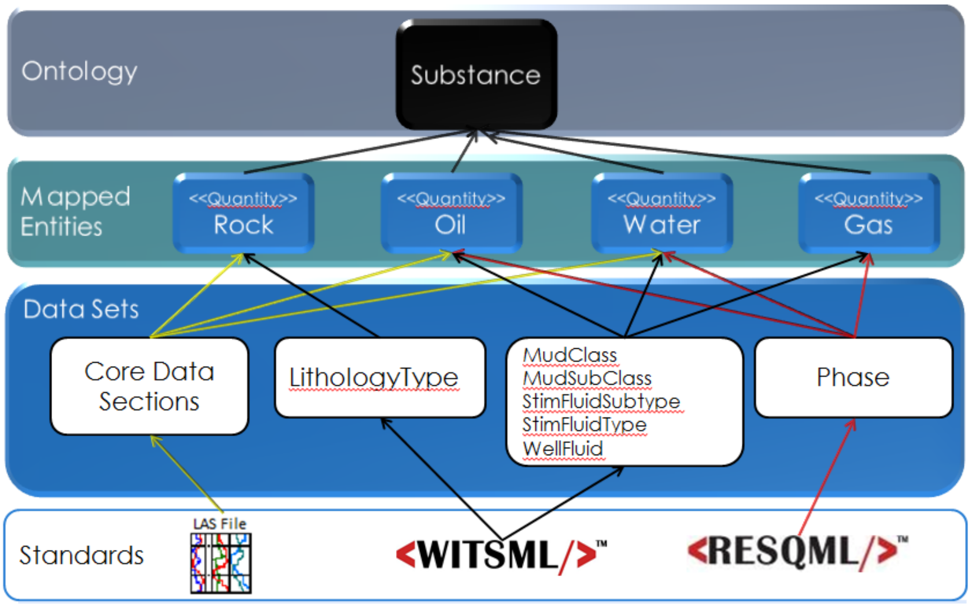


Figure 7.10: Possible mappings to entities from *Substance* sub-ontology.

Figure 7.11 provides the mappings to entities from Geological Unit sub-ontology. In this case, all analyzed standard data formats have data sets that enumerate concepts mapped to entities from the *Geological Unit* sub-ontology. Note that between the concepts enumerated in the data sets *Phase* and *FluidContact*, from RESQML standard data format, only the concept *Seal* can be mapped to *Lithostratigraphic Unit*.

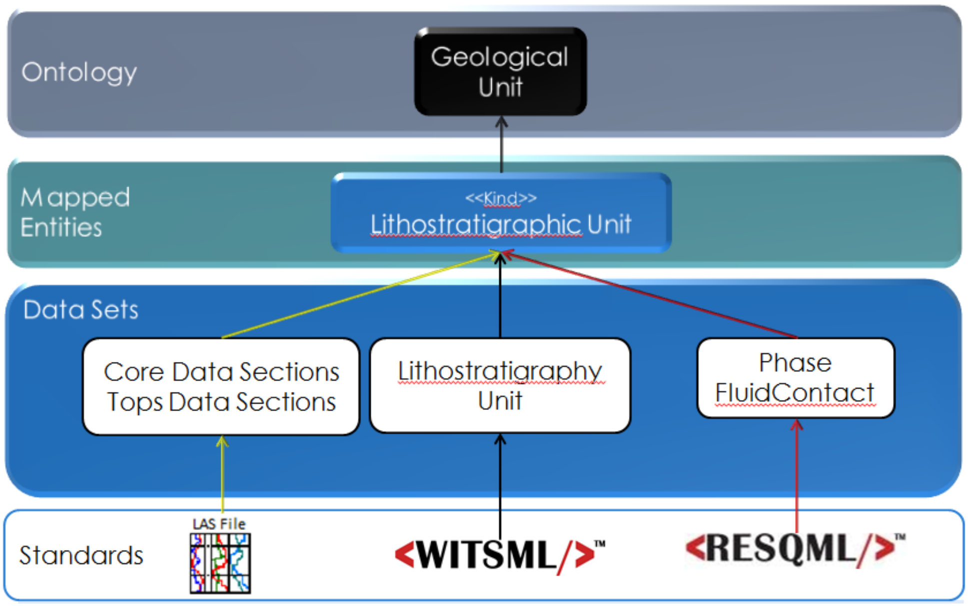


Figure 7.11: Possible mappings to entities from *Geological Unit* sub-ontology..

Figure 7.12 provides the mappings to entities from Geological Boundary sub-ontology. In this case, only the data set *GeneticBoundaryKind*, from RESQML standard data format, enumerates concepts mapped to entities from Geological Boundary sub-ontology.

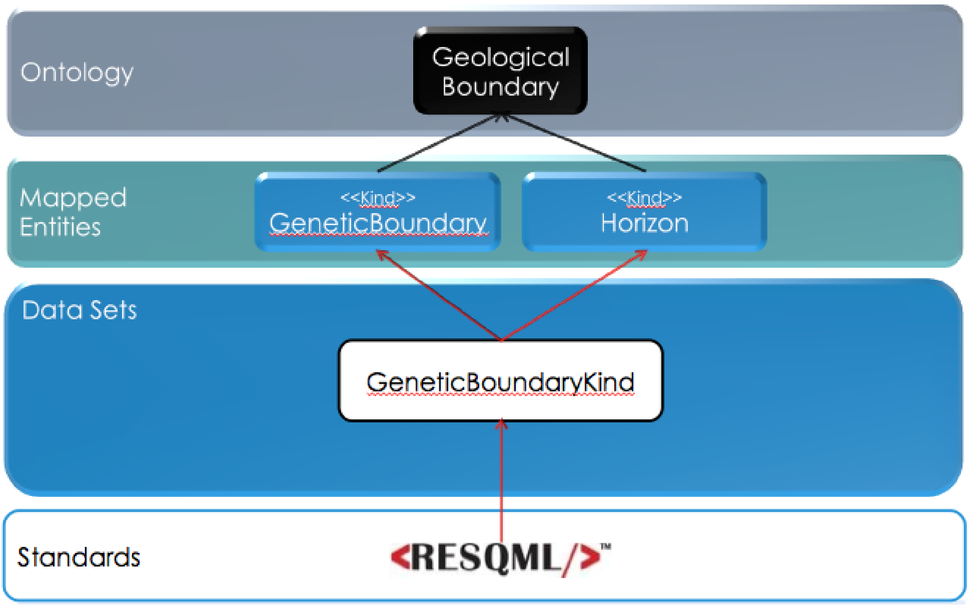


Figure 7.12: Possible mappings to entities from *Geological Boundary* sub-ontology.

Finally, Figure 7.13 provides the mappings to entities from Geological Object sub-ontology. In this case, LAS and RESQML standards data formats have data sets that enumerate concepts mapped to entities from the *Geological Object* sub-ontology.

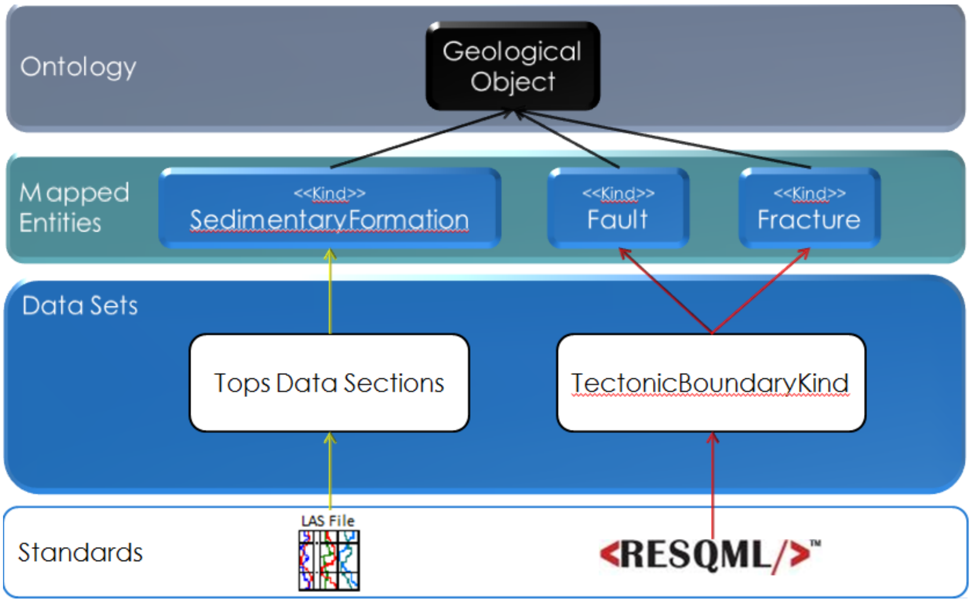


Figure 7.13: Possible mappings to entities from *Geological Object* sub-ontology.

7.5 Mapping validation and evaluation

In this section, we will present how the identified mappings were validated and the profiles of the professionals that have done that. Furthermore, we also intend to evaluate the possibility of automation of the proposed methodological approach by using a formal representational language to formalize the mappings between the standards and the chosen ontology.

In relation to the validation of the identified mappings, we had the honour of the participation of two highly qualified professionals with skills and expertise in ontologies and earth modeling. One of them, Jean-François Rainaud, is Doctor in Geophysics. He is presently Senior Project Manager at Institute Français du Petroleum et Energies Nouvelles (IFPEN), in France, in the field of Information Technology. In the last 15 years, he has become one of the biggest authority in earth modeling systems, which a large set of publications and practical projects. He presently plays a major role in the Energistics RESQML Special Interest Group (SIG), which gathers representatives of oil companies and software providers for creating a data exchange standard for geomodeling and develop the WITSML, RESQML, PRODML, MICRO ML standards.

The other one, Michel Perrin, is Doctor in Geology with extensive experience as a researcher in various fields of geology (structural geology, petrology, geochemistry), he dedicated the last twenty last years of professional activity on the subject of geomodeling. He intensely collaborated on this subject with BRGM, IFPEN and with various research groups specialized in computer science in France, Switzerland and Brazil, by means of ten joint doctoral works and of joint participations in various research projects supported by governmental authorities. His main concern has been developing practical solutions for putting geological knowledge at the center of the geomodeling process.

They all validate the mappings individually more than three times, emphasizing the classifications of the identified entities from the analyzed standard data formats and the mappings of them to the chosen ontology. Thus, we changed our classifications and mappings of the identified geological concepts many times until we get on the current situation.

In relation to the evaluation about the possibility of automation of the proposed methodological approach, we ended up that steps 1 and 2 are basically an activity of knowledge acquisition. Thus, since there are no syntactic aspects in the information about geological concepts that allow the identification of rigid and anti-rigid types, the automation of these steps is not possible to achieve in the current state of art.

We also evaluated that steps 3 and 4 could be automated if the rigid types and their essential properties were previously identified in the analyzed standard data formats. The current limitation refers to the fact that any mapping would be extremely dependent of syntactic aspects, such as the necessity of keeping the same name of concepts, the same name of attributes and their domain values. Thus, considering the present state of art, achieve this goal is still far.

Finally, in order to allow an automatic mapping between the standard data formats and the chosen ontology, using inference and reasoning mechanisms, a formal representational language should be used in order to allow the processing of the mappings.

8 CONCLUSIONS

In this work, we presented a methodology for ontological analysis of earth models, aiming in providing support for information integration and software interoperability. We founded our analysis in the philosophical background of ontological studies about the nature of the existence of being, updated with recent development in Computer Sciences.

Our claim is that the interoperability among earth models built and manipulated by different professionals and systems can be achieved by making apparent the intended meaning of the geological objects represented in the models. We described the ontological meaning of Geology concepts and relationships previously modeled in the Basic Geology ontology. Our work does not intend to cover the whole Geology domain in proper extension and detail. Thus, the few concepts that have been analyzed are those which are more frequently found in Earth models and which are of central importance for anchoring petroleum exploration models into raw data and into entities existing in the reality.

We showed that few ontological properties inhered to concepts - identity, rigidity, essentiality, dependence and unity - are enough to clarify the meaning of the modeled entities and to define the similarities and identities between models and data. These properties elucidate which are the concepts that are essential to the modeled reality and can be used for providing a unified view over this reality. In addition, the understanding of the modeling principles can lead to clean the future models of the entities that have no proper existence in reality (i.e., they are not rigid and have no instances). This allows modelers to develop naturally integrable models based on a common framework of the essential rigid concepts.

We also discussed several misconceptions about relationship meaning commonly found in Earth models. From these, the more significant are: the mix of the concepts geological unit (a delimited object in 3D space and time) and rock (the substance of what this object is made), the relation between geological unit and their spatial limits (boundaries, faults) and internal structure (geological structures, sedimentary structures), which are rigid concepts that are inseparable parts of lithological units and not properties of them.

The approach proposed in this work are a contribution towards a complete semantic integration of communication standard formats most used in the reservoir modeling chain. The analysis performed in this study corresponds to an essential step in the integration process. Clarifying the semantic of the geological concepts most used in petroleum exploration models, using a theoretical background of foundational ontologies, assists in solving problems of semantic heterogeneity. Among the several issues related to the integration of communication standard formats most used to exchange information in the earth modeling process, we particularly addressed here those related to:

- **Identify and understand the geological concepts most used in petroleum exploration modeling**. Firstly, we made an immersion in the literature in order to identify the major gaps in information integration and the newest solutions to deal with semantic integration (ontologies). Then, we analyzed the use of foundational ontologies to deal with semantic heterogeneity problems and earth models interoperability. Finally, we studied the reservoir modeling workflow in order to identify and understand the most used geological concepts.

- **Identify the ontologies developed for Geosciences and earth modeling**. We realized a survey in the literature to discover if the available ontologies for Geosciences and earth modeling have the identified geological concepts.
- **Map the geological concepts to the found ontologies**. We exemplified the mapping capability by analyzing the communication standard formats most used in the modeling chain (LAS, WITSML and RESQML), searching for entities semantically related with the geological concepts described in those ontologies, in order to make explicit the nature and properties of the geological objects found in each format.
- **Classify the geological concepts according to foundational ontologies primitives**. We classified the geological concepts found in each format, according to ontological meta-properties (identity, rigidity, essentiality and unity).

8.1 Future Work

The present work must be considered an approach for elucidate the semantics of the standards for data exchange in the earth modeling process. This work has opened various perspectives concerning what remains to be done in order to develop a complete framework for earth models integration.

Significant work remains to be done for studying the ontological properties of all the objects that have to be considered when building earth models. We have only analyzed some main concepts and properties attached to geological objects leaving apart some other important properties (permeability, granulometry, age, etc.). Also, we presented only a subset of the Geology concepts required to model a petroleum prospect. A deeper analysis will be detailed in further reports related to our project.

Once all geological concepts are identified and classified, completing the ontological analyzes started in this work, a complete conceptual model can be defined. Then, in order to allow an automatic mapping between the standards and the defined conceptual model, using inference and reasoning mechanisms, a formal representational language should be used. In this way, the mappings may be processed. Languages used to represent mappings between ontologies can be a source of inspiration for this challenge.

Finally, others standard data formats, such as GeoSciML and NADM, which were briefly discussed in this analysis, should deserve a deeper study in further studies due their growing importance for geological ontologies.

REFERENCES

[Abel 2001]ABEL, M. *The study of expertise in Sedimentary Petrography and its significance for knowledge engineering*. Thesis (PhD) — Computer Pos-Graduation Program, UFRGS, Porto Alegre, 2001.

[Abel et al. 2004]ABEL, M. et al. How to model visual knowledge: A study of expertise in oil-reservoir evaluation. In: GALINDO, F.; TAKIZAWA, M.; TRAUNMÜLLER, R. (Ed.). *Database and Expert Systems Applications*. Galindo, fernando and takizawa, makoto and traunmüller, roland. [S.l.]: Springer Berlin Heidelberg, 2004, (Lecture Notes in Computer Science, v. 3180). p. 455–464.

[Alonso et al. 2004]ALONSO, G. et al. *Web Services Concepts, Architectures and Applications*. [S.l.]: Springer-Verlag, Berlin, 2004.

[Batini, Ceri and Navathe 1992]BATINI, C.; CERI, S.; NAVATHE, S. B. *Conceptual database design: an Entity-relationship approach*. [S.l.]: Benjamin/Cummings Redwood City, CA, 1992.

[Belkin and Croft 1992]BELKIN, N. J.; CROFT, W. B. Information filtering and information retrieval: two sides of the same coin? *Communications of the ACM*, v. 35, n. 12, p. 29–38, 1992.

[Benerecetti, Bouquet and Ghidini 2000]BENERECETTI, M.; BOUQUET, P.; GHIDINI, C. Contextual reasoning distilled. *Journal of Experimental & Theoretical Artificial Intelligence*, v. 12, n. 3, p. 279–305, 2000.

[Bernstein and Haas 2008]BERNSTEIN, P. A.; HAAS, L. M. Information integration in the enterprise. *Communications of the ACM*, v. 51, n. 9, p. 72–79, 2008.

[Biju-Duval 2002]BIJU-DUVAL, B. *Sedimentary geology: sedimentary basins, depositional environments, petroleum formation*. [S.l.]: Editions Technip, 2002.

[Buccella, Cechich and Fillottrani 2009]BUCCELLA, A.; CECHICH, A.; FILLOTTRANI, P. Ontology-driven geographic information integration: A survey of current approaches. *Computers & Geosciences*, v. 35, n. 4, p. 710–723, 2009.

[Buccella et al. 2011]BUCCELLA, A. et al. Building a global normalized ontology for integrating geographic data sources. *Computers & Geosciences*, v. 37, n. 7, p. 893–916, 2011.

[Calì 2003]CALÌ, A. Reasoning in data integration systems: why lav and gav are siblings. In: *Foundations of Intelligent Systems*. [S.l.]: Springer, 2003. p. 562–571.

[Calì et al. 2003]CALÌ, A. et al. On the expressive power of data integration systems. In: *Conceptual Modeling—ER 2002*. [S.l.]: Springer, 2003. p. 338–350.

[Calvanese and Giacomo 2005]CALVANESE, D.; GIACOMO, G. D. Data integration: A logic-based perspective. *AI magazine*, v. 26, n. 1, p. 59, 2005.

[Calvanese et al. 2011]CALVANESE, D. et al. The mastro system for ontology-based data access. *Semantic Web*, v. 2, n. 1, p. 43–53, 2011.

[Calvanese, Giacomo and Lenzerini 2002]CALVANESE, D.; GIACOMO, G. D.; LENZERINI, M. A framework for ontology integration. In: IOS PRESS. *The Emerging Semantic Web—Selected Papers from the First Semantic Web Working Symposium*. Stanford, USA, 2002. p. 201–214.

[Calvanese et al. 2007]CALVANESE, D. et al. Mastro-i: Efficient integration of relational data through dl ontologies. In: CEUR ELECTRONIC WORKSHOP PROCEEDINGS. *Description Logics*. Brixen-Bressanone (near Bozen-Bolzano), Italy, 2007.

[Carbonera 2012]CARBONERA, J. L. *Raciocínio sobre conhecimento visual: Um estudo em Estratigrafia Sedimentar*. Dissertation (Master) — UFRGS, 2012.

[Chalupsky 2000]CHALUPSKY, H. Ontomorph: A translation system for symbolic knowledge. In: *KR*. [S.l.: s.n.], 2000. p. 471–482.

[Chaudhuri and Dayal 1997]CHAUDHURI, S.; DAYAL, U. An overview of data warehousing and olap technology. *ACM Sigmod record*, v. 26, n. 1, p. 65–74, 1997.

[Chum 2007]CHUM, F. *Use Case: Ontology-Driven Information Integration and Delivery A Survey of Semantic Web Technology in the Oil and Gas Industry*. 2007. Available from Internet: <http://www.w3.org/2001/sw/sweo/public/UseCases/Chevron/>.

[Cox and Richard 2005]COX, S. J.; RICHARD, S. M. A formal model for the geologic time scale and global stratotype section and point, compatible with geospatial information transfer standards. *Geosphere*, v. 1, n. 3, p. 119–137, 2005.

[Deny et al. 2013]DENY, L. et al. Delivering key witsml measured and interpreted well information through resqml v2 for reservoir characterization and flow simulation. In: SOCIETY OF PETROLEUM ENGINEERS. *SPE Annual Technical Conference and Exhibition*. New Orleans, Louisiana, USA, 2013.

[Doan and Halevy 2005]DOAN, A.; HALEVY, A. Y. Semantic integration research in the database community: A brief survey. *AI magazine*, v. 26, n. 1, p. 83, 2005.

[Endres et al. 2013]ENDRES, D. M. et al. How resqml version 2 can facilitate the update of a reservoir model. In: SOCIETY OF PETROLEUM ENGINEERS. *EAGE Annual Conference & Exhibition incorporating SPE Europec*. London, UK, 2013.

[Energistics and SIG 2014]ENERGISTICS; SIG, R. *RESQML Technical Usage Guide*. [S.l.], 2014.

[Energistics and SIG 2012]ENERGISTICS; SIG the W. *WITSML STORE Application Program Interface (API)*. [S.l.], 2012.

[Euzenat, Shvaiko et al. 2007]EUZENAT, J.; SHVAIKO, P. et al. *Ontology matching*. [S.l.]: Springer, 2007.

[Fagin et al. 2005]FAGIN, R. et al. Data exchange: semantics and query answering. *Theoretical Computer Science*, v. 336, n. 1, p. 89–124, 2005.

[Feijo 2010]FEIJO, F. J. *Noções Básicas de Geologia do Petróleo*. [S.l.], 2010.

[Fileto and Medeiros 2003]FILETO, R.; MEDEIROS, C. B. *A survey on information systems interoperability*. [S.l.], 2003.

[Firestone 2003]FIRESTONE, J. M. *Enterprise information portals and knowledge management*. [S.l.]: KMCI/Butterworth-Heinemann, 2003.

[Friedman et al. 1999]FRIEDMAN, M. et al. Navigational plans for data integration. *AAAI/IAAI*, p. 67–73, 1999.

[Gagnon 2007]GAGNON, M. Ontology-based integration of data sources. In: IEEE. *10th International Conference on Information Fusion, FUSION 2007*. Québec, Canada: IEEE, 2007. p. 1–8.

[Ge et al. 2011]GE, J. et al. Petroleum exploration domain ontology-based knowledge integration and sharing system construction. In: IEEE. *Network Computing and Information Security (NCIS), 2011 International Conference on*. Washington, DC, USA: IEEE Computer Society, 2011. (NCIS '11, v. 1), p. 84–88.

[Goh et al. 1999]GOH, C. H. et al. Context interchange: New features and formalisms for the intelligent integration of information. *ACM Transactions on Information Systems (TOIS)*, v. 17, n. 3, p. 270–293, 1999.

[Gruber 2003]GRUBER, T. *Ontology progress - What is an ontology*. 2003. Software World.

[Gruber 2008]GRUBER, T. Ontology (to appear in the encyclopedia of database systems). *Encyclopedia of Database Systems*, 2008.

[Gruber 1992]GRUBER, T. R. *Ontolingua: A mechanism to support portable ontologies*. [S.l.]: Stanford University, Knowledge Systems Laboratory, 1992.

[Guarino 1998]GUARINO, N. *Formal ontology in information systems: Proceedings of the first international conference (FOIS'98), June 6-8, Trento, Italy*. [S.l.]: IOS press, 1998.

[Guarino and Welty 2009]GUARINO, N.; WELTY, C. A. An overview of ontoclean. In: *Handbook on ontologies*. [S.l.]: Springer, 2009. p. 201–220.

[Guizzardi 2005]GUIZZARDI, G. *Ontological foundations for structural conceptual models*. [S.l.]: CTIT, Centre for Telematics and Information Technology, 2005.

[Guizzardi, Falbo and Guizzardi 2008]GUIZZARDI, G.; FALBO, R. de A.; GUIZZARDI, R. S. Grounding software domain ontologies in the unified foundational ontology (ufo): The case of the ode software process ontology. In: *CIbSE*. [S.l.: s.n.], 2008. p. 127–140.

[Guizzardi and Wagner 2010]GUIZZARDI, G.; WAGNER, G. Using the unified foundational ontology (ufo) as a foundation for general conceptual modeling languages. In: *Theory and Applications of Ontology: Computer Applications*. [S.l.]: Springer, 2010. p. 175–196.

[Guizzardi et al. 2013]GUIZZARDI, G. et al. Towards ontological foundations for the conceptual modeling of events. In: *Conceptual Modeling*. [S.l.]: Springer, 2013. p. 327–341.

[Guizzardi et al. 2004]GUIZZARDI, G. et al. An ontologically well-founded profile for uml conceptual models. In: SPRINGER. *Advanced Information Systems Engineering*. [S.l.], 2004. p. 112–126.

[Guizzardi and Guizzardi 2011]GUIZZARDI, R. S.; GUIZZARDI, G. Applying the ufo ontology to design an agent-oriented engineering language. In: SPRINGER. *Advances in Databases and Information Systems*. [S.l.], 2011. p. 190–203.

[Guizzardi et al. 2007]GUIZZARDI, R. S. et al. Towards an ontological account of agent-oriented goals. In: *Software Engineering for Multi-Agent Systems V*. [S.l.]: Springer, 2007. p. 148–164.

[Haas 2006]HAAS, L. Beauty and the beast: The theory and practice of information integration. In: *Database Theory–ICDT 2007*. [S.l.]: Springer, 2006. p. 28–43.

[Halevy et al. 2005]HALEVY, A. Y. et al. Enterprise information integration: successes, challenges and controversies. In: ACM. *Proceedings of the 2005 ACM SIGMOD international conference on Management of data*. Baltimore, Maryland, 2005. p. 778–787.

[Härder, Sauter and Thomas 1999]HÄRDER, T.; SAUTER, G.; THOMAS, J. The intrinsic problems of structural heterogeneity and an approach to their solution. *The VLDB Journal*, v. 8, n. 1, p. 25–43, 1999.

[Hasselbring 2000]HASSELBRING, W. Information system integration. *Communications of the ACM*, v. 43, n. 6, p. 32–38, 2000.

[Heslop et al. 2000]HESLOP, K. et al. *Log ASCII Standard Document – File Structures*. [S.l.], 2000.

[Hyne 2012]HYNE, N. J. *Nontechnical Guide to Petroleum Geology, Exploration, Drilling & Production*. [S.l.]: PennWell Books, 2012.

[Kalfoglou and Schorlemmer 2003]KALFOGLOU, Y.; SCHORLEMMER, M. Ontology mapping: the state of the art. *The knowledge engineering review*, v. 18, n. 01, p. 1–31, 2003.

[Kashyap and Sheth 1996]KASHYAP, V.; SHETH, A. Semantic and schematic similarities between database objects: a context-based approach. *The VLDB Journal—The International Journal on Very Large Data Bases*, v. 5, n. 4, p. 276–304, 1996.

[Kim and Seo 1991]KIM, W.; SEO, J. Classifying schematic and data heterogeneity in multidatabase systems. *Computer*, v. 24, n. 12, p. 12–18, 1991.

[Kimball and Caserta 2004]KIMBALL, R.; CASERTA, J. *The data warehouse ETL toolkit*. [S.l.]: John Wiley & Sons, 2004.

[Klein 2001]KLEIN, M. Combining and relating ontologies: an analysis of problems and solutions. In: *IJCAI-2001 Workshop on ontologies and information sharing*. Seattle, Washington, USA: [s.n.], 2001. p. 53–62.

[Litwin and Abdellatif 1986]LITWIN, W.; ABDELLATIF, A. Multidatabase interoperability. *Computer*, v. 19, n. 12, p. 10–18, 1986.

[Lorenzatti et al. 2011]LORENZATTI, A. et al. Ontological primitives for visual knowledge. In: *Advances in Artificial Intelligence–SBIA 2010*. [S.l.]: Springer, 2011. p. 1–10.

[Lorenzatti et al. 2009]LORENZATTI, A. et al. Ontology for imagistic domains: Combining textual and pictorial primitives. In: *Advances in Conceptual Modeling-Challenging Perspectives*. [S.l.]: Springer, 2009. p. 169–178.

[Madhavan, Bernstein and Rahm 2001]MADHAVAN, J.; BERNSTEIN, P. A.; RAHM, E. Generic schema matching with cupid. In: *VLDB*. [S.l.: s.n.], 2001. v. 1, p. 49–58.

[Mastella 2010]MASTELLA, L. S. *Semantic Exploitation of Engineering Models: Application to Petroleum Reservoir Models*. Thesis (PhD) — Ecole Nationale Superieure des Mines De Paris, 2010.

[Melnik, Adya and Bernstein 2007]MELNIK, S.; ADYA, A.; BERNSTEIN, P. A. Compiling mappings to bridge applications and databases. In: ACM. *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*. [S.l.], 2007. p. 461–472.

[Morgenthal and Kernochan 2005]MORGENTHAL, J.; KERNOCHAN, W. *Enterprise Information Integration: A Pragmatic Approach*. [S.l.]: Lulu.com, 2005.

[Noy 2004]NOY, N. F. Semantic integration: a survey of ontology-based approaches. *ACM Sigmod Record*, v. 33, n. 4, p. 65–70, 2004.

[Omdal 2006]OMDAL, S. The integrated information platform (iip) for resevoir and subsea production systems. In: *Technical report, Norwegian Oil Industry Association (OLF) 2005, POSC IntOPS SIG Regional Meeting*. [S.l.: s.n.], 2006.

[Perrin 1998]PERRIN, M. Geological consistency: an opportunity for safe surface assembly and quick model exploration. 3d modeling of natural objects. *GOCAD ENSG Conference - 3D Modeling of Natural Objects : A Challenge for the 2000's*, v. 3, p. 4–5, June 1998.

[Perrin et al. 2011]PERRIN, M. et al. Geological time formalization: an improved formal model for describing time successions and their correlation. *Earth Science Informatics*, v. 4, n. 2, p. 81–96, 2011.

[Perrin and Rainaud 2013]PERRIN, M.; RAINAUD, J.-F. *Shared Earth Modeling: Knowledge Driven Solutions for Building and Managing Subsurface 3D Geological Models*. [S.l.]: Editions Technip, 2013.

[Perrin et al. 2007]PERRIN, M. et al. Knowledge related challenges for efficient data fusion. *Data Fusion: Combining Geological, Geophysical and Engineering Data, SEG/AAPG/SPE Joint Workshop*, October 2007.

[Perrin et al. 2005]PERRIN, M. et al. Knowledge-driven applications for geological modeling. *Journal of Petroleum Science and Engineering*, Elsevier, v. 47, n. 1, p. 89–104, 2005.

[Rahm and Bernstein 2001]RAHM, E.; BERNSTEIN, P. A. A survey of approaches to automatic schema matching. *the VLDB Journal*, Springer, v. 10, n. 4, p. 334–350, 2001.

[Rainaud et al. 2005]RAINAUD, J.-F. et al. Innovative knowledge-driven approach for shared earth model building. In: SOCIETY OF PETROLEUM ENGINEERS. *SPE Europec/EAGE Annual Conference*. Madrid, Spain, 2005.

[Raskin and Pan 2005]RASKIN, R. G.; PAN, M. J. Knowledge representation in the semantic web for earth and environmental knowledge representation in the semantic web for earth and environmental terminology (sweet). *Computers & Geosciences*, 2005.

[Richard 2006]RICHARD, S. M. Geoscience concept models. *Geological Society of America Special Papers*, v. 397, p. 81–107, 2006.

[Sandsmark and Mehta 2004]SANDSMARK, N.; MEHTA, S. Integrated information platform for reservoir and subsea production systems. In: PDT. *Proceedings of the 13th Product Data Technology Europe Symposium (PDT 2004), October, Stockholm*. Stockholm, 2004.

[Sheth and Larson 1990]SHETH, A. P.; LARSON, J. A. Federated database systems for managing distributed, heterogeneous, and autonomous databases. *ACM Computing Surveys (CSUR)*, v. 22, n. 3, p. 183–236, 1990.

[Sinha 2006]SINHA, A. K. *Geoinformatics: data to knowledge*. [S.l.]: Geological Society of Amer, 2006.

[Soma et al. 2008]SOMA, R. et al. Semantic web technologies for smart oil field applications. In: *Intelligent Energy Conference and Exhibition*. [S.l.: s.n.], 2008.

[Souza 2006]SOUZA, R. G.-S. *Agent-oriented constructivist knowledge management*. Thesis (PhD) — Centre for Telematics and Information Technology University of Twente, 2006.

[Studer, Benjamins and Fensel 1998]STUDER, R.; BENJAMINS, V.; FENSEL, D. Knowledge engineering: Principles and methods. *Data & knowledge engineering*, v. 25, n. 1 - 2, p. 161 – 197, March 1998.

[Uschold and Gruninger 1996]USCHOLD, M.; GRUNINGER, M. Ontologies: Principles, methods and applications. *The knowledge engineering review*, v. 11, n. 02, p. 93–136, 1996.

[Visser et al. 1997]VISSER, P. R. et al. An analysis of ontology mismatches; heterogeneity versus interoperability. In: AAAI. *AAAI 1997 Spring Symposium on Ontological Engineering, Stanford CA., USA*. Stanford CA., USA, 1997. p. 164–72.

[Wache et al. 2001]WACHE, H. et al. Ontology-based integration of information-a survey of existing approaches. In: CITESEER. *IJCAI-01 workshop: ontologies and information sharing*. [S.l.], 2001. v. 2001, p. 108–117.

[Werlang et al. 2014]WERLANG, R. et al. Ontological foundations for petroleum application modeling. In: *18th International Conference on Petroleum Data, Integration and Data Management*. [S.l.: s.n.], 2014.

[Wiederhold 1992]WIEDERHOLD, G. Mediators in the architecture of future information systems. *Computer*, v. 25, n. 3, p. 38–49, 1992.

[Wiederhold 1994]WIEDERHOLD, G. An algebra for ontology composition. In: CITESEER. *Proceedings of 1994 Monterey Workshop on Formal Methods*. Monterey, CA: U.S. Naval Postgraduate School, 1994. v. 56, p. 61.