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

LUAN FONSECA GARCIA

GeoCore Ontology: A Core Ontology for Geological Knowledge Description

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

Advisor: Prof. Dr. Mara Abel

Porto Alegre
July 2021

CIP — CATALOGING-IN-PUBLICATION

Garcia, Luan Fonseca

GeoCore Ontology: A Core Ontology for Geological Knowledge Description / Luan Fonseca Garcia. – Porto Alegre: PPGC da UFRGS, 2021.

101 f.: il.

Thesis (Ph.D.) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR–RS, 2021. Advisor: Mara Abel.

1. Ontology. 2. Geology Core Geology. 3. Geological Knowledge. 4. GeoCore Ontology. 5. Material Constitution. 6. Core Ontology. 7. Semantic Interoperability. I. Abel, Mara. II. Título.

I dedicate this work to my sons Renan and Pedro, who are so young and yet have taught me so much.

"The means by which knowledge is conveyed
are every bit as important as that knowledge itself."

— Katherine Munn

ACKNOWLEDGMENTS

I'm thankful to the CNPq, the CAPES, and the Petrobras and Geosiris companies for sponsoring this project in different moments. I'm also thankful to the Informatics Institute of UFRGS for providing excellent infrastructure for the development of my work.

I owe much of my academic career to my dear supervisor Mara Abel. This work was only possible thanks to all efforts that she dedicated to helping me. I also like to think that I'm a better person today than I was ten years ago, and undoubtedly she has a great part in that.

I'm thankful to my unofficial supervisor, Michel Perrin, who helped me understand this beautiful and complicated science that is Geology. Like Mara, Michel's advice crossed the professional aspect and helped me greatly in my personal life.

I'm thankful to Jean-François and his wife, dear Marie-Laure, for taking care of me like family during the time I spent in France. I really enjoyed all the moments we had together.

Thanks to all my colleagues during my time in the BDI group at UFRGS. All of you have made this place very pleasant to work in.

A special thanks to Fabrício Rodrigues, Joel Carbonera, and Cauã Antunes. I'm sure that all our onto-philosophical discussions are somehow materialized in this thesis. A special thanks also to Renata Alvarenga, which is crucial in reducing the gap between computer science and geology.

There are not enough words to express how grateful I'm for having the family I have. I'm who I'm today only thanks to my mother Laura and my sister Júlia. I love both of you.

Last but not least, I'm grateful for having my beloved wife, Rosi, at my side all this time. She is my emotional stronghold and the one who gave me the best gifts anyone could give, my sons Renan and Pedro. I love you so much, my dear.

ABSTRACT

This work proposes the GeoCore Ontology, a core ontology for Geology that captures the most common entities mentioned across the whole domain. It is composed of eleven classes representing entities shared in specific geological subdomains. We have developed GeoCore to facilitate the development and integration of domain ontologies and conceptual models that support information systems in the geological domain.
The GeoCore Ontology is the result of an ontological analysis based on a strong assumption of realism. We assume that the entities observed by the geoscientists exist independently of anyone's view or belief.
For developing GeoCore, we have considered a view in which an object and the amount of matter constituting it are distinct entities. We based this view on the philosophical theory of material constitution. This strong assumption under the material constitution view allows us to treat properties related to the object and properties related to the matter separately.
The top-level ontology we have based our work on is the Basic Formal Ontology (BFO). The BFO suits well the geological domain since the authors have developed the top-level ontology for describing scientific domains. Furthermore, it affords a well-defined set of classes for representing material entities, such as those we deal with in the Geology domain.
We validated the GeoCore Ontology by analyzing the competency questions, checking the consistency of our implementation with the support of an automated reasoner, and representing a real-world geological use case.
Our contributions to state-of-the-art are both at the application level and the conceptual level. At the application level, we provide a computational artifact containing a set of classes and axioms. At the conceptual level, we provide the ontological analysis of the geological domain based on solid philosophical assumptions and ontology engineering methodologies and principles. We also propose an ontological pattern for the material constitution in the geological domain based on a philosophical theory of material constitution.

Keywords: Ontology. Geology Core Geology. Geological Knowledge. GeoCore Ontology. Material Constitution. Core Ontology. Semantic Interoperability.

GeoCore Ontology: Uma Ontologia Core para Descrição de Conhecimento Geológico

RESUMO

Este trabalho apresenta a ontologia GeoCore, uma ontologia core para Geologia que captura as entidades referidas mais comuns em todo domínio. Ela é composta por onze classes representando entidades compartilhadas em subdomínios geológicos específicos. Nós desenvolvemos a GeoCore para facilitar o desenvolvimento e integração de ontologias de domínio e modelos conceituais que dão suporte a sistemas de informação no domínio geológico.

A ontologia GeoCore resulta de uma análise ontológica baseada em uma visão filosófica realista. Nós assumimos que as entidades observadas pelos geocientistas existem independentemente das visões ou crenças de qualquer um.

Para desenvolver a GeoCore, nós consideramos também uma visão em que um objeto e a matéria que o constitui são entidades distintas. Nossa visão foi baseada em uma teoria filosófica de constituição material. Esta visão de constituição material nos permite tratar propriedades relacionadas ao objeto e propriedades relacionadas à matéria separadamente.

A ontologia de topo em que nosso trabalho foi baseado é a *Basic Formal Ontology*(BFO). A BFO se adapta bem ao domínio geológico porque seus autores a desenvolveram para descrever domínios científicos. Além disso, ela contém um conjunto de classes bem definidas para representação de entidades materiais, como aquelas que são considerados na Geologia.

Nós validamos a ontologia GeoCore analisando as questões de competência, verificando a consistência de nossa implementação com o suporte de um raciocinador automático, e representando um caso de uso geológico com dados reais.

As contribuições ao estado-da-arte deste trabalho estão tanto no nível de aplicação quanto no nível conceitual. No nível de aplicação, nós desenvolvemos um artefato computacional contendo um conjunto de classes e axiomas. No nível conceitual, nós realizamos uma análise ontológico do domínio geológico baseada em visões filosóficas e princípios e metodologias de engenharia de ontologias. Além disso, propomos um padrão ontológico para a constituição material no domínio geológico com base em uma teoria filosófico de constituição material.

Palavras-chave: Ontologia, Ontologia Core para Geologia, Conhecimento Geológico,

Constituição Material, Ontologia Core, Interoperabilidade Semântica.

LIST OF FIGURES

CONTENTS

1 INTRODUCTION

We developed this work in the scope of the Computer Science area and the Knowledge Engineering discipline. Our goal is to contribute to the semantic interoperability problem in the Geology domain. We propose a core ontology for Geology that represents the general relevant entities in the domain.

The semantic interoperability problem is related to the difficulty in integrating resources developed using different vocabularies and different perspectives on the data (HEFLIN; HENDLER, 2000). In (NEUHAUS et al., 2011), the authors says that data and information are often designed to address local needs and in the context of specific applications, which difficult their reuse for new purposes. Furthermore, different sets of data do not cumulate, and thus possible benefits of data integration are lost. In this context, semantically-interoperable systems must be capable of exchanging data where the precise meaning of the data is readily accessible, and any system can translate the data itself into a form that it understands (HEFLIN; HENDLER, 2000).

In the geological domain, the challenge of integrating data and making systems interoperable is nothing new. During an IT conference in 2004, according to (RAINAUD, 2005), the IT director of a big petroleum company reported that his company spent 59% of its budget on IT applications, where 43% of this amount they have spent on system integration and data interoperability.

Almost a decade later, according to (EDISON; BRANTLEY; EDWARDS, 2011), researchers interviewed more than a hundred high-level executives from the petroleum industry. In general, the executives saw the integration becoming even more critical to connect various disciplines and innovations with people and processes. They also said that integration is crucial to delivering the right information to the right people at the right time for better decision-making.

More recently, in 2019, the Brazilian National Agency of Petroleum, Natural Gas and Biofuels (ANP) published a report containing a Plan for Open Data (GUTMAN; RIBEIRO, 2019). Among others, the objectives of this plan are promoting open access to the agency's public data, increasing its data quality, making data available in open formats, and promoting semantic interoperability. They expect that this leads to an easier and faster access to the data, better information management by public agencies, increased system development by independent companies, and better decision-making by the managers. However, in this document, the agency also presents its biggest challenges in making the data available: 1) the massive amount of data; 2) stored in several different databases; 3) represented in incompatible data formats.

The interoperability problem in the geological is likely far from being solved. On the contrary, the digital transformation in the field - i.e., replacing non-digital or manual activities with digital processes - only increased the magnitude of the problem.

In the last thirty years, many authors have proposed ontologies as tools to deal with the interoperability problem (GRUBER, 1991; CHANDRASEKARAN; JOSEPHSON; BENJAMINS, 1999; NILES; PEASE, 2001; NOY; MCGUINNESS et al., 2001; GÓMES-PÉREZ; FERNÁNDEZ-LÓPEZ; CORCHO, 2004; USCHOLD; GRUNINGER, 2004; GUIZZARDI, 2005; SURE; STAAB; STUDER, 2009; GUARINO, 2009; HERRE, 2010; ARP; SMITH; SPEAR, 2015; KEET, 2018). In the Computer Science context, an ontology is a representational artifact, comprising a taxonomy of terms and the relationships among them, with their definitions both in natural language for human understandability and machine-readable language for machine understandability.

Because ontologies explicitly represent the domain entities, their properties, and their relationships, they are used to span heterogeneously structured databases and multiple systems with comparable semantics (OBRST, 2003). They clarify ontological assumptions and provide a common framework across applications for analyzing what entities their data describes (PARTRIDGE, 2002). They provide a common mode of access to the data and help identify incompatibilities between different bodies of data (ARP; SMITH; SPEAR, 2015).

Ontologies are important in a scientific domain like Geology because they formalize the terms shared by geoscientists to describe their domain, making explicit their intended meaning. Having the precise meanings of domain terms defined reduces the ambiguity and possible misinterpretations of the geoscientists' descriptions. However, the process of developing a domain ontology is expensive. It is resource-consuming, and it may keep a whole team dedicated for years refining and improving the knowledge represented (ABEL et al., 2015). Thus, assuring the reuse of different domain ontologies into a common integrated framework is critical.

The role of a core ontology for Geology like the one we propose in this work is to facilitate the development of new domain ontologies and the integration of existing domain ontologies. The core ontology should define general terms shared across various Geology subdomains, such as Stratigraphy, Petrology, Sedimentology, or Structural Geology, allowing the ontologists to anchor their ontologies on unambiguously defined high-level terms that facilitate the communication between distinct ontologies.

In the remainder of this chapter, we contextualize and motivate our research within the geological domain (section 1.1), introduce the interoperability problem for geological knowledge and motivate our work within the Computer Science domain (section 1.2), present the objectives and assumptions we have in developing a core ontology for Geology (1.3), describe the main contributions of this thesis (1.4), and present the text structure of this thesis (section 1.5).

1.1 Geological Context and Motivation

Geology is a science that studies the material entities that exist in the subsurface of the Earth and eventual aerial exposures of this subsurface and tries to interpret their creation and evolution. Geoscientists are concerned with describing Earth's rocks and objects and the characteristics and processes associated with them (PERRIN; RAINAUD, 2013). They must deal with entities of distinct ontological nature and scale of analysis, ranging from millimeters to kilometers, where direct observation is only possible on a restricted portion of the Earth's surface and on small samples of rock from the subsurface that are expensive to obtain.

In (FÁVERA, 2001), the author draws a parallel of geological interpretation with holograms. He says that we may compare each scale of analysis to a fragment of a hologram. The more fragments we bring together, the sharper the image becomes. In a multidisciplinary analysis such as a geological interpretation, each discipline and scale of analysis focuses only partially on the study subject. The gathering of all disciplines and scales will provide a clearer view of the problem.

According to Della Fávera, the geological scale consists of seven scales of analysis that correspond to the scales of geoscientists' different representations, ranging from a magnitude order of 10^{-6} meters (the scale of a scanning electron microscope) to 10^{7} meters (the scale of basin maps).

In a scenario with this scale variability and with multiple different sorts of professionals involved in producing data for geological interpretation, interoperability is essential for efficient information retrieval and for knowledge discovery in studies and applications using geological data (MA, 2011). The problem is that although geological reality is always the same, independently of who observes it, different professionals may have different backgrounds and schools of thought and do not use the terms that represent the entities in reality in a consistent way (ABEL et al., 2016). This inconsistency is a problem because geoscientists can only use this information if there is a precise way for ensuring its progressive integration with the information already existing and for making it readily available in formats understandable to both computers and human beings.

The difficulty in acquiring the data, the scale variability in the domain, the necessity of gathering information from a variety of sources to comprehend the whole scenario, the inconsistencies and ambiguities in the data description, and the high costs of integrating data and making systems interoperable are aspects that make research in the topic of semantic interoperability in the geological domain crucial. In the next section, we present what is needed to ensure that geological data is properly integrated and the research efforts that we have up to this day to deal with that.

1.2 Semantic Interoperability in the Geological Domain

The first step to integrate and reuse geological data correctly is to ensure that geoscientists share consistent meanings of the terms they are using to talk and represent their subject of analysis, not only in their natural language but also in the systems they use. The terms they use to refer to the domain entities must have explicit definitions that can be unambiguously understandable both by them and by their computer systems. In this way, people can develop models, systems, and data standards that represent entities that are more likely to be correctly understood by the information consumers.

Geoscientists have done extensive work in formalizing the knowledge attached to Earth Sciences in the last decades. We can classify the resulting models into three different categories.

The first corresponds to loosely structured knowledge models that cover the full domain of Earth Sciences, like SWEET (RASKIN; PAN, 2005) or restricted parts of it (SINHA et al., 2007; BOUTEILLER et al., 2019).

The second category corresponds to highly structured knowledge models, developed to fit the expectations of geological map editors. The North America Geological-Map Data Model defined a first conceptual model (COMMITTEE et al., 2004), further refined in (RICHARD, 2006). Based on Richard's model, a consortium involving few major geological surveys under the cap of the International Union of Geological Sciences[1](IUGS) has issued successive versions of the GeoSciML[2] model. Geoscientists consider GeoSciML 4.0 as a de facto standard for exchanging geological map data. Richard's model and the GeoSciML were accordingly regarded as major references sources by the Energistics Consortium[3] for developing the data exchange standards RESQML[4] and WITSML[5], dedicated to earth model building and the study of well engineering respectively.

Whatever their degree of formalization, the models on these two categories describe geological entities according to the geologist's point of view. Thus, they do not systematically refer to definite categories of reality and do not consistently represent entities from different scales.

The third category consists of ontologies that focus on the various subfields of Geology. As we have seen already, ontologies are potentially a good solution for securing semantic interoperability within the domain of Geology. However, today's reality is that most ontologies for this domain are very specialized and hardly rests on the same conceptual basis. For instance, in table 1.1, we classify domain ontologies according to their geological subfield and the top-level framework they use. We have three different top-level ontologies: Basic Formal Ontology (ARP; SMITH; SPEAR, 2015), Unified Foundational Ontology (GUIZZARDI, 2005), and SUMO (NILES; PEASE, 2001); two reference models: GeoSciML and Temporal Hierarchical Ordinal Reference System - THORS (COX; RICHARD, 2005); and ontologies that do not use any top-level framework.

The variability in the top-level framework used by the different domain ontologies is a problem. Ontological choices based on distinct goals and ontologies built by specific communities require significant adaptations that are hard to achieve (KOP, 2011). Groups focused on their specific local needs usually create incompatible domain ontologies, and, in the absence of a common top level, these ontologies are developed in relative isolation from each other, resulting in new information siloes. (ARP; SMITH; SPEAR, 2015).

Another problem that arises is that geological language is not the result of rational construction but uses the natural language that emerged from various mental representations and working habits. Resting on definitions that are often context-dependent, using shortcuts and approximations, such language is often ambiguous and misleading for all those who do not master it. Thus, the different entities considered in Geology and the

[1]http://www.iugs.org/
[2]http://www.geosciml.org/
[3]http://www.energistics.org
[4]https://www.energistics.org/resqml-current-standards/
[5]https://www.energistics.org/portfolio/witsml-data-standards/

Table 1.1: Geological domain ontologies and their top-level framework.

Author	**Top-level Framework**
Sedimentary Geology	
(LORENZATTI et al., 2010)	UFO
(ABEL et al., 2012)	UFO
(CARBONERA; ABEL; SCHERER, 2015)	UFO
(GARCIA et al., 2017)	UFO
(RODRIGUES et al., 2019)	UFO
(CICCONETO, 2021)	BFO
Structural Geology	
(BABAIE et al., 2006)	No explicit top-level
(ZHONG; AYDINA; MCGUINNESS, 2009)	No explicit top-level
(BABAIE; DAVARPANAH, 2018)	No explicit top-level
Geological Time	
(PERRIN et al., 2011)	No explicit top-level
(COX; RICHARD, 2015)	THORS
(RADEMAKER et al., 2019)	SUMO
Geological Visual Information Artifacts	
(ABEL et al., 2019)	BFO
Geological Mapping	
(BOYD, 2016)	No explicit top-level
(MANTOVANI; LOMBARDO; PIANA, 2020)	GeoSciML
Geological Relations	
(BABAIE, 2011)	No explicit top-level
(CICCONETO et al., 2020)	BFO

relationships among these entities are often not consistently represented. What is the relationship between a rock and a geological unit? Are they the same kind of entity? How are grains and rocks related? Is a fault structure an object? These are questions that sometimes might have different answers depending on the context and the scale of analysis.

The result is that although we have more precise geological models (as ontologies), we still face problems when integrating them. We consider that the integration could be easier if the domain ontologies were conformant to an ontology of a higher level of abstraction that provides definitions for common geological terms that can be applied equally shared in different geological scales and subdomains and that provides definitions for the ontological relations that exist among the entities represented by these terms. Unfortunately, we are not aware of any ontology that can fill this gap between different domain ontologies for Geology.

Due to the lack of an ontology that can act as the integrating point of domain ontologies in the geological domain, we propose in this work GeoCore Ontology, a core ontology for Geology. In the next section, we briefly introduce the GeoCore Ontology.

1.3 A Core Ontology for Geology

In the previous section, we identified the need for an ontology of a higher level of abstraction to facilitate the development and integration of domain ontologies for Geology. For that reason, we propose the GeoCore Ontology, a core ontology for Geology. In this section, we present our goals, basic premisses, and other aspects we took into account during the development of GeoCore.

The goal of GeoCore Ontology is to serve as a basis for developing new domain ontologies and integrating existing domain ontologies within the geological domain. To achieve that, we identified and formally defined general terms shared in different geological subfields and scales of analysis and also the relationships among the entities these terms represent.

In this work, we rest on the strong assumption of philosophical realism. We assume that the entities we consider and the properties that characterize them exist independently of anyone's beliefs, linguistic practices, or conceptual views (MILLER, 2016). For instance, some geologists may decide that a particular set of layers constitute a geological unit. In contrast, others may believe that these layers are only a part of some bigger unit. Nevertheless, we assume that this set of layers exists independently of any view that geologists might have. Additionally, some might argue that there are vague objects in Geology, and thus, sometimes, it is not possible to precisely demarcate the boundaries of an object observed in Geology. We assume here a contrary view, named the *de dicto* view (VARZI, 2006), according to which the vagueness lies not in the object but in the term (or predicate) used to refer to this object in reality[6].

We defined the ontology's scope with the support of domain experts. Based on a set of competency questions proposed by them, we identified that the ontology should at least cover general terms shared among the geological disciplines of Sedimentology, Stratigraphy, Petrology, and Structural Geology. According to the experts, these are the major disciplines dealing with the natural geological material entities that geologists can objectively analyze and describe. The entities studied in these disciplines range from the scanning electron microscope scale (the smallest scale) to the basin maps scale (the largest scale of analysis) on Della Fávera's geological scale.

Considering the definitions from the Encyclopedia Britannica, Sedimentology is

[6]The classic example is Mount Everest. In the *de dicto* view, there are many different crisp aggregates of matter referred to by the name Mount Everest, which is considered a vague term. In the *de re* view, Mount Everest itself (not the term) is a vague object with no crisp boundaries.

the discipline concerned with sedimentary rocks' physical and chemical properties and the processes involved in their formation (BRITANNICA, 2018b); Stratigraphy is the scientific discipline concerned with the description of rock successions and their interpretation in terms of a general time scale (BRITANNICA, 2014); Structural Geology is the scientific discipline that deals with the geometric relationship of rocks and geologic features in general (BRITANNICA, 2018c); and Petrology is the scientific study of rocks that deals with their composition, texture, structure, occurrence and distribution, and origin in relation to physicochemical conditions and geologic processes (BRITANNICA, 2018a). It was not our goal to develop a complete ontology covering every domain entity of these different disciplines. Nevertheless, our ontology should provide a set of general classes to serve as a basis for developing domain ontologies that may cover the specific entities of these domains.

When we analyze these different disciplines, undoubtedly, rock is one of the central entities linking them. Our work assumes a view in which we separate the amount of matter (the amount of rock) from the object it constitutes (some portion of rock). We were inspired here by the Constitution View (BAKER, 2007), a metaphysical view of concrete entities in the natural world that accepts the possibility of two different material things existing at the same place simultaneously. In other words, this is the view that assumes that an object is different from the matter that constitutes it, that they are spatially coincident, and that a relation of material constitution holds between them. The Constitution View is helpful because it allows us to treat properties of an amount of rock separately from the properties of the object that it constitutes, each having its defining characteristics. It also help us to understand why geoscientists sometimes observe similar characteristics in distinct objects (when they are constituted by the same amount of rock).

A critical aspect of the development of GeoCore is the top-level ontology we used to develop it. In our work, we choose the Basic Formal Ontology (BFO) (ARP; SMITH; SPEAR, 2015). BFO suites our goals because it is an ontology developed for supporting data integration of scientific domains. It has a realistic approach - i.e., the terms in the ontology should refer to real entities of the world -, which is aligned with our basic philosophical assumption of realism. It is also an ontology that fits the descriptions of geological knowledge adequately because it has a well-defined set of classes, relations, and axioms regarding material entities and their location in the space and their relation to processes they participate. Furthermore, it is one of the leading top-level ontologies used in scientific and other contexts (SMITH, 2012), it part of ISO/IEC Standard 21838 and has extensive documentation and an active user community.

For dealing with the boundaries and spatial distribution aspects of geological entities, we combined the classes and relations defined by BFO and the set of spatial relations defined by Cicconeto (CICCONETO et al., 2020). Cicconeto's relations are relevant for us because they were defined specifically for Geology and were based on the BFO ontology as well.

In the next section, we present what we believe to be the contributions of this work.

1.4 Contributions

The major contribution of this thesis is the GeoCore Ontology, a core ontology that defines a set of classes representing the most general entities shared in the Geology domain and the relationships between them. The ontology may support the development and integration of domain ontologies in the geological domain. It may also serve as a framework to analyze and align legacy models for data integration.

The ontological analysis over the geological domain is also a contribution of this work. Furthermore, we consider that the approach we applied during the analysis can bring helpful lessons for those who wish to replicate the ontology development in other domains.

Considering the Constitution View within the Geology field and the ontological pattern for material constitution in the geological domain are also novel contributions.

1.5 Text structure

This thesis is structured in the following chapters:

1. In chapter 2, we define the term ontology within Computer Science and present the principles and methodologies used in this work.

2. In chapter 3, we present the Basic Formal Ontology, the top-level ontology that we based our core ontology.

3. In chapter 4, we present Della Fávera's geological scale and Cicconeto's set of spatial relations, two geological framework that we based our work on.

4. In chapter 5, we review the state of the art in the philosophical literature of material constitution and present the philosophical view which we follow.

5. In chapter 6, we present the ontology requirements specification document that guided our development, the ontological analysis that we performed to decide the relevant entities that we should consider in our ontology, and an ontological pattern for material constitution in Geology.

6. In chapter 7, we present all the natural language and formal definitions of the GeoCore Ontology, a core ontology for Geology.

7. In chapter 8, we validate the GeoCore Ontology considering the competency questions, our ontology implementation consistency, and a real-world use case.

8. In chapter 9, we present our conclusions and the future works.

2 ONTOLOGY AND COMPUTER SCIENCE

Ontology is a term that comes from Philosophy. It is the discipline that studies what are the things that exist. Lynne Baker once wrote that an ontology is an inventory of what exists in reality (BAKER, 2007). Although related, the term ontology in Computer Science has a different meaning. Definitions of ontologies in Computer Science vary depending on whether the author of the definition believes that ontologies should represent the conceptualizations of a domain (a world-view or mental abstraction of someone over the domain) or the real domain entities per se (SMITH, 2004).

In the remainder of this chapter, we present the main differences in ontology definitions and the specific types of ontologies within the Computer Sciences in section 2.1. After that, in section 2.2, we introduce some concepts and methodologies of Ontology Engineering we used in this work.

2.1 Defining ontology

One of the most cited definitions for ontology within Computer Sciences is the definition given by Studer in (STUDER; BENJAMINS; FENSEL, 1998). Studer combines the definitions of (GRUBER, 1993) and (BORST, 1997) to create the following definition:

An ontology is a formal, explicit specification of a shared conceptualization.

According to Studer, *conceptualization* refers to an abstract model of some phenomenon in the world that identifies the concepts relevant to (for example) the scientific treatment of that phenomenon. *Explicit* is related to the fact that the type of concepts used and the constraints on their use are explicitly defined. *Formal* means that the ontology should be machine-readable. Finally, *shared* is related to the fact that an ontology captures consensual knowledge of a group, not what is private to some individual.

The problem with Studer's definition lies in the use of the term conceptualization. In his view, the role of an ontology is to represent the concepts existent in our minds that might refer to entities in the real world. Although it is true that, in many contexts, ontologists use concepts as tools to gain cognitive access to the corresponding entities in reality, the role of an ontology is to represent reality, not what people's minds think of it (SMITH, 2004).

For that reason, we consider that a better definition the one proposed in (ARP;

SMITH; SPEAR, 2015):

An ontology is a representational artifact, comprising a taxonomy as proper part, whose representations are intended to designate some combination of universals, defined classes, and certain relations between them.

This definition contains a set of terms that require definitions themselves. An *ontology* is a *representational artifact*, i.e., something that has been deliberately produced or constructed by human beings to represent other entities. An *entity* is anything that exists, including objects, processes, and qualities. *Taxonomies* are hierarchies consisting of terms denoting types (or universals or classes) linked by subtype relations. *Terms* in an ontology are the linguistic expressions used by human experts in the corresponding discipline to represent the world. Therefore, IDs like those used in programming languages or the letter 'L' are not considered terms in this context.

Now, defining what is a *universal* requires some context. The authors of the ontological realist view hold that the question "what is that is general in reality?" is roughly the question of what it is that makes scientific generalization and law-like statements true. In their view, universals are entities in reality responsible for the structure, order, and regularity (the similarities) that are to be found there. It is what all members of a natural class or natural kind such as heart, rock, or cell, have in common. Thus, universals are repeatable, and they can be instantiated by more than one object and at more than one time. Throughout the rest of this text, we use the terms universal and type as synonyms.

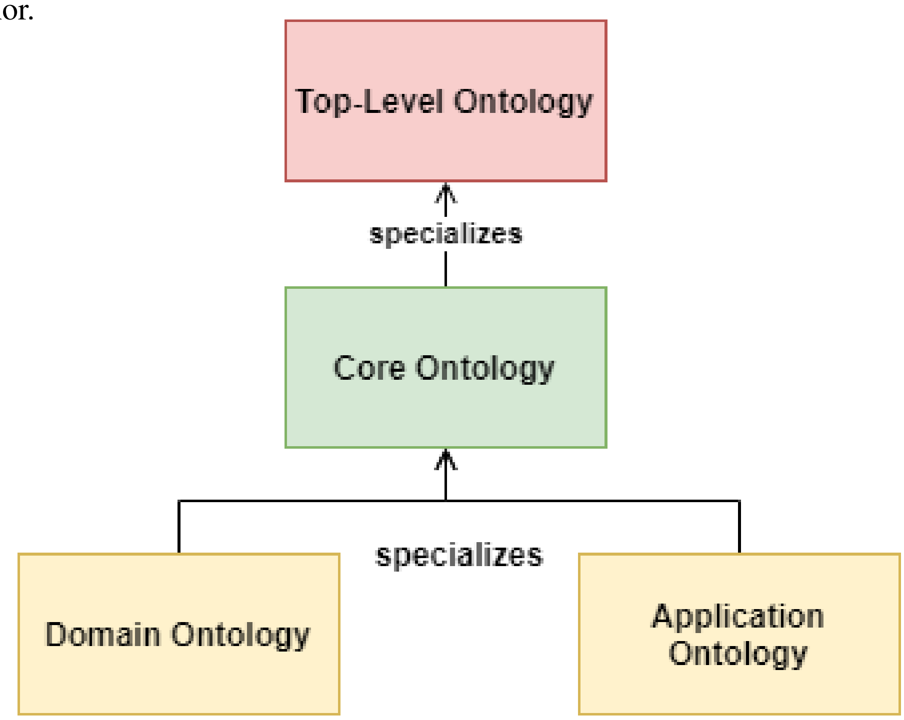
On the other hand, particulars, such as a specific amount of rock, my dog, or the city of Porto Alegre, are nonrepeatable. I.e., particulars can exist only in one place at any given time, and they are instances of universals. They cannot have instances themselves.

A *class* is a maximal collection of particulars falling under a given general term. When the general term denotes a universal, the class's extension is the same extension of the universal. In this case, we might say that this is a "natural class". However, sometimes scientists use general terms to refer to particulars in reality that don't correspond to a single universal, e.g., the smokers who live in Brazil, the people with blue eyes. In these cases, we say that the class is a *defined class*.

The literature offers different classifications of types of ontologies. One that is well accepted is the classification according to the level of generality of the ontology proposed in (PRESTES et al., 2013). In this classification, there are top-level ontologies, core ontologies, domain ontologies, and application ontologies (figure 2.1).

Top-level ontologies, or yet upper-level or foundational ontologies, define com-

Figure 2.1: Ontology classification schema according to their level of generality. Source: the author.



mon general entities that are domain-independent, such as space, time, object, process, and relationships of these entities, such as parthood and subsumption. Well-known top-ontologies in literature are DOLCE (GANGEMI et al., 2002), SUMO (NILES; PEASE, 2001), GFO (HERRE, 2010), UFO (GUIZZARDI, 2005), BORO (PARTRIDGE; STEFANOVA, 2003), and BFO (ARP; SMITH; SPEAR, 2015). Top-level ontologies are important because they define basic terms and relationships, allowing the ontologist to define the terms of his domain.

A *core ontology* defines the central terms from a specific field (Geology, Medicine, Robotics), usually for linking the general terms of a top-level ontology to the more domain-specific concepts from a sub-field (Stratigraphy, Dermatology, Robot Surgery). They are located in a layer between top-level ontologies and domain ontologies (OBERLE, 2006). The number of classes that a core ontology contains isn't necessarily high. Still, in conjunction with the classes defined in the top-level ontology that it specializes, the core provides a framework for developing and integrating different domain ontologies. In this sense, in the same way that top-level ontologies serve as a basis for developing core ontologies, core ontologies serve as a basis for developing domain ontologies (SCHERP et al., 2011).

Domain and *application* ontologies represent the vocabulary related to a specific domain (like Structural Geology or Petrography) or a specific application. They are often developed based on a top-level ontology to have some root for defining their specific terms.

The discipline investigating the principles, methods, and tools for initiating, developing, and maintaining ontologies is *Ontology Engineering* (SURE; STAAB; STUDER, 2009). In the next section, we present the methodology, frameworks, and principles we followed to develop the GeoCore Ontology.

2.2 Ontology Engineering

There are many different methodologies for developing ontologies, such as TOVE (GRÜNINGER; FOX, 1995), METHONTOLOGY (FERNÁNDEZ-LÓPEZ; GÓMEZ-PÉREZ; JURISTO, 1997), On-To-Knowledge (SURE; STAAB; STUDER, 2009), and NEON (SUÁREZ-FIGUEROA, 2010). They are usually composed of the same major steps, varying on how they divide these steps or minor aspects such as the specific activities or tools for knowledge acquisition or documentation. These steps are specifying the ontology purpose and goals, knowledge acquisition, ontology implementation, and ontology validation. Although we didn't follow one of these specific methodologies for developing GeoCore, our development process had very similar steps. In the remainder of this section, we present the steps of the methodology we followed to develop our ontology. After that, we present the metaproperties we considered and the principles we based on during its development.

The methodology is composed roughly of five steps:

1. *Creating an ontology requirements specification document (ORSD):*

   This document includes the purpose, scope, and implementation language of the ontology, the target group, and intended uses of the ontology, and the set of requirements that the ontology should fulfill (SUÁREZ-FIGUEROA et al., 2012). It guides the whole development process and helps to validate the ontology.

2. *Domain knowledge acquisition and modeling:*

   In this step, we capture the knowledge from domain experts and collect it from the literature to create the ontology's first conceptual model. The philosophical notions of essence, identity, unity, and dependence and the BFO top-level ontology support the ontological analysis of the domain.

3. *Formalizing the ontology:*

   We create aristotelean natural language definitions and formalize the ontology using a first-order like language and refine relations between the entities.

4. *Implementing the ontology in a computational language:*

   We implement the ontology in a machine-processable language, such as OWL. In this way, it was possible to use automated reasoners for consistency checking of the ontology.

5. *Validating the ontology:*

   We validate the ontology, verifying if the requirements of the ORSD are satisfied, checking if the ontology is consistent, and, if possible, providing a real-world use case of the ontology.

These are the major steps of the methodology we use in our ontology development. Nevertheless, the whole process is iterative, and it is common to navigate back and forth between the steps.

The ontological analysis occurs mostly in steps 2 and 3. With the support of domain experts, we analyze the domain to raise the relevant domain entities. During the analysis, the philosophical notions of essence, identity, unity, and existential dependence help clarify the domain entities' ontological nature and relationships. These notions are what the authors in (GUARINO; WELTY, 2000) call ontological metaproperties. They are formal properties[1] (domain-independent) that impose some constraints on the subsumption relation, helping to build better taxonomies. In the following, we define each one of the metaproperties.

- **Essence:** we say that a property is essential to an entity only if it must hold for every possible instance of this entity. The property *being crystalline* is essential to **Minerals** since all instances of minerals are necessarily crystalline. However, *being crystalline* is not essential to **Gemstones** since there are gemstones made of non-crystalline material, such as **Amber**.

[1] Properties should not be misunderstood with attributes of objects, as usually adopted in UML language. Here, the word **property** corresponds to unary predicates in Logic, such as *being a rock*, *being a fault*, or *being a reservoir*. All instances of the property *being a fault* that exists are members of the class **Fault** (and thus instances of the corresponding universal **Fault**), and the extension of the class **Fault** are all instances of *being a fault*.

- **Rigidity:** it is a special kind of essentiality. A rigid property is essential to all its instances. *Mineral* and *Person* are rigid properties because no mineral instance or person instance can cease to be an instance of mineral or person and continue to exist.

- **Anti-rigidity:** a property that is not essential to all its instances. *Gemstone* and *Student* are anti-rigid properties because every instance of gemstone could lose its cultural value as a precious stone and would still exist as a stone, and every instance of Student could end his studies ceasing to be a student and still exist as a Person.

Deciding whether a property that defines an entity is rigid or anti-rigid is fundamental for constructing taxonomies because a class defined by an anti-rigid property cannot subsume classes (i.e., be the superclass of) defined by rigid properties (GUIZZARDI; WAGNER, 2005; ABEL et al., 2016).

- **Identity:** it is related to distinguishing a specific instance of a certain class from other instances using some characteristic property, which is unique for it (the whole instance). The identity criterion is important because it is based on it that geoscientists can identity if different objects are constituted by the same amount of rock or not.

- **Unity:** is related to the problem of distinguishing the parts of an instance from the rest of the world by means of a unifying relation that binds these parts together and under what conditions these parts form a whole. The unity criteria is geoscientists use to understand if an entity is a whole geological unit or not.

- **Existential Dependence:** we say that a property a is existentially dependent on another property b if, necessarily, an instance of b in which a depends on must exist when a exists. The color of the amount of rock can only exist while the amount of rock exists, my height only exists while I exist.

We used the meta properties to clarify the domain entities' ontological nature and support our decision of which entities of the Basic Formal Ontology would subsume our domain entities in the hierarchy.

Naturally, an ontology is more than a taxonomy. Definitions are also one of the most critical ontologies components. They are why an ontology can support consistent use across multiple communities and disciplines and support computational reasoning. Thus, we followed a set of good principles of definitions writing proposed in (ARP; SMITH; SPEAR, 2015). We followed these principles because they are the principles followed during the development of BFO top-ontology, the ontology we based on our work. In the following, we enumerate the most relevant principles that we followed.

1. Provide all non-root terms with definitions.
   Terms that are not at the root of the taxonomy should have a definition in the form of a statement of a set of necessary and sufficient conditions.

2. Use Aristotelian definitions. An Aristotelian definition has the form:
   S =def. a G that Ds,
   where "G" is the immediate parent term of "S" in the ontology for which the definition is being created. "D" tells us what is it about Gs in virtue of which they are Ss.

3. Use essential features in defining terms.
   A term's definition captures what we can think of as the essential features of the entities that are instances of the designated type. The essential features of a thing are those features without which the thing would not be the type of thing that it is.

4. Start with the most general terms in your domain.
   Define the terms in an ontology using a top-down approach. Following the Aristotelian template for definitions from rule 2, we start defining the most general universals first and then working downward through the *is a* hierarchy toward progressively more specific terms. Ideally, these general terms should have as "parents" terms defined in a top-level ontology.

5. Avoid circularity in defining terms.
   A circular definition happens when the term to be defined (or a near-synonym of that term) occurs in the term's definition itself. E.g., rock =def. anything that is made of rock.

6. Use simpler terms than the term you are defining.
   The terms in a definition should be more scientifically, logically, or ontologically basic than the term you define. This ensures the intelligibility of definitions and promotes their utility to human beings, allowing people who are not familiar with the particular technical terms to understand the definition.

7. Do not create terms for universals through logical combinations.
   Boolean combinations such as the negation of a term, or the union or disjointness of two or more terms, should be avoided.

8. Definitions should be unpackable.
   If we define an A as "a B that Cs," then we should be able to replace every occurrence of "an A" in a sentence with "a B that Cs," and the result will have the same meaning. This process of substitution is called "unpacking."

In this chapter, we introduced the definition of ontology and its classifications and the methodology and principles we followed for developing GeoCore Ontology. In the next chapter, we introduce the Basic Formal Ontology, the top-level ontology we used to develop our ontology.

3 BASIC FORMAL ONTOLOGY

In this chapter, we present the Basic Formal Ontology and detail some of its parts that are needed to understand some of the ontological choices during the modeling of GeoCore Ontology.

BFO is a small top-level ontology developed for supporting data integration of scientific research (ARP; SMITH; SPEAR, 2015). According to the authors, the goal of an ontology focused on describing a scientific domain like BFO is *"to encapsulate the knowledge of the world that is associated with the general terms used by scientists"* in this domain. Recently, BFO has been under the process of becoming an ISO Standard[1]. BFO provides a common top-level structure that allows integrating the information compiled by various domain ontologies in a common framework for categorization and reasoning. Using BFO, we can use the most generic entities defined by it (material and immaterial entities, boundaries, processes, qualities, spatial regions, etc.) as the basis for defining our entities. It also defines relations, such as the ones from parthood theory, that we can use to define the relationships among the entities we need to represent.

Figure 3.1 presents a view of the main categories of BFO. In the following, we focus on the entities highlighted in this figure because they are the ones that we used to develop the GeoCore Ontology. However, this does not mean that when someone uses GeoCore Ontology as basis to develop a domain ontology that he has to be restricted only to these categories specialized by the core ontology.

The BFO entities belong to two broad categories: *continuants*, which persist through time, and *occurrents*, that occur in time. Both are distinct and complementary views of reality. Examples of *continuants* are a **person**, a **person's height**, a **dog**, a **dog's color**. A person is wholly present as long as it exists. It may have grown, lost its hair, or traveled world wide, but it keeps its identity unless it disappears. *Occurrents* can be processes such as a **handshake**, a **surgery**, or a time interval, such as the **year 2020**.

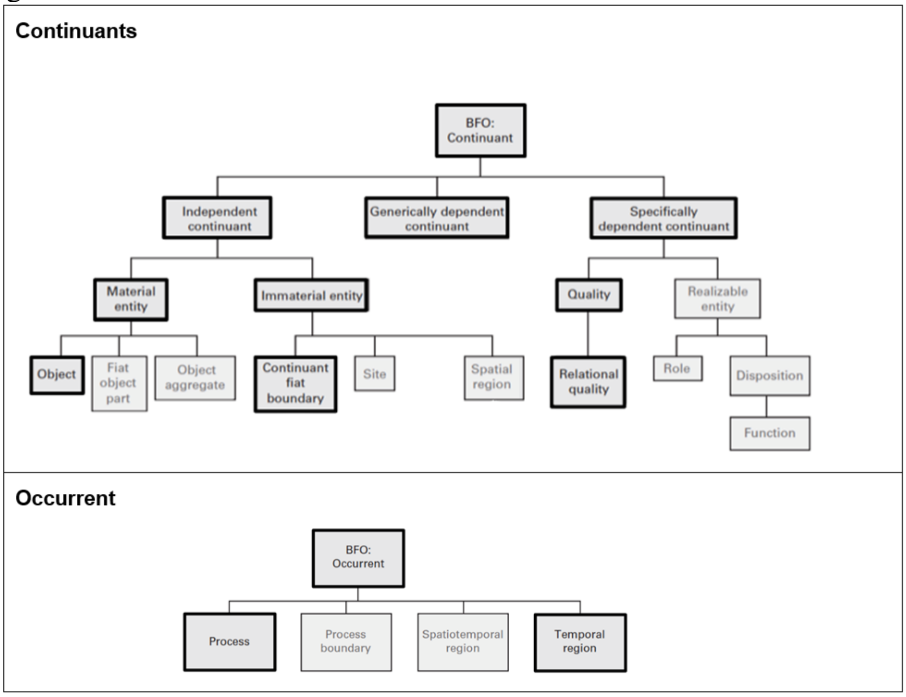
BFO defines different but similar relations of parthood for continuants and occurrents. The relation *continuant part of* holds between *continuants*, and the relation *occurrent part of* holds between *occurrents*.

Continuants are divided in *independent continuants*, *generically dependent continuants* and *specifically dependent continuants*.

Independent continuants are such that their identity and existence can persist through

[1] https://www.iso.org/standard/74572.html

Figure 3.1: The BFO main entities utilized in this work. Source: the author.



the gain and loss of parts. They are also said to be bearers of qualities, i.e., if a continuant entity *a* is the bearer of a quality *b*, then we say that *b* *inheres in* *a*. The **height of a person** is *inherent* to the **person**, i.e., it will not exist unless that person exists. Independent continuants can be of two kinds: material entities and immaterial entities.

Material entities are independent continuants that have some portion of matter as a part. Specific types of material entities are objects and objects aggregates.

Objects are material entities that hold some unifying relation between their parts and are maximally self-connected. Examples of objects are a **car**, a **laptop**, a **building**.

An *object aggregate* is a material entity consisting exactly of a plurality of objects as **member parts** which together form a unit. Exactly means that aggregates have no other parts than their members. It is also possible that the aggregate at a certain time consists of exactly one object, but it must at some time have a plurality of member parts. Examples of aggregates are a **symphony orchestra**, a **collection of sand grains**, a **fire brigade**.

Immaterial entities are those entities that do not have material entities as parts. Specific types are continuant fiat boundaries, sites, and spatial regions.

Continuant fiat boundaries are immaterial entities of zero (*fiat point*), one (*fiat line*), or two dimensions (*fiat surface*), whose location is determined in relation to some material entity, and there is no time in which they have a spatial region as a part. Examples are the **geographic North Pole** (fiat point), the **Equator** (fiat line), the **surface of the**

Earth (fiat surface).

Sites are three-dimensional immaterial entities whose boundaries either partially or wholly coincide with the boundaries of one or more material entities or that have locations determined in relation to some material entity. Examples are a hole in a **portion of cheese**, a pore in a **portion of rock**.

A *spatial region* is a continuant entity of zero, one, two, or three dimensions that is a continuant part of the spatial projection of a portion of spacetime at a given time. Spatial regions are at rest by definition. Examples are the **spatial region occupied at some time instant by the Nort Pole** (0d), an **edge of a cube-shaped portion of space** (1d), the **surface of a sphere-shaped part of space** (2d), a **cube-shaped region of space** (3d).

BFO defines relations to handle spatial localization. Every *independent continuant* that is not a *spatial region* *occupies* some *spatial region* and can be *located in* or be the *location of* another *independent continuant* that is not a *spatial region*. The **BDI lab** is *located in* the **second floor of the 43425 building** that *occupies* the **UFRGS' Informatics Institute's** spatial region.

There are two groups of dependent continuants entities in BFO: specifically-dependent and generically-dependent entities.

A *specifically dependent continuant* *inheres in* an *independent continuant* that is not a *spatial region*, and an *independent continuant* is the *bearer of* a *specifically dependent continuant*. A specifically dependent continuant inheres in an independent continuant that is not a spatial region, and an independent continuant is the bearer of a specifically dependent continuant. A specifically dependent continuant has an existential dependence on a specific instance of the independent continuant on which it inheres. Thus, if the specific instance of independent continuant ceases to exist, then the instance of specifically dependent continuant also ceases to exist.

A *quality* is a *specifically dependent continuant* that is always fully exhibited, manifested, or realized in its bearer, without the need for any further process to occur. Examples are the **dog's color** or the **person's height**.

A *relational quality* is a *quality* that inheres in two or more independent continuants. Examples of relational qualities are **a marriage bond**, **a social contract between one person and another**.

On the other hand, a *generically dependent continuant* depends on some instance of *independent continuant* to exists, but not a specific one. *Generically dependent contin-*

uants exist in virtue of the fact that there is at least one of what may be multiple copies. They are the *content* or the *pattern* that the multiple copies share. Examples are the **content** shared by a string of dots and dashes written on a page and the transmitted Morse code signal, the **Coca-cola logo**, and the **chess pattern**.

To say that a generically dependent continuant *b* *g-depends on* an independent continuant *c* means that there is a *specifically dependent continuant* that *inheres in* *c* that *concretizes* b. An *independent continuant* is the *bearer of* a *generically-dependent continuant*.

Occurrents represent entities that occur, happen, unfold or develop in time. These entities are occurrences of change, i.e., they are the present participles' ontological counterparts (running, walking, cutting, etc.). In this thesis, we focus on two specific types of occurrents: process and temporal region.

A *process* is an occurrent that exists in time by occurring or happening, has some *temporal proper part*, and has some *material entity* as *participant*. A *process boundary* is a *temporal part* of a *process* and has no *proper temporal parts*.

Processes or *process boundaries* *occur in* some *material* or *immaterial entity*, and a *material* or an *immaterial entity* *environs* a *process* or *process boundary*. Examples of *processes* are a **handshake**, and the **life of an organism**. Examples of a *process boundary* is the **boundary between the years 2020 and 2021**.

A *temporal region* is an occurrent over which processes can unfold. Time intervals are *one-dimensional temporal regions* while time instants are *zero-dimensional temporal regions*. An example of a *one-dimensional temporal region* is the **Mesozoic Era**, while **today at 14:00** is an example of a temporal instant.

4 GEOLOGICAL SCALES AND SPATIAL RELATIONS

Geology is concerned with describing the Earth's natural surface and its subsurface material objects, the amounts of rock that constitute them, and their history. The interpretation of their history is supported mainly by two different types of evidence: 1) the amounts of rock intrinsic properties; 2) the spatial distribution of objects constituted by these amounts of rock.

Geoscientists may investigate entities ranging from the size of a mineral grain to the size of a continental tectonic plate. Part of the semantic interoperability problem in the geological domain arises from the fact that software applications consolidate models that do not clearly distinguish the relations among entities in different scales. Thus, understanding the different scales of analysis is crucial for adequately relating these different entities. In section 4.1, we present Della Fávera's list of geological scales of analysis (FÁVERA, 2001). Della Fávera's scales are relevant because he defines the different scales of analysis using a discrete numerical scale corresponding to the scales of geoscientists' different representations.

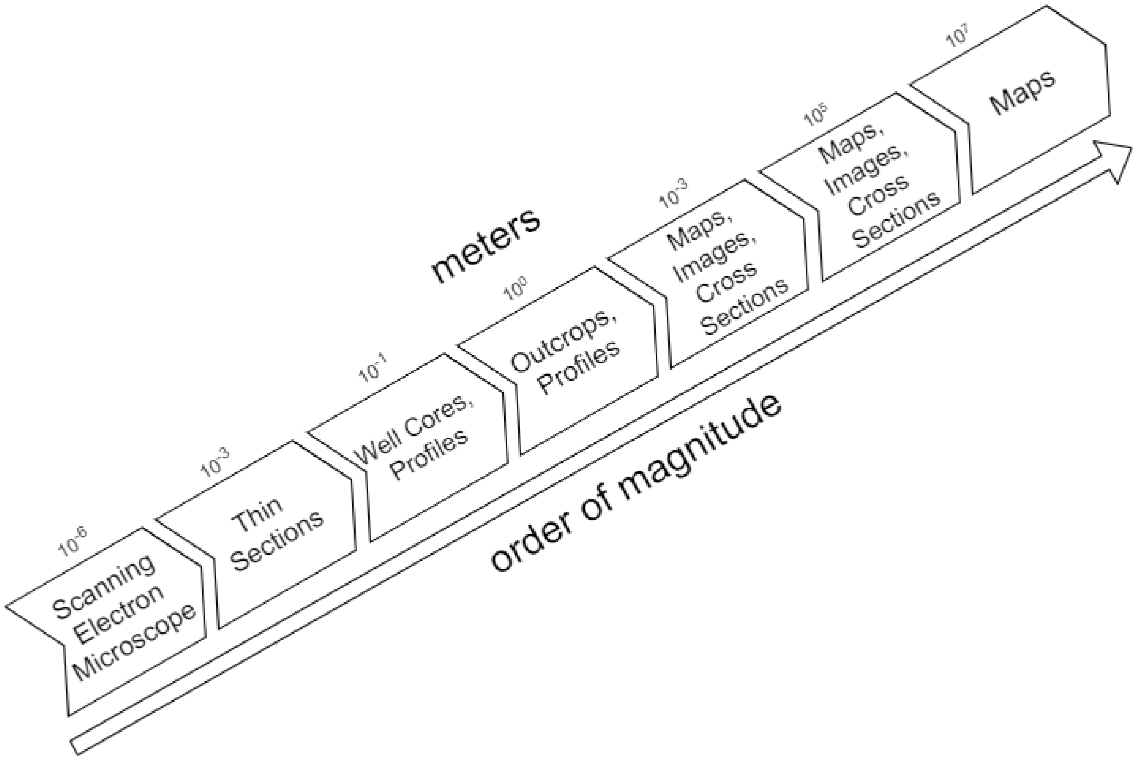
A computable model for Geology should represent how geological objects are spatially related to others since architectural arrangements and the stacking of rock layers support the geoscientist interpretation of the geological processes that occurred in the past. For that reason, we consider in our work Cicconeto's spatial relations for Geology (CICCONETO et al., 2020). We present this set of relations in section 4.2. Cicconeto's relations are relevant for our work because he defined his relations specifically for representing the spatial distribution of geological objects. Furthermore, he also based this set of relations on the BFO top-level ontology, which facilitates their integration with GeoCore Ontology.

4.1 Della Favera's geological scale

In (FÁVERA, 2001), the author divides the geological scale into seven different possible scales of analysis, where each one has its own set of possible representations. The scale ranges from a magnitude order of 10^{-6} meters to 10^{7} meters (figure 4.1). In the following, we explain what major entities of interest in each scale are.

- **10^{-6} scale:** This is the scale of the Electron Microscopy. The geoscientists analyze

Figure 4.1: Della Favera geological scale. Adapted from (FÁVERA, 2001).



various signals over rock samples to study rocks, minerals, crystals, and grains.

- **10^{-3} scale:** This is the scale of Thin Sections. They are a laboratory preparation of a thin sliver of rock sample for use with a polarizing petrographic microscope. They are used to investigate the optical properties of the minerals of a rock. It is also possible to investigate the pores, grains, and crystals present in the rock.

- **10^{-1} scale:** This is the scale of Well Cores and Wireline Well Logs. Well cores are cylindrical samples of rock extracted from boreholes on the Earth's surface, and Wireline Well Logs are graphical records of the measured or computed physical characteristics of the rock section found in a well, plotted as a continuous depth function. This is the first scale where direct observation is possible. The geoscientist investigates characteristics of the rock, its structures, and possible fossils present in the rock sample.

- **10^{-0} scale:** This is the scale of Outcrops and Lithological Profiles. On this scale, the geoscientists investigate stacks of different rock layers and their geological structure to interpret the processes that generated and operated over a specific geological unit.

- **10^{3}, 10^{5}, 10^{7} scales:** These are the scales of Maps, Cross-Sections, and Satellite

Images. The geoscientist investigates the stratigraphic relationships of different geological units and tries to compose bigger in the context of the big basins of the Earth.

For the remainder of this thesis, we will refer as in the macroscopic scale all the entities that are in the well core scale or higher, while entities in the microscopic scale are in the thin section scale or lower.

4.2 Cicconeto's spatial relations

In (CICCONETO et al., 2020), the authors define a set of spatial relations among geological entities based on the BFO top-level ontology and different theories of topology and mereology (parthood theory). Having this set of relations is vital because geological interpretation depends on the proper representation of geological objects' spatial distribution.

For the authors, a relation is a binary predicate that connects two entities having a specific direction. Thus if a relation *r* connects *x* to *y*, it does not necessarily imply that *r* connects *y* to *x*. Following what BFO defines, a relation can relate a universal to a universal, a particular to a particular, or a particular to a universal. They also use the notions of *transitivity*, *symmetry*, *reflexivity*, and *inversion*, and the relations *located in*, *location of*, and *adjacent to* from BFO, and the notion of *subrelation*.

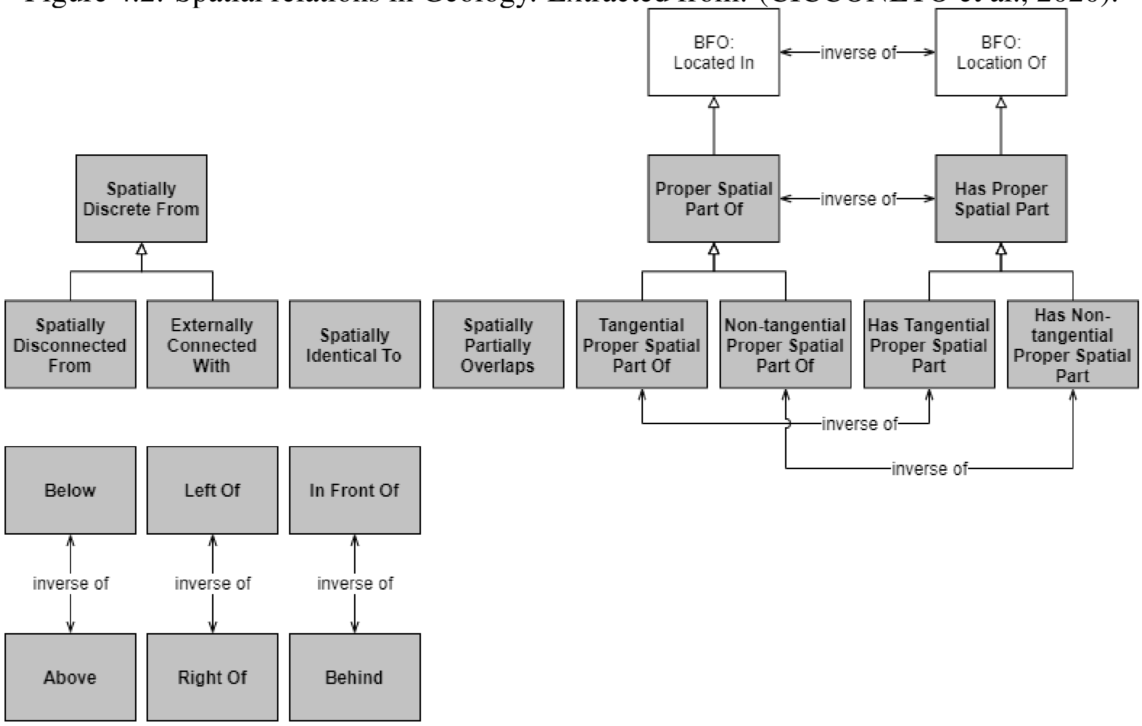
The authors divide their relations into two main categories: 1) mereotopological relations and (2) directional relations. All relations that they define connect particulars that instantiate the BFO type *Independent Continuant*. Figure 4.2 shows all the relations defined by them.

In the following, we present the eleven mereotopological relations that they defined:

1. **Spatially Discrete From:** A symmetric spatial relation between two *independent continuants* in which both do not share the same *spatial region*, either wholly or partially.

2. **Spatially Disconnected From:** A symmetric subrelation of **Spatially Discrete From** between two *independent continuants* whose external boundaries are not adjacent.

Figure 4.2: Spatial relations in Geology. Extracted from: (CICCONETO et al., 2020).



3. **Externally Connected With:** A symmetric subrelation of **Spatially Discrete From** between two *independent continuants* whose external boundaries are adjacent.

4. **Spatially Identical To:** A symmetric spatial relation between two *independent continuants* in which both occupy precisely the same spatial region.

5. **Spatially Partially Overlaps:** A symmetric spatial relation between two *independent continuants* in which both share a part of the spatial regions they occupy.

6. **Proper Spatial Part Of:** A transitive subrelation of *located in* between two *independent continuants*, *x* and *y*, in which the spatial region that *x* occupies is entirely inside the spatial region that *y* occupies. It is the inverse relation of **Has Proper Spatial Part**.

7. **Tangential Proper Spatial Part Of:** A subrelation of **Proper Spatial Part Of** between two *independent continuants* whose external boundaries are adjacent. It is the inverse relation of **Has Tangential Proper Spatial Part**.

8. **Non-Tangential Proper Spatial Part Of:** A subrelation of **Proper Spatial Part Of** between two *independent continuants* whose external boundaries are not adjacent. It is the inverse relation of **Has Non-Tangential Proper Spatial Part**.

9. **Has Proper Spatial Part:** A transitive subrelation of *location of* between two *independent continuants*, *x* and *y*, in which the spatial region that *y* occupies is entirely inside the spatial region that *x* occupies. It is the inverse relation of **Proper Spatial Part Of**.

10. **Has Tangential Proper Spatial Part:** A subrelation of **Has Proper Spatial Part** between two *independent continuants* whose external boundaries are adjacent. It is the inverse relation of **Tangential Proper Spatial Part Of**.

11. **Has Non-Tangential Proper Spatial Part:** A subrelation of **Has Proper Spatial Part** between two *independent continuants* whose external boundaries are not adjacent. It is the inverse relation of **Non-Tangential Proper Spatial Part Of**.

The authors considered the observer's relative location as the frame of reference for defining the following six directional relations:

1. **Below:** A spatial relation between two *Independent Continuants*, *x* and *y*, in which *x* has a location lower than the location of *y* in the vertical axis corresponding to the same frame of reference. It is the inverse relation of **Above**.

2. **Above:** A spatial relation between two *Independent Continuants*, *x* and *y*, in which *x* has a location higher than the location of *y* in the vertical axis corresponding to the same frame of reference. It is the inverse relation of **Below**.

3. **Left Of:** A spatial relation between two *Independent Continuants*, *x* and *y*, in which *x* has a location to the east of the location of *y* in the horizontal axis corresponding to the same frame of reference. It is the inverse relation of **Right Of**.

4. **Right Of:** A spatial relation between two *Independent Continuants*, *x* and *y*, in which *x* has a location to the west of the location of *y* in the horizontal axis corresponding to the same frame of reference. It is the inverse relation of **Left Of**.

5. **In Front Of:** A spatial relation between two *Independent Continuants*, *x* and *y*, in which *x* has a location that makes it nearer than *y* in the longitudinal axis corresponding to the same frame of reference. It is the inverse relation of **Behind**.

6. **Behind:** A spatial relation between two *Independent Continuants*, *x* and *y*, in which *y* has a location that makes it nearer than *x* in the longitudinal axis corresponding to the same frame of reference. It is the inverse relation of **In Front Of**.

In the next chapter, we introduce The Constitution View, a metaphysical view of concrete entities in the natural world that accepts the possibility of existing two different material objects at the same place at the same time.

5 THE MATERIAL CONSTITUTION PROBLEM

In (EVNINE, 2011), constitution is defined as the relation between something and what it is made of. In this sense, we can define material constitution as the relation between some material entity and the physical matter that it is made of. Although this definition looks simple, there is much of a debate about the material constitution relation.

Consider the example of the statue and the clay extracted from (WASSERMAN, 2017). A sculptor buys a lump of clay on Monday and names it 'Lump'. On Tuesday, he sculpts the clay into the form of a statue of the biblical king David and names it 'David'. It is possible to see that Lump differs from David. Lump existed on Monday, David didn't. Lump could survive being squashed, while David couldn't. They also differ in their kinds, since Lump is primarily a lump of clay, while David is primarily a statue. Thus, the following argument is possible:

1. David did not exist on Monday (and it exists on Tuesday).
2. Lump did exist on Monday (and continues to exist on Tuesday).
3. If 1 and 2 are true, then David is not identical to Lump.
4. We conclude then that David is not identical to Lump.

The problem is that this implies that two distinct spatially coincident objects exist, which seems absurd at first glance. This paradox can be broadly analyzed in five perspectives (WASSERMAN, 2017):

1. Accepting that Lump and David are different entities and have different properties (the Constitution View).
2. Denying that 1 is true either because David already existed on Monday, stating that it never existed (the Eliminativist View), or saying that David is a role of an aggregate of molecules.
3. Denying that 2 is true by either denying the existence of Lump (Eliminativist), denying that Lump could survive the shape transformation (the Dominant Kinds View), or saying that David is just a proper temporal part of Lump (the Temporal Parts view).

4. Denying that 3 is true by rejecting the standard formulation of Leibniz's Law[1].

5. Insisting that the underlying issues are in some sense verbal and there is no matter of fact about which premise is false (the Deflationist View).

In this thesis, we assume the first philosophical view, the so-called The Constitution View, because it is the only one that allows us to differentiate properties from the matter (the amount of rock) from properties of the object constituted by it (a geological unit constituted by that amount of rock, for instance).

The Constitution View (BAKER, 2007), is a metaphysical view of concrete entities in the natural world that accepts the possibility of existing two different material objects at the same place at the same time. In other words, this is the view that accepts that some object is different from the matter that constitutes it, that they are spatially coincident and that a relation of material constitution holds between them. This philosophical standpoint brings to light properties that are essential for one entity while are only contingent to another entity. For instance, being shaped like a man is an essential property for David, the statue, while it is only contingent for the lump of clay that constitutes it. Suppose that the statue is melted until it has not a man-shape. Them the David statue would cease to exist, while Lump, the lump of clay, would remain the same.

For Geology, we can demonstrate the importance of separating the amount of rock from the object that it constitutes by considering the following example:

An amount of rock is a consolidated amount of matter created by the same event. Consider that *r1* is a particular instance of an amount of rock of the type sandstone generated by a particular instance of a geological process *e1*. The process *e1* also created the portion of rock *p1*, which *r1* constitutes entirely. In this context, a portion is a maximally connected amount of matter.

Now consider that *r2* is a different particular instance of an amount of rock of the type sandstone, generated by a different particular instance of a geological process *e2*. The process *e2* also created the portion of rock *p2*, which *r2* constitutes entirely. Again, since *p2* is a portion, it is necessarily maximally connected.

Finally, suppose that a process *e3* that occurred after the processes *e1* and *e2* cracked the portion *p1* into two parts: *p1'* and *p1"*.

For geologists, what we have is the following: *p1'* and *p1"* are two distinct portions

[1] the Leibniz' Law, also known as *the Identity of Indiscernibles*, is a principle which says that no two objects have exactly the same properties, i.e., for every property *F*, object *x* has *F* if and only if object *y* has *F*, then *x* is identical to *y* (FORREST, 2016).

of rock because a portion is necessarily maximally connected, and *p1'* and *p1"* are not. *p1'* and *p1"* are constituted by the same amount of rock *r1* because *r1*'s identity is not related to being maximally connected but rather to being created by the same process (process e1 in this case). *r1* and *r2* are instances of the same type of rock *sandstone*. Sandstone is a type of rock with a specific mineralogical composition that may have many different instances in the world. *r1* and *r2* are different amounts of rock because different processes have created them (process *e1* and process *e2*).

This example shows how the Constitution View helps us identify two distinct types of entities, the amount and the portion of rock, and the relation between them - the material constitution.

It is important to note that even within the Constitution View there are distinct approaches. (EVNINE, 2011) separates three different approaches according to the relationship between constitution and composition, where composition is the relation holding between something and its parts. All these three views of constitution assume that the entities involved in the relation of constitution are three-dimensional. It also implies that constitution is relative to a specific instant of time - what constitutes something may change over time.

The first approach considers that objects are constituted by their parts, so constitution is identified as composition (i.e. a parthood relation). This is the view of (FINE et al., 1982; FINE, 1999; JOHNSTON, 2005; JOHNSTON, 2006; KOSLICKI, 2008). The second approach admits that constitution and composition are distinct relations, but it defines constitution in terms of parthood. Usually, in this view, the constituent and constituted entity share all the same parts, but the constituted is more loosely tied to these parts regarding its identity (ZIMMERMAN, 1995; THOMSON, 1998). The third approach, which is the approach we follow in this work, says that constitution is not composition and if *x* constitutes *y* at time *z*, then *x* is not a part of *y* at time *z*. This approach is defended by (BAKER, 2000; BAKER, 2007) and will be detailed in the following.

According to Baker, the fundamental idea of constitution is the following. When something of a certain primary kind is in certain circumstances, something of a different primary kind (a new thing, with new causal powers) comes into existence. Baker says that everything is of a single primary kind. She says that an object's primary kind answers the question: What is x most fundamentally? In this sense, the primary kind of the Lump would be *being a lump of clay* and, when it is sculpted in a certain way, there are certain circumstances that brings to existence a new thing - David -, with new causal powers and different primary kind, i.e. *being a statue*.

Baker defines 6 conditions for an entity to constitute another at a certain time:

1. The constituent and the constituted entity are from distinct primary kinds.

2. The constituent and the constituted are co-localized in space and nothing can be constituted by two things of the same primary kind at the same time.

3. There must exist a set of favorable circumstances which the constituent (circumstances that are needed to exist the new constituted entity) must meet in order for the constitution relation to exist.

4. Whenever the constituent meets the favorable circumstances, the constitution relation must exist.

5. There must exist a possible situation where the constituent is not constituting anything of the same primary kind as the constituted entity, that is, whenever it is not in favorable circumstances.

6. Constitution only holds between things of the same basic kind of stuff (material things to material things, immaterial to immaterial, etc.).

Furthermore, Baker concludes that the constitution relation is asymmetric, irreflexive and contingent. For instance, there could be an aggregate of grains of quartz and feldspar that didn't meet favorable circumstances because they are far away from each other and thus do not constitute a sandstone rock.

In the next chapter, we present the GeoCore Ontology development.

6 DEVELOPING THE GEOCORE ONTOLOGY

In this chapter, we present the results of the methodology we followed to develop the GeoCore Ontology. Firstly, in section 6.1, we summarize the ontology requirements specification document that guided our development. Then, in section 6.2, we explain the rationale behind our ontological analysis of the domain. We also present the natural language definitions of the terms that we considered in the GeoCore Ontology.

6.1 The Ontology Requirements Specification Document for GeoCore Ontology

Following the methodology presented in section 2.2, we created an ontology requirements specification document (ORSD) for GeoCore Ontology to guide our development. We created this document with the support of domain experts, based on the document proposed on (SUÁREZ-FIGUEROA et al., 2012). In the following sections, we summarize the ORSD we created.

6.1.1 Purpose

The ontology's purpose is to be a core ontology that can serve as a basis for developing and integrating ontologies in the geological domain.

6.1.2 Implementation language

The ontology should be developed using the Basic Formal Ontology as a top-level ontology and should be implemented using the Web Ontology Language (OWL).

6.1.3 Intended end users

The intended end-users are:

- Ontology developers;
- IT professionals;

- Geoscientists.

6.1.4 Intended uses

The intended uses of the ontology are:

- Serve as a basis for developing and integrating ontologies for Geology;
- Serve as a basis for analyzing and aligning legacy ontologies from the domain;
- Support the development of ontology-based applications;
- A high-level reference framework to support the geological interpretation.

6.1.5 Ontology requirements

The requirements are aspects that we have to consider in the development of the ontology. Non-functional requirements refer to general aspects or characteristics that are not related to the content of the ontology, such as the naming conventions or if the ontology should be multilingual or not. On the other hand, functional requirements refer to the particular knowledge that the ontology should represent (SUÁREZ-FIGUEROA et al., 2012).

6.1.5.1 Non-functional requirements

Natural language definitions must have an English version. The terms' definitions must follow good practices of ontological definitions as presented in section 2.2.

6.1.5.2 Functional requirements

We asked the domain experts to provide a set of competency questions they expect the ontology to answer for the functional requirements. For formulating the questions, they had in mind that the ontology should have a high level of abstraction and cover the entities shared in geological subdomains. The experts provided a set of six competency questions that we present below:

1. Does the ontology allow representation of the entities that occur in at least more than one scale of analysis?

2. Does the ontology allow representation of a geological area in such detail to allow the interpretation of the spatial distribution of the objects contained on it?
3. Does the ontology allow representation of a geological area in such detail to allow sedimentological interpretation?
4. Does the ontology allow representation of a geological area in such detail to allow stratigraphic interpretation?
5. Does the ontology allow representation of a geological area in such detail to allow structural interpretation?
6. Does the ontology allow representation of a geological area in such detail to allow petrological interpretation?

Based on this set of competency questions, we defined the ontology scope. The ontology should at least cover general terms shared among the geological disciplines of Sedimentology, Stratigraphy, Petrology, and Structural Geology. We do not choose these disciplines randomly, but rather because they are the major disciplines dealing with the natural geological material entities that geologists can objectively analyze and describe. The entities studied in these disciplines range from the scanning electron microscope scale (the smallest scale) to the basin maps scales (the largest scale of analysis) on Della Fávera's geological scale. In the next section, we present the result of the ontological analysis we did for developing GeoCore Ontology.

6.2 Ontological Analysis of the Geological Domain

With the purpose, the scope, and the intended uses for the ontology defined, we need to make an ontological analysis of the domain that we want to represent. In the following sections, we investigate the relevant entities in the geological domain, analyze their ontological nature, and clarify how they are related.

6.2.1 The case of rocks, objects and material constitution

Analyzing the different disciplines with the domain experts' support, we could identify that geoscientists use the term rock in all the different disciplines and scales, but differently. They analyze the grains and crystals to identify the minerals that make a rock on a microscopic scale. However, they analyze objects made of rock on the macroscopic scale, such as well cores, outcrops, rock layers, lithological unities, and geological unities. Thus, although they consider the microscopic scale that a rock is made of different grains and crystals, they consider that objects are made of homogeneous amounts of rock on the macroscopic scale.

An amount of rock presents roughly uniform properties - such as porosity and density - across its entire spatial extent. However, we can only observe amounts of rock when they constitute some object, such as a rock sample or a lithological unit. Amounts of rock have their identity defined by their composition and the process that generated them, but no unifying criterion for their parts. They are a kind of what we call amounts of matter (GANGEMI et al., 2002). An amount of matter is a material entity that not necessarily holds a unifying relation among its parts, it is not necessarily self-connected, and that is mereologically invariant.

Analogously to the amount of rock, an amount of mineral also presents homogeneous properties, such as chemical composition, melting point, and behavior under pressure, provides identity criteria, but not necessarily holds unity.

On the other hand, a grain, a crystal, a well core, a lithologic unit, or a geologic unit are necessarily maximally-self-connected entities that have a unifying relation among their parts and have their own identity criteria. They are Objects as defined by BFO. Thus, geoscientists are interested in properties such as size, shape, and spatial position.

The correlation between these entities of different scales, however, is usually imprecise. A usual definition for rocks is that they are "made of" grains and crystals, and properties such as grain-size distribution are attributed to the rock. However, these properties contradict the homogeneous nature of rocks. Something else collects the microscopic grains and crystals, allowing the analysis of their properties as a group, but without the homogeneity characterizing the amount of rock. We are dealing with three fundamental kinds of material entities: objects, amounts, and aggregates. The aggregate is the entity that "collects" the microscopic grains and crystals, allowing the analysis of their properties as a group.

In the same way that an object has properties that are not directly derived from the amount of rock that constitutes it, the rock properties do not arise directly from the properties of the aggregate of grains and crystals that constitutes it. Instead, those properties emerge, in a certain sense, from the way the grains and crystals are arranged. The same aggregate, if differently arranged – e.g., due to diverse pressure conditions under the rock formation – would yield a rock with different material properties - e.g., higher pressure would result in 'less permeability' of the amount of rock.

By distinguishing these three main kinds of entities, we can elucidate which properties belong to each type of entity. For example, spatial properties such as position, size, and shape always come from objects. A geoscientist is interested in a lithological unit's boundaries, not of the amount of rock that constitutes it. He is interested in the shape of the grain, not of some mineral amount. Similarly, material properties, like permeability, density, and melting point, are always properties of amounts of matter. For example, an amount of rock has a certain porosity that results from the arrangement of the aggregate of grains and crystals that constitute it. An amount of mineral presents a certain melting point that results from the particular way the aggregate of molecules that constitute it is bounded together.

In these cases, we have properties of the "upper level" entity that do not arise directly from the underlying entity's properties. These new properties suggest that there may be some constitution relations playing a part here. More specifically, there seems to be constitution relations between:

A Amount of rock and object.

B Aggregate of grains and crystals and amount of rock.

C Amount of mineral and crystal.

To verify whether or not this is the case, we can check them against Baker's six conditions presented in chapter 5.

Some of Baker's conditions can be seem immediately to be fulfilled by all cases.

First of all, we are dealing with entities of distinct primary kinds in each of the considered relations, which fulfills condition (1).

It is also clear that, in each of the cases, the related entities are spatially coincident. Additionally, amounts of rock and minerals have their identities determined by that of their parts. If two instances of the same kind of them spatially coincide, they would share all the same parts, and then they would be the same entity. The case of the aggregate of grains and crystals is similar. Thus, it would not be possible to have two spatially coincident amounts (or object aggregates) giving rise to the same constituted entity, which fulfills condition (2).

Moreover, the relations involve just material entities, which arguably corresponds to a single basic type of stuff, fulfilling condition (6).

Now we need to verify conditions (3), (4), and (5).

According to condition (3), the constituent entity must be in some favorable circumstance that gives rise to the constituted entity. Additionally, by condition (4), whenever something of the constituent's type is in such favorable circumstance, it must give rise to the corresponding constituted entity. Finally, by condition (5), there must be some possible situation in which a potential constituent doesn't give rise to the corresponding constituted entity.

For case (A), let us consider a lithological unit constituted by an amount of rock. A lithological unit is a body of rock constituted by a single amount of rock with its lithology sufficiently distinctive from its surroundings. Thus, an amount of rock gives rise to a lithological unity whenever it is in some condition that results in the referred "sufficient distinctiveness". This distinctiveness is provided by some material discontinuity between the rock and its surroundings, such as being delimited by a geological fault or being surrounded by rocks of different types. By the definition of lithological unity, the rock that constitutes it is always in such condition, what fulfills condition (3).

Likewise, whenever an amount of rock is in such circumstances, it becomes sufficiently distinctive and gives rise to some lithological unity, which fulfills the condition (4).

Finally, whatever sufficiently large rock we consider contains smaller, inner amounts of rock of the same type since it is homeomerous. There is no material discontinuity for such inner amounts that can make it distinguishable from the larger rock amount. Thus, in these cases, the inner amounts cannot be said to give rise to any lithological unit, providing a possible case to fulfill the condition (5).

In case (B), for an aggregate of grains and crystals to give rise to a rock, they must present some high degree of consolidation (i.e., they must be tied strongly enough together). Such degree of consolidation is the favorable circumstance required by condition (3). Additionally, this degree of consolidation gives rise to the properties that rocks exhibit (e.g., it fixes the grains in the specific structure that results in the rock's porosity). Therefore, whenever an aggregate of grains reaches such an adequate degree of consolidation, an amount of rock comes into existence - what meets the condition (4).

For condition (5), we can think what is the case of when we already have the collection of grains and crystals gathered at the same place right after the sediment deposition (i.e., when the grains are deposited at some surface), and when the lithification process (i.e., the process that consolidates the sediment into rock) did not yet occur. In this situation, we do not have an amount of rock, but only an aggregate of grains and crystals - which fulfills the condition (5).

Case (C) is analogous to case (A). A crystal is an individuated portion of a mineral. Thus, a crystal arises from an amount of mineral when it presents a material discontinuity with its surroundings (e.g., due to a crack in a larger amount of mineral, the two amounts are no longer tied together by the chemical bonds characteristic of the mineral type). This material discontinuity is the favorable circumstance required by condition (3).

Whenever an amount of mineral is surrounded by a material discontinuity, it becomes an individuated portion of mineral, i.e., a crystal - what meets condition (4). Analogously to the case (A), every sufficiently large mineral amount contains some inner non-individuated amount of the same mineral that lacks the required material discontinuity and doesn't give rise to a crystal - as required by condition (5).

In contrast to from the entirely new properties in cases (A), (B), and (C) - which do not derive from the underlying constituent entities -, no new property arises when grouping objects in an aggregate. An aggregate's properties are simply statistical data over the objects they collect, derived from those objects' properties.

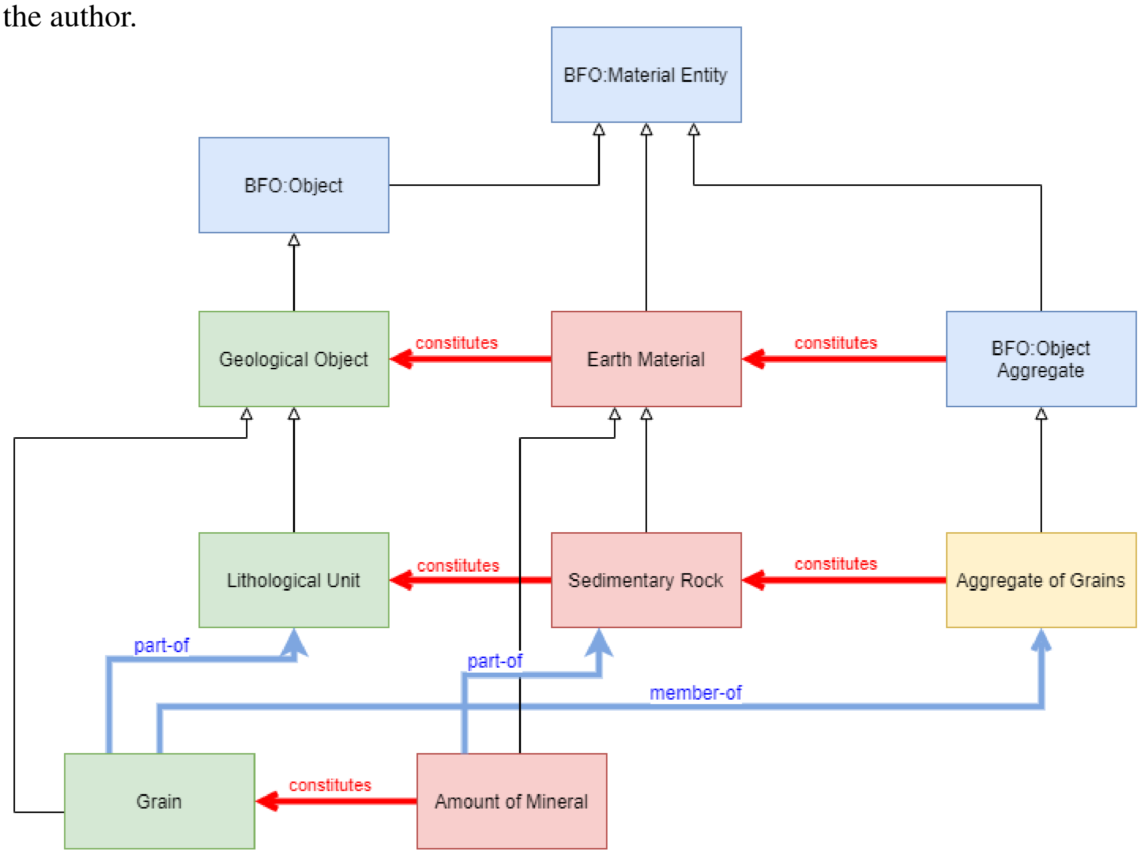
In our example, the mode, the average, and the distribution of the grains' size are properties of the collection derived from each grain's size. There is a clear distinction in the relation between the aggregate and the objects it groups and the relations of constitution holding between objects and amounts or amounts and aggregates.

Additionally, if the relation between the aggregate and each object were a constitution relation, it would contradict our definition of constitution by having more than one constituent for a single constituted entity and having a constituent entity that does not spatially coincide with the constituted entity. The relation of an aggregate and each of its objects is one of membership, a specific type of parthood[1].

The grains and crystals also hold a different relation of parthood with the object of an upper constitutional level, i.e., they are a part of the bigger object, not the amount of rock. They are related with the rock through the constitution but only gathered with other grains and crystals in the aggregate. Therefore, we can think of grains and crystals in some lithological unit, even considering that it is constituted by something homogeneous like an amount of rock.

[1]Here we assume that parthood relations between the grain and crystal and the aggregate, and the grain and crystal and the object are of different nature (GUIZZARDI, 2005). The parthood between the aggre-

Figure 6.1: Ontological pattern for material constitution in the Geology domain. Source: the author.



Given our definition of the domain's entities, their classification into three upper-level types, and the identification of the constitution and parthood relations among them, we summarize the full model in figure 6.1. The pattern: (1) integrates different scales of analysis in the same model (e.g. dealing with amounts of mineral, grains and crystals in the microscopic scale; amounts of rock and objects in the macroscopic scale), (2) supports the distinction between objects and the matter that constitutes them (e.g., lithological units constituted by sedimentary rock and grains constituted by mineral amounts), and (3) clarifies how some entities present different properties in different scales (i.e., by acknowledging the existence of object aggregates as constituents of amounts, we establish a link between discrete particles that we observe in smaller scales, e.g., grains, with the homogeneous amounts of matter we observe in larger scales, e.g., amounts of rocks).

---

gate and the grain is one of member-collection that holds between a singular entity and either a plural or collective entity. The parthood between the grain and the unit is one of component and functional complex, where the part (grain) has some functional role in the whole (object).

6.2.2 Independent Geological Entities

In the initial analysis, we could identify that 'amount of rock' refers to an entity that appears in all the different scales either by 1) constituting some object in the macroscopic scale; 2) being constituted by an aggregate of grains and minerals in the microscopic scale. In both cases, the amount is related to objects in some way.

However, what is it that the objects of the domain share in common? Firstly, they are all Objects as defined in BFO. Thus, they have some unifying relation for their parts, and they are maximally self-connected. They also have their own identity that may vary, depending on their size, shape, and the process that generated them. More importantly, there is always some natural amount of matter constituting them, such as an amount of rock or an amount of mineral. With all that in mind, we can see that they are a specific category of objects. They are what we call Geological Objects.

A first attempt at defining what a Geological Object is:

Geological Object=def a *BFO:Object* that is constituted by some natural amount of matter.

Thus, geological units, rock layers, the planet Earth, grains, and crystals are all objects constituted by natural amounts of matter, and so are all geological objects. But a well core, a thin section, a tree chunk could also be considered geological objects, which is not our intention.

Let's start with the case of well cores and thin sections.

What differentiates a well core or a thin section from a geological unit, a grain, or a crystal? The obvious difference is that the unit, the grain, and the crystal are natural objects generated by some natural process. Conversely, the well core and the thin section are artifacts, objects deliberately produced or constructed by humans beings. This difference leads to our conclusion then that geological objects are only those generated by natural processes.

A second attempt at defining what a Geological Object is below:

Geological Object=def a *BFO:Object* that is constituted by some natural *amount of matter* and that is generated by a natural process.

With this definition, we can successfully exclude artifacts constituted by amounts of matter from the category of geological objects, but the tree case remains open. A tree chunk is an object that is constituted by wood, arguably a natural amount of matter, and thus still fits the above definition. The question here is that geoscientists are not interested in *any* natural objects that exist, only in those constituted by what we call Earth Materials, such as amounts of rock, mineral, petroleum, soil, water. They are not also interested in any natural process, only those macro processes that generate, transform, deform, transport, or destroy other entities relevant in Geology in the course of Geological Time, what we call a Geological Process.

Now, we are in a better position for defining what a Geological Object is:

Geological Object=def. a *BFO:Object* that is <u>*constituted by*</u> *Earth Material* and that is <u>*generated by*</u> a *Geological Process*, where an Earth Material is:

Earth Material=def. is a *BFO:Material Entity* that is a natural amount of matter <u>*generated by*</u> some *Geological Process*.

Earth Materials are natural amounts of matter. They come into existence by nature, without any artificial aid. Since they are amounts, they don't hold unity criteria. Thus, Earth Materials and BFO Objects are disjoint. They are either solid, fluid, or unconsolidated. We usually observe solid earth materials when constituting other objects, such as an amount of rock constituting a geological unity or an amount of sand constituting a dune.

And a Geological Process is:

Geological Process=def. is a physical, or chemical, or biological, naturally occurring *BFO:Process* that <u>*occurs*</u> on the Earth's surface or subsurface and <u>*occupies*</u> some *Geological Time Interval*.

The Geological Time Scale is a system for chronological dating widely used in Geosciences. In the context of BFO, instances of this scale are *Time Intervals*. We can define it as the following:

Geological Time Interval=def. is a *BFO:Temporal Interval* that corresponds to a time interval within the Geologic Time Scale.

The Geological Time Scale is a time scale based on the International Chronostratigraphic Chart, developed and maintained by the International Commission on Stratigraphy, under the International Union of Geological Sciences cap (COHEN et al., 2013).. The scale hierarchy consists of - from bigger unit to smaller - eon, era, period, epoch, and age, starting roughly 4.5 billion years ago.

Unfortunately, we can have the bizarre case with our geological object definition where an object is constituted purely by some fluid like water or petroleum. So, to rule out these cases, we finally define a geological object as the following:

Geological Object=def. is an *BFO:Object* that is <u>*generated by*</u> some *Geological*

Process and has at least one part constituted by some *Earth Material* that is not an *Earth Fluid*.

And we define Earth Fluid:

Earth Fluid=def. is an *Earth Material* that is fluid.

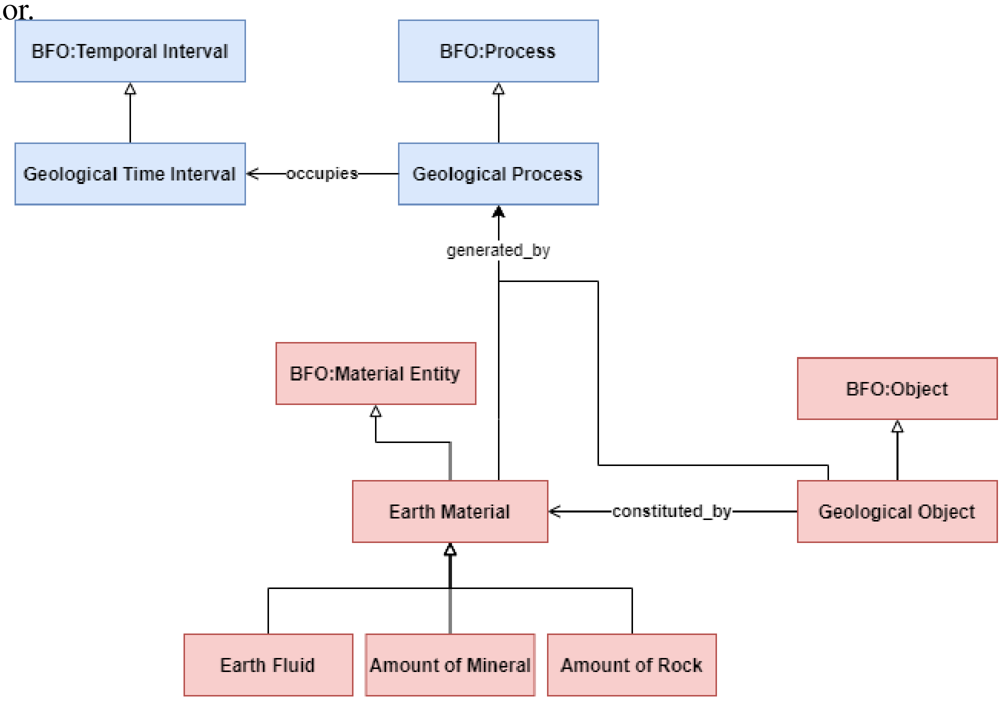
Earth fluids can be water, oil, gas, or a mixture of those fluids.

From the definitions of geological process, earth material, and geological object, it is natural that we conclude that the continuant counterpart of the geological time interval in which a process operated is the geological age of the object or earth material. Thus, we defined geological age as the following:

Geological Age =def. a *BFO:Quality* that *inheres in* a *Geological Object* or *Earth Material* that corresponds to the Geological Time Interval in which the *Geological Process* that *generated* them *occupied*.

In figure 6.2, we have the relations of subsumption, constitution, and generation between the entities we defined in this section. Blue entities are occurrents, while red entities are continuants.

Figure 6.2: GeoCore Ontology overview and subsumption relationships. Source: the author.



The categories of Earth Material and Geological Object englobe all the relevant independent material entities relevant in Geology. Analogously, the category of Geological Process englobes all the natural processes considered in Geology. These are the main categories that ontology developers should use as roots when building ontologies for the geological domain. These are also the key categories that they can use to analyze the terms defined in legacy ontologies to support their integration. However, although geoscientists are primarily interested in the independent material entities, they are so because of these entities' specific characteristics. In the next section, we will analyze the relevant dependent entities of the domain.

6.2.3 Spatial relations of Geological Objects

Geoscientists are not only interested in describing the geological objects they observe and the material that constitutes them. They also want to understand how the objects are spatially distributed to group entities from smaller scales to understand bigger scales' objects and processes. In the previous sections, we already saw that the constitution relation relates objects and amounts, and that the generation relation relates material entities and processes. However, these are relations between either 1) maximally colocalized material entities; or 2) occurrents and material entities. A question that remains is:

How are two distinct objects spatially related?

The first step for a geoscientist to verify how two objects are spatially related is to identify the objects' spatial boundaries in question. Geological objects are maximally self-connected. Thus, we can say that for every instance of a geological object, in any scale, there exists a fiat surface (as defined in BFO) that perfectly bounds the exterior surface of this object. This fiat surface is located on the object and coincides with a physical discontinuity that delimits the object's spatial extension. This fiat surface is what we call a Geological Boundary.

We define a Geological Boundary as the following:

Geological Boundary =def. a *BFO:Fiat Surface* that is <u>located in</u> the external surface of a *Geological Object*.

Once the geoscientist recognizes the objects' boundaries, he can then localize the spatial regions that these objects occupy and check what kind of spatial relation these objects hold. Considering the spatial relations of Cicconeto presented in section 4.2, we have two possible scenarios for spatially relating objects.

1) The two objects are *spatially discrete from* each other, and in this case, they hold no parthood relation; 2) One of the objects is a *proper spatial part* of the other, and thus they hold a relation of proper parthood.

In the first case, if both objects are *spatially disconnected from* each other, they do not have any kind of direct spatial relationship. However, if both objects are *externally* *connected with* each other, their geological boundaries are in physical adjacency, and they are in what we call a *geological contact*.

We define a geological contact as the following:

Geological Contact =def. is a *BFO:Relational Quality* that *inheres in* two distinct *geological objects* that are *externally connected with* each other.

In the second case, when one of the objects is a Proper Spatial Part of the other, we rather say that one is a *continuant proper part of* the other, not that they share a geological contact. However, it is perfectly acceptable that two distinct objects are continuant proper parts of a bigger object but have geological contact between them.

We can observe geological contacts in different scales, such as the contact between two grains that are continuant proper parts of a bigger rock layer. In this case, they are part of the layer because they are *members of* the aggregate that constitutes the amount of rock that constitutes the layer.

Now we can answer the question left over from the start of the section. Two distinct geological objects are spatially related either by one of the three cases below: 1) A geological contact, when their geological boundaries are adjacent; 2) A relation of proper parthood, when the spatial region that one object occupies is entirely inside the spatial region that the other object occupies; 3) They are spatially distant because both objects are *spatially disconnected from* each other.

6.2.4 Geological Structures

Commonly, the objects that the geoscientists observe today are thousands of years old. When they need to describe the processes that generated or transformed a geological object that is so old, they rely on interpretations of the objects' intrinsic characteristics.

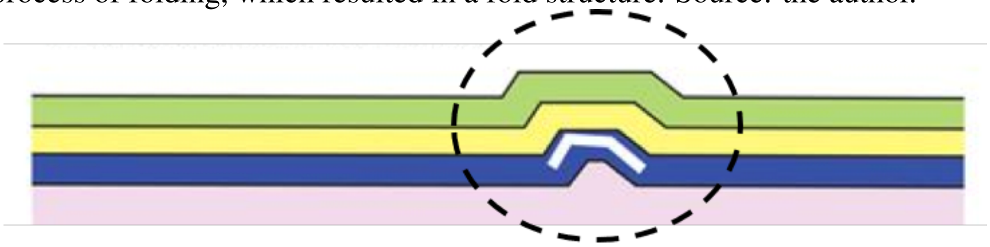
For instance, in a sedimentary environment, due to the law of superposition[2], a geoscientist would normally expects that the rock layers that are part of a bigger geological unit are arranged horizontally, one layer deposited on top of the other. Thus, when he observes an arrangement of the geological unit's layers like in the simplified schema in figure 6.3, he identifies the general pattern of bedding (one bed of rock on top of the other) and concludes that a sedimentation process generated the unit.

[2]The law of superposition says that within a sequence of layers of sedimentary rock, the oldest layer is at the base and that the layers are progressively younger with ascending order in top of each other, in the sequence of their deposition.

Figure 6.3: A set of rock layers that is part of a geological unit, in a bedding pattern. Source: the author.

If the geoscientist observes a case like in figure 6.4, in addition to the bedding pattern, he identifies a pattern of folding in the arrangement of the layers. He then concludes that after the sedimentation process generated the unit, a tectonic process of folding transformed the layers' arrangement.

Figure 6.4: A set of rock layers that is part of a geological unit that has suffered a geological process of folding, which resulted in a fold structure. Source: the author.



Here, we can see that the arrangement of a non-atomic object's internal parts is not completely arbitrary but results from the geological processes that have generated or transformed it. Furthermore, due to the uniformitarian principle[3], we know that different types of geological processes are repeatable along geological time, so we can expect that the patterns resulting from them also repeat in nature.

Geological Structures are these patterns of the internal arrangement of geological objects. Structures are BFO *generically dependent continuants* concretized by some complex quality that inheres in the *geological object* that is its *carrier*.

We can define a geological structure as the following:

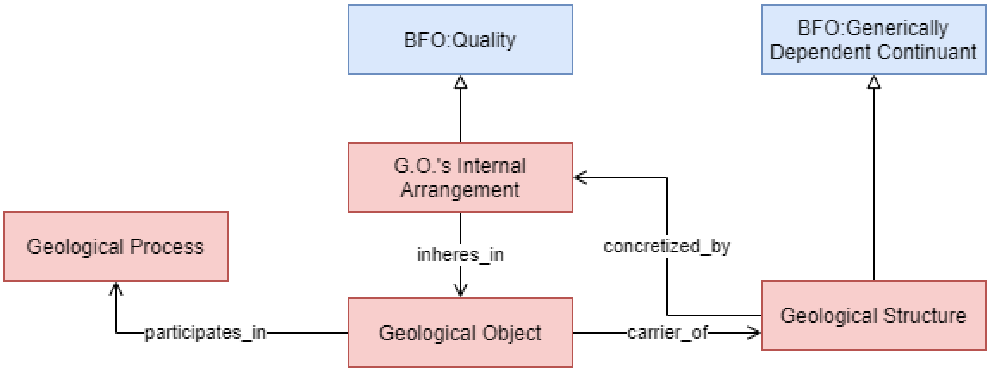
Geological Structure =def is a *BFO:Generically Dependent Continuant* that is the pattern of a non-atomic *Geological Object's* internal arrangement.

A bedding structure is an instance of a Geological Structure that is the pattern of a vertical stack of rock layers. This instance of bedding structure is concretized in the many different instances of geological units that have a complex quality of the arrangement of its parts being in a vertical stack of layers. Figure 6.5 summarizes the relationships of a

[3]Uniformitarianism is the assumption that the same natural laws and processes that operate in our present-day scientific observations have always operated in the universe in the past and apply everywhere in the universe.

geological structure.

Figure 6.5: Geological structures are concretized by some quality inhering in a geological object. Source: the author.



In the next chapter, we present an extensive list of all the terms defined in this chapter and their formalization in equivalent axioms of a fol-like language.

7 THE GEOCORE ONTOLOGY: A CORE ONTOLOGY FOR GEOLOGY

The GeoCore Ontology is the result of a deep ontological analysis over the geological domain. We based our work on philosophical theories to clarify the entities shared across the different scales and subdomains of Geology and the relationships among them. We used the BFO top-level ontology, the geological scale of Della Fávera, and the spatial relations of Cicconeto as the basis for our natural language and formal definitions.

Figure 7.1 shows all the classes we defined in the ontology and the respective BFO class that each of them is specializing. Figure 7.2 shows the core relations that hold among the classes. We shared the OWL implementation of the GeoCore Ontology in a web repository[1].

Figure 7.1: Ontology overview and subsumption relationships. Source: The author.

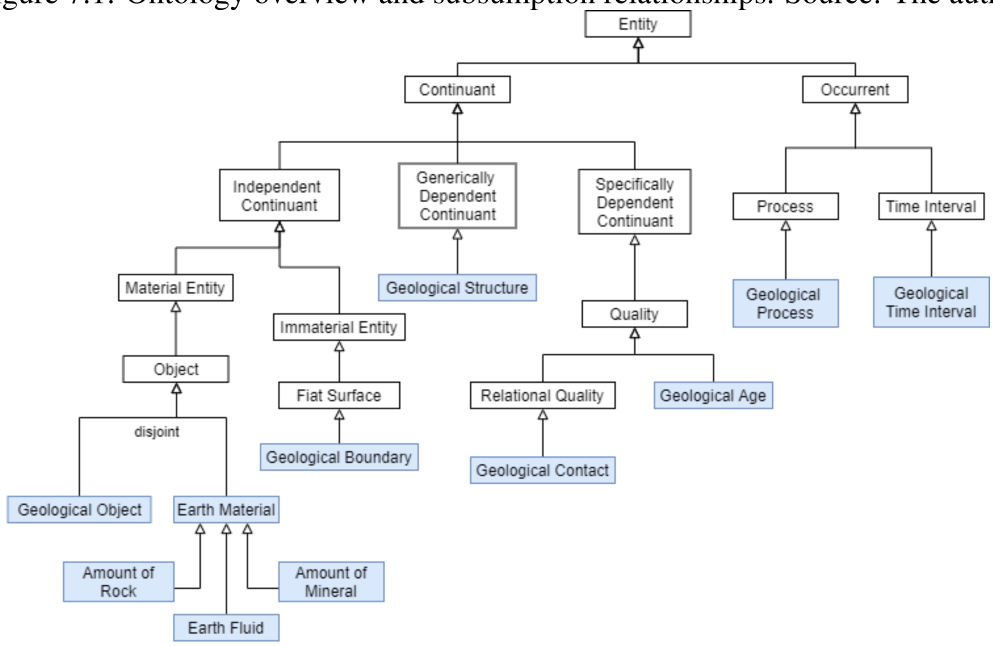
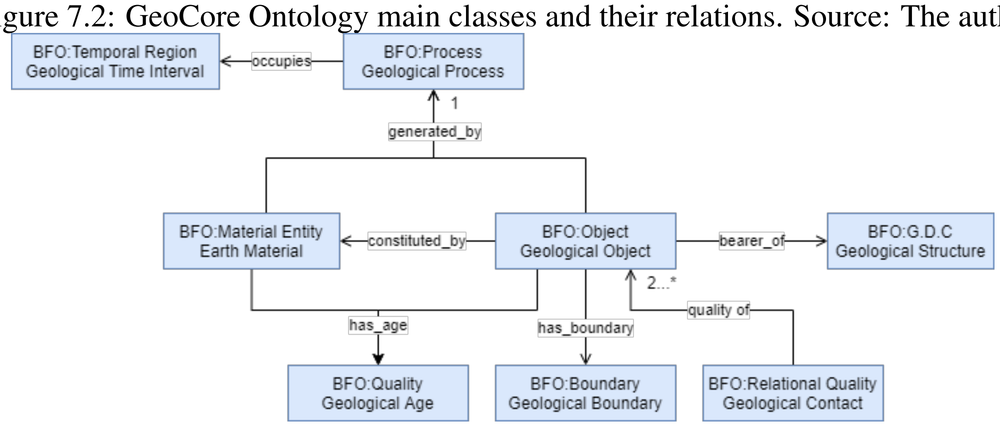


Figure 7.2: GeoCore Ontology main classes and their relations. Source: The author.



[1]https://github.com/bdi-ufrgs/geocoreontology

In section 7.1, we present the natural language and formal definitions of all the classes that compose the ontology. After that, in section 7.2, we present a simple use case of how we can use the GeoCore Ontology to analyze and represent the geological domain.

7.1 Ontology Definitions

In this section, we present our ontological definitions. We used Aristotelian definitions for the natural language part and a fol-like language for the formal part. In this language, the predicates corresponding to classes are represented by a term starting with a capital letter. In contrast, the predicates corresponding to n-ary relations are represented by a term starting with a lowercase. Variables in this language are represented by single letters also in lowercase. We also use logical operators from first-order logic: $\forall$ (universal quantification), $\exists$ existential quantification, $\wedge$ conjunction, $\vee$ disjunction, $\rightarrow$ implication, and $\neg$ negation. Our axiomatization is inspired by BFO's axiomatization for the ISO Standard[2], thus, predicates are time-indexed.

In our language, the axiom below says that the particular x is an instance of the class Object in the time t.

$\forall x, t(Object(x, t))$

The axiom above is thus equivalent to the following axiom:

$\forall x, t(instanceOf(x, Object, t))$

7.1.1 Generation relation

generated by =def. is a BFO:participatesIn relation where the participation of an *Earth Material* or a *Geological Object* is in the virtue of their generation.

Domain: Earth Material or Geological Object.

Range: Geological Process.

Only earth materials and geological objects can be generated by geological pro-

[2]https://github.com/BFO-ontology/BFO-2020

cesses:

$$\forall x, y, t(generatedBy(x, y, t) \\ \rightarrow ((EarthMaterial(x, t) \vee GeologicalObject(x, t)) \wedge GeologicalProcess(y, t))) \tag{7.1}$$

If something is generated by a process, then it participates in that process:

$$\forall x, y, t(generatedBy(x, y, t) \rightarrow participatesIn(x, y, t)) \tag{7.2}$$

When a geological process creates a new geological object or earth material, we say that it generated this new geological object or earth material. Thus, every instance of earth material and geological object has a unique generation relation with one instance of geological process.

7.1.2 Constitution relation

constituted by =def. is the relationship between something and what it is made of.

Domain: Geological Object, Earth Material.

Range: Geological Object, Earth Material, Aggregate.

This is a material constitution relation intended to represent the relationship between material objects, such as geological objects and earth materials, or aggregates of them.

Earth materials constitute geological objects and aggregates of geological objects constitute earth materials:

$$\forall x, y, t(constitutedBy(x, y, t) \rightarrow \\ (GeologicalObject(x, t) \wedge EarthMaterial(y, t))) \vee \\ (EarthMaterial(x, t) \wedge AggregateOfGeologicalObject(x, t))) \tag{7.3}$$

If y constitutes x then x and y are spatially identical to:

$$\forall x, y, t(constitutedBy(x, y, t) \rightarrow spatiallyIdenticalTo(x, y, t)) \tag{7.4}$$

7.1.3 Geological Process

Geological Process =def. is a physical, or chemical, or biological, naturally occurring *BFO:Process* that *occurs on* the Earth's surface or subsurface and *occupies* some *Geological Time Interval.*

Geological processes are macro processes that generate, transform, deform, transport, or destroy geological objects and earth materials. These processes are not necessarily atomic and may have other geological processes as parts.

Geological Process is a BFO:Process:

$$\forall x, t(GeologicalProcess(x,t) \rightarrow Process(x,t)) \tag{7.5}$$

For every Geological Process, there exists at least one Geological Object in which the process occurs and at least one Geological Time Interval that the process occupies:

$$\forall x, t(GeologicalProcess(x,t) \rightarrow \exists y, z(GeologicalObject(y,t) \\ \wedge GeologicalTimeInterval(t) \wedge occursIn(x,y,t) \wedge occupiesTemporalRegion(x,t))) \tag{7.6}$$

Examples are the process of deposition, the process of folding, a tectonic process, a sedimentary process.

7.1.4 Earth Material

Earth Material=def. is a *BFO:Material Entity* that is a natural amount of matter *generated by* some *Geological Process*.

Earth Materials are natural amounts of matter. Thus, they come into existence by nature, without any artificial aid. Since they are amounts, they don't hold unity criteria, but they are ontologically rigid and provide an identity criteria. Earth Materials and BFO Objects are disjoint. They are either solid, fluid, or unconsolidated. We usually observe earth materials when constituting other objects, such as an amount of rock constituting a geological unity or an amount of sand constituting a dune.

Earth Material is a BFO:Material Entity:

$$\forall x, t(EarthMaterial(x,t) \rightarrow MaterialEntity(x,t)) \tag{7.7}$$

Every Earth Material is generated by a Geological Process:

$$\forall x, t, t'(EarthMaterial(x,t) \rightarrow \exists y(GeologicalProcess(y,t') \wedge generatedBy(x,y,t') \wedge precedes(t',t)))) \tag{7.8}$$

Earth Material is either solid, fluid, or unconsolidated, but not all at the same time:

$$\forall x, t(EarthMaterial(x,t) \rightarrow (Solid(x,t) \vee Fluid(x,t) \vee Unconsolidated(x,t)) \tag{7.9}$$

$$\forall x, t(EarthMaterial(x,t) \wedge Solid(x,t) \rightarrow \neg(Fluid(x,t) \vee Unconsolidated(x,t)) \tag{7.10}$$

$$\forall x, t(EarthMaterial(x,t) \wedge Fluid(x,t) \rightarrow \neg(Solid(x,t) \vee Unconsolidated(x,t)) \tag{7.11}$$

$$\forall x, t(EarthMaterial(x,t) \wedge Unconsolidated(x,t) \rightarrow \neg(Solid(x,t) \vee Fluid(x,t)) \tag{7.12}$$

Since Earth Materials don't have unity criteria and BFO:Objects do, we can say that their instances are disjoint:

$$\forall t \neg \exists x(EarthMaterial(x,t) \wedge Object(x,t)) \tag{7.13}$$

Examples are an amount of sandstone, an amount of petroleum, an amount of natural gas.

7.1.5 Earth Fluid

Earth Fluid def=. is an Earth Material that is fluid.

Earth fluids can be amounts of water, oil, gas or a mixture of those fluids.

Earth Fluid is a fluid Earth Material:

$$\forall x, t(EarthFluid(x,t) \rightarrow (EarthMaterial(x,t) \wedge Fluid(x,t))) \tag{7.14}$$

7.1.6 Amount of Mineral

Amount of Mineral =def. is an Earth Material that is a naturally occurring, inorganic, solid, homogeneous chemical compound with a crystalline structure.

Amounts of mineral lack unity criteria as every other Earth Material. We usually observe them when they are constituting objects such as crystals or grains.

An amount of mineral is a solid earth material:

$$\forall x, t(AmountOfMineral(x, t) \rightarrow (EarthMaterial(x, t) \wedge Solid(x, t))) \quad (7.15)$$

Examples are the amount of quartz that constitutes a grain, the amount of feldspar that is part of an amount of rock.

7.1.7 Amount of Rock

Amount of Rock =def. a solid consolidated Earth Material that is constituted by an aggregate of particles made of mineral matter or material of biological origin.

Geologists define rocks at a scale of observation where they consider them homogeneous, even though an aggregate of solid particles constitutes it. These particles are usually geological objects, such as grains or crystals, or the rest of dead animals or plants. Amounts of rock, like other earth materials, are independent rigid entities that do not hold any unity criteria. We can observe them in nature when they are constituting objects such as geological unities.

Amount of rock is a consolidated earth material:

$$\forall x, t(Consolidated(x, t) \rightarrow Solid(x, t)) \quad (7.16)$$

$$\forall x, t(AmountOfRock(x, t) \rightarrow (EarthMaterial(x, t) \wedge Consolidated(x, t))) \quad (7.17)$$

Amount of rock is constituted by an aggregate of geological objects:

$$\forall x, t(AmountOfRock(x, t) \rightarrow \exists y(AggregateOfGeologicalObject(y, t) \wedge constitutedBy(x, y, t))) \quad (7.18)$$

Examples of amounts of rocks are the amount of sandstone that constitutes a well core, the amount of conglomerate that constitutes a geological unit.

7.1.8 Geological Object

Geological Object=def. is an *BFO:Object* that is generated by some *Geological Process* and has at least one part constituted by some *Earth Material* that is not an *Earth Fluid*.

A geological object is a naturally occurring entity because a geological process generates it, and some earth material constitutes it. Thus, we can differentiate geological objects from artificial objects, such as a well-core, because even though some earth material constitutes artificial objects, they are human-made rather than generated by some geological process. Furthermore, Geological Objects are specializations of BFO Objects, meaning they must necessarily have some unity criteria. The unity is what differentiates geological objects from Earth Materials.

A geological object is a BFO:Object generated by a geological process:

$$\forall x,t(GeologicalObject(x,t) \rightarrow (Object(x,t) \wedge \exists y(GeologicalProcess(y,t) \wedge generatedBy(x,y,t)))) \quad (7.19)$$

A Geological Object has at least one part that is constituted by some Earth Material that is not an Earth Fluid.

$$\forall x,t(GeologicalObject(x,t) \rightarrow \exists y,z(GeologicalObject(y,t) \wedge continuantPartOf(y,x,t) \wedge EarthMaterial(z,t) \wedge \neg EarthFluid(z,t) \wedge constitutedBy(y,z,t))) \quad (7.20)$$

The disjointness of Earth Material and Geological Object is already covered by the axiom 7.13.

Examples are the Earth, a geological unit, a grain, a crystal.

7.1.9 Geological Boundary

Geological Boundary =def. a *BFO:Fiat Surface* that is *located in* the external surface of a *Geological Object*.

The geological boundary of an object coincides with the complete physical discontinuity that delimits a Geological Object.

Geological boundary is a BFO:Fiat Surface located in a geological object:

$$\forall x,t(GeologicalBoundary(x,t) \rightarrow FiatSurface(x,t)) \tag{7.21}$$

$$\forall x,t(GeologicalBoundary(x,t) \rightarrow \exists y(GeologicalObject(y,t) \wedge locatedIn(x,y,t))) \tag{7.22}$$

Examples are the complete boundary of a grain, the upper boundary of a sedimentary layer.

7.1.10 Geological Contact

Geological Contact =def. is a *BFO:Relational Quality* that *inheres in* two distinct *Geological Objects* that are *externally connected with* each other.

A geological contact exists when two distinct geological objects are externally connected with each other, i.e., their external boundaries are physically adjacent. Objects that are in contact do not have any kind of proper parthood relationship between them.

Geological contact is a BFO:Relational Quality:

$$\forall x,t(GeologicalContact(x,t) \rightarrow RelationalQuality(x,t)) \tag{7.23}$$

Geological contact occurs between at least two distinct geological objects that are externally connected with each other:

$$\begin{aligned}&\forall x,t(GeologicalContact(x,t) \rightarrow \\ &\quad \exists y,z(GeologicalObject(y,t) \wedge GeologicalObject(z,t) \wedge \neg(y=z) \\ &\quad \wedge specificallyDependsOn(y,x,t) \wedge specificallyDependsOn(z,x,t) \\ &\quad \wedge externallyConnectedWith(z,y,t)))\end{aligned} \tag{7.24}$$

Examples are a geological contact between the bottom layer and the upper layer of a geological unit.

7.1.11 Geological Structure

Geological Structure =def. is a *BFO:Generically Dependent Continuant* that is the pattern of a non-atomic *Geological Object's* internal arrangement.

Geological structures are general material patterns repeated in many geological objects. The pattern comprises the material configuration and the mutual relationships of the object's different parts. Structures result from one or a series of geological processes that generated or transformed the geological object they *generically depend on*. Thus, there is some historical dependence relation between the structure and this geological process, but what concretizes the structure is some complex quality inhering in the object, not the process.

Geological structure is a BFO:Generically Dependent Continuant:

$$\forall x, t(GeologicalStructure(x, t) \rightarrow GenericallyDependentContinuant(x, t)) \tag{7.25}$$

A geological structure is concretized in a geological object by some BFO:Quality that inheres in the object:

$$\forall x, t(GeologicalStructure(x, t) \rightarrow \\ \exists y, z(GeologicalObject(y, t) \wedge Quality(z, t) \wedge inheresIn(z, y, t) \wedge concretizes(z, x, t))) \tag{7.26}$$

An example is a sedimentary planar stratification structure, resulting from a deposition process that generated a geological unit (a kind of geological object). The general pattern is the vertical stack of different layers (the geological structure), one on top of the other, concretized by the specific arrangement of the particular layers (the complex quality) that are part of this geological unit.

7.1.12 Geological Time Interval

Geological Time Interval =def. is a *BFO:Temporal Interval* that corresponds to a time interval within the Geologic Time Scale.

The Geological Time Scale is a system for chronological dating widely used in Geosciences.

Geological time interval is a BFO:Temporal Interval:

$$\forall t(GeologicalTimeInterval(t) \rightarrow TemporalInterval(t)) \tag{7.27}$$

Examples are the Cenozoic era, the Quaternary period.

7.1.13 Geological Age

Geological Age =def. a BFO:Quality that *inheres in* a Geological Object or Earth Material that corresponds to the Geological Time Interval in which the Geological Process that *generated* them *occupied*.

Geological age is the continuant counterpart of the geological time interval in which the process that generated it occupied.

Geological age is a BFO:Quality that inheres in a geological object or an earth material:

$$\begin{aligned}\forall x, t(GeologicalAge(x,t) \rightarrow (Quality(x,t) \\ \wedge \exists y((GeologicalObject(y,t) \vee EarthMaterial(y,t)) \wedge inheresIn(x,y,t))))\end{aligned} \tag{7.28}$$

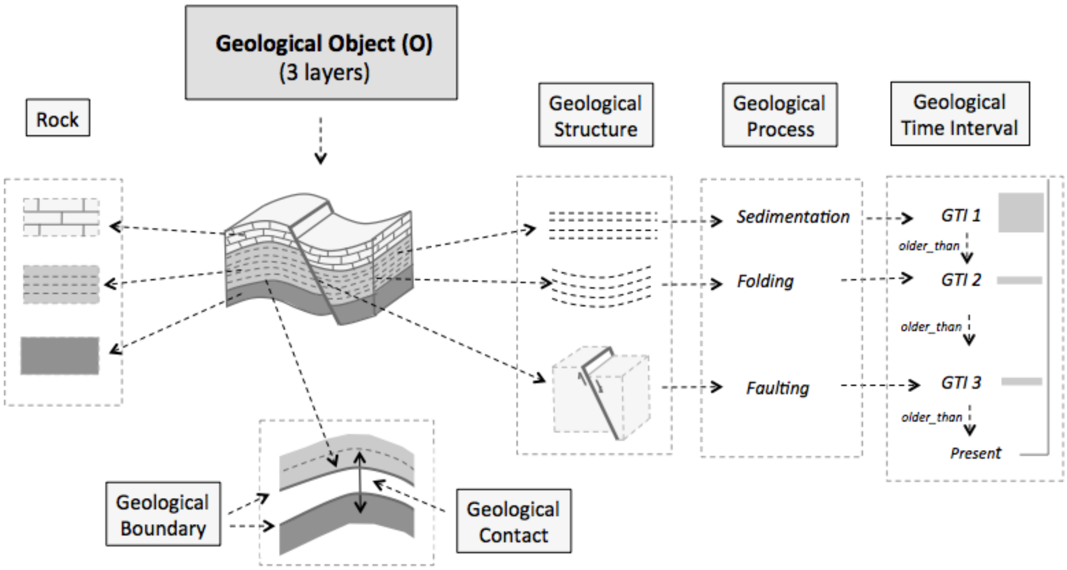
An example is the Albian geological age that inheres to a particular geological unit.

In the following section, we present a simple use case of the GeoCore Ontology.

7.2 Describing geological entities with GeoCore

GeoCore allows identifying and separating the different components of a geological object whatever its complexity. A *Geological Object* such as the object **(O)** represented in figure 7.3 can be decomposed in several elements. Object **(O)** consists in three layers (each one a *Geological Object as well*) in addition of their top and bottom boundaries (*Geological Boundaries*) related by *Geological Contacts*. Each layer is constituted by an instance of *Amount of Rock*.

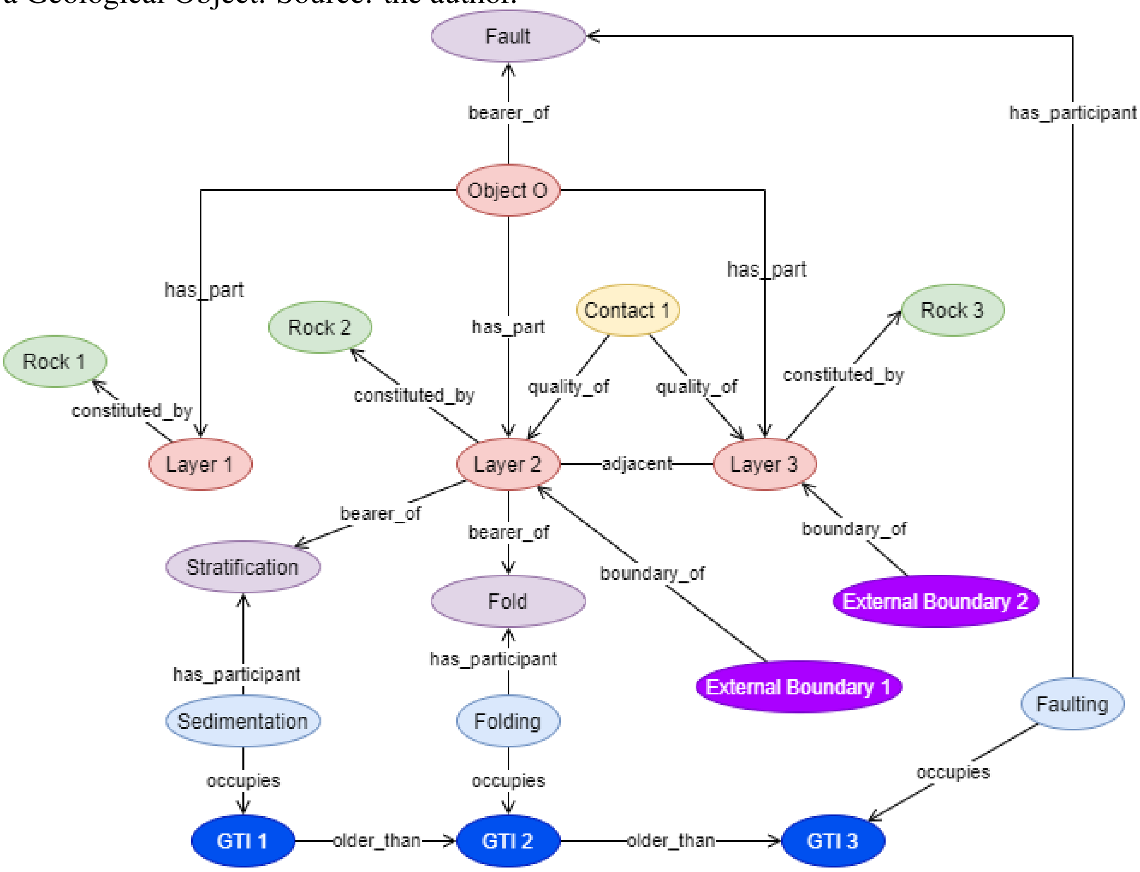
Figure 7.3: Entities related to Geological Object (Object O). Source: the author.



The internal arrangement of a Geological Object is defined by its genetic organization: in the case of sedimentary objects (such as Object **O**), they can be superposed layers each possibly having a specific stratification. In this case, this original structure is the result of the sedimentation process that generated Object **O**. However, for metamorphic or igneous environments (distinct from sedimentary environments), the internal arrangement corresponds as the internal zones or bands of the objects.

In this case, Object **O** was later modified by tectonic processes (folding and faulting), which resulted in the different structures (fold and fault). Each of these processes occurred during a specific *Geological Time Interval*. Figure 7.4 represents the relationships existing between the instances of different ontological entities related to object **O**. Red ellipses are instances of *Geological Objects*, green ellipses are instances of *Amount of Rock*, light purple ellipses are instances of *Geological Structures*, light blue ellipses are instances of *Geological Processes*, dark purple ellipses are instances of *Geological Boundaries* and dark blue ellipses are instances of *Geological Time Intervals*.

Figure 7.4: Relationships provided by GeoCore and BFO between the entities related to a Geological Object. Source: the author.



8 VALIDATING GEOCORE ONTOLOGY

GeoCore is at a high level of abstraction, and its goal is to serve as a central hub for developing more specialized ontologies and integrating existing ontologies. Thus, we do not expect that one can directly represent every geological domain using GeoCore purely. Nevertheless, it should provide a common framework that ontologists can specialize for representing the geological data from their geological domains.

The approach we used to validate was composed of three main steps. The first one was to consider the competency questions provided by domain experts as requirements during the ontology development. If the ontology can fulfill all the competency questions, we can assume that we achieved at least partially our goal. The second step was to implement the ontology using the OWL language and check its consistency using an automated reasoner. The ontology should not contain any inconsistency in its implementation. The final step was to use GeoCore to represent a real use case provided by the domain experts. GeoCore should be able to provide the basic framework to represent the use case. If we are successful in all three steps, then we can consider that GeoCore is validated.

In the following sections, we analyze the competency questions that guided our work (section8.1), explainining how we checked our implementation consistency (section 8.2), presenting a real-world use case (section 8.3) and concluding with our final remaks on GeoCore Ontology validation (section 8.4).

8.1 Analyzing the competency questions

In this section, we analyze each of the six competency questions provided by the domain experts that we presented in section 6.1.5.2. Typically, these kinds of questions come together with expected answers, which allows the modeler to verify if its ontology does answer the question correctly. However, due to the high level of abstraction of GeoCore Ontology, the answers to the competency questions are not straightforward. Thus, we used them as requirements that our ontology had to fulfill rather than queries over the ontology. Nevertheless, GeoCore Ontology should meet all the requirements imposed by the competency questions.

Below, we enumerate the competency questions:

1. Does the ontology allow representation of the entities that occur in at least more

than one scale of analysis?

2. Does the ontology allow representation of a geological area in such detail to allow the interpretation of the spatial distribution of the objects contained on it?

3. Does the ontology allow representation of a geological area in such detail to allow petrological interpretation?

4. Does the ontology allow representation of a geological area in such detail to allow sedimentological interpretation?

5. Does the ontology allow representation of a geological area in such detail to allow stratigraphic interpretation?

6. Does the ontology allow representation of a geological area in such detail to allow structural interpretation?

Question 1 is relevant because it is related to the scale of analysis problem. An ontology willing to deal with entities from different geological domains should deal with entities that occur in different scales. GeoCore Ontology can deal with such entities thanks to the ontological pattern for material constitution presented in section 6.2.1. The pattern allows us to separate the matter from the object that it constitutes and shows that entities from minor scales, such as grains, when gathered together in some favorable circumstance, constitute entities of greater scales, such as an amount of rock that constitutes some geological object.

Question 2 is related to spatial distribution. It is an expected question for an ontology proposing to deal with material entities. The GeoCore Ontology deals with that problem in different but complementary ways. Firstly, since GeoCore is an extension of the BFO top-level ontology, every independent continuant defined on it, be it material or immaterial, may occupy some Spatial Region of zero, one, two, or three dimensions. This relation allows us to represent the spatial position of the independent geological entities. Furthermore, we can represent the specific boundaries of some geological object with the Geological Boundary class, the specific contacts between two distinct objects with the Geological Contact class, and spatial relationships between the objects with the set of relations presented in section 4.2.

Questions 3 to 6 are all related to different disciplines of Geology. They are crucial to GeoCore because these disciplines are the ones that study the natural entities in Geology. Each of these disciplines is highly complex on its own and undoubtedly deserves its domain ontology. Regardless, the goal of GeoCore Ontology is to provide a common framework for developing and integrating such domain ontologies. For that reason, in the following, we analyze each question considering the definition given by Britannica Encyclopedia[1] for each discipline. We may consider that GeoCore meets the requirements imposed by the competency question if it can provide the basic top-level classes to develop these domain ontologies.

Question 3 is related to Petrology. In (BRITANNICA, 2018a), Petrology is defined as the scientific study of rocks that deals with their composition, texture, structure, occurrence and distribution, and origin in relation to physicochemical conditions and geologic processes. Petrology is a science that is more concerned with amounts of rock than the objects that they constitute. Nevertheless, an ontology for Petrology will have at its core specializations of the classes amount of rock from GeoCore, the qualities that inhere in them, and the geological processes associated with them. In addition to that, the ontology would need specializations of the class mineral and their relation to amounts of rock, which can be dealt with by considering the ontological pattern for material constitution. We may consider then that GeoCore provides the basic classes needed to develop an ontology of Petrology.

Question 4 is related to the Sedimentology discipline. In (BRITANNICA, 2018b), Sedimentology is defined as the scientific discipline that is concerned with the physical and chemical properties of sedimentary rocks and the processes involved in their formation. Sedimentary rocks are the rocks formed by the accumulation or deposition of mineral or organic particles at the Earth's surface. Thus, any instance of sedimentary rock is indeed an instance of a GeoCore amount of rock. Furthermore, sedimentary rocks' physical and chemical properties are dependent entities that can be represented by specializing the quality class from BFO. Finally, the processes that create sedimentary rocks are GeoCore geological processes and related to amounts of rock using the generated by relation. Therefore, GeoCore Ontology, combined with BFO, provides the main classes required by an ontology of Sedimentology to represent knowledge of this discipline.

Question 5 is related o Stratigraphy. In (BRITANNICA, 2014), the authors defined Stratigraphy as the scientific discipline concerned with the description of rock successions and their interpretation in terms of a general time scale. Rock successions are bodies of rock deposited one on top of the other following the law of superposition. An ontology of Stratigraphy would need to define specific types of rock bodies, processes and related

[1] https://www.britannica.com

them to the geological time scale. We can represent more specific rock bodies in GeoCore by specializing the geological object class and the amounts of rock that constitute them. In addition to that, GeoCore covers the time aspect of Stratigraphy with the classes of geological age and geological time interval. This framework allows us to represent a specific age and relate the object to a specific time interval within the geological time scale in which a geological process occupied that this object participated. Consequently, GeoCore provides the required classes to develop an ontology of Stratigraphy.

Finally, question 6 is related to Structural Geology. In (BRITANNICA, 2018c), Structural Geology is defined as the scientific discipline that deals with the geometric relationship of rocks and geologic features in general. In the view we adopt with GeoCore, the object that an amount of rock constitutes is the entity that provides its geometric characteristics. In GeoCore, the internal geometric arrangement of a geological object is its geological structure. Geometric relationships between distinct objects may be either geological contacts or spatial relations, covered respectively by the geological contact class and the geological spatial relations. General geological features of amounts of rock are dependent entities that characterize them, therefore, qualities from BFO. Thus, GeoCore ontology combined with BFO can provide the basic classes required to develop an ontology of Structural Geology.

Analyzing each of the questions that the domain experts provided, we conclude that GeoCore Ontology combined with BFO meets its modeling requirements. However, to fully validate it, we must yet verify its implementation consistency and show its adequacy to a real-world use case. In the next section, we offer the results of GeoCore consistency checking with the support of an automated reasoner.

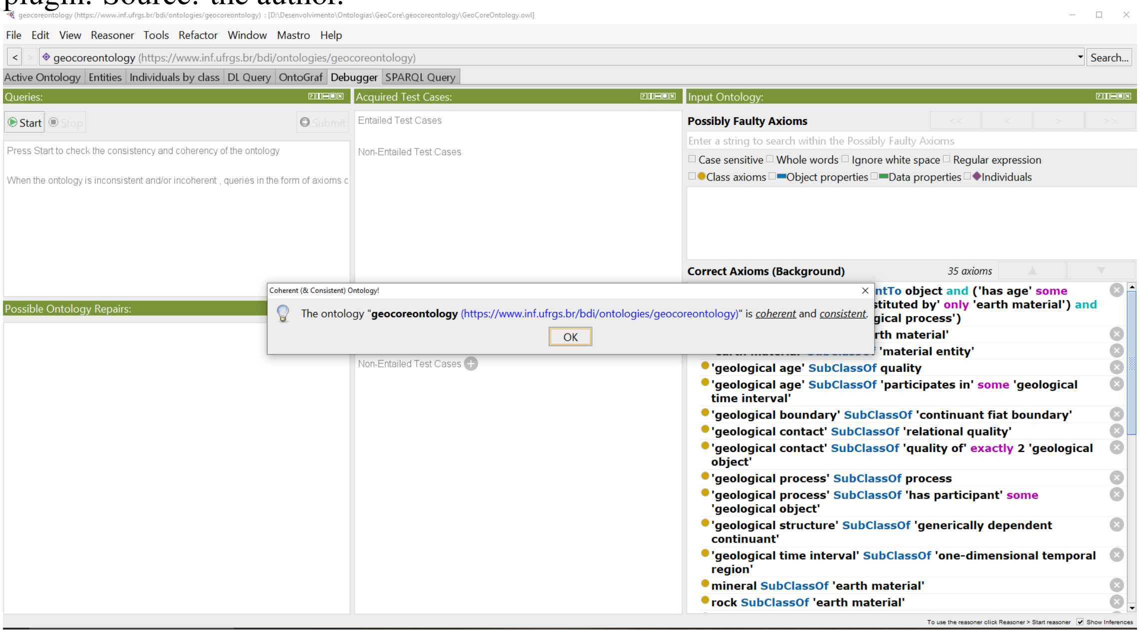
8.2 Consistency checking

In order to verify the consistency of GeoCore Ontology, we have implemented it using the Web Ontology Language (OWL). We choose OWL for a variety of reasons. The first one is because OWL is a Semantic Web language designed to represent knowledge. Secondly, OWL is a computational logic-based language that allows automated reasoners to verify the consistency of the represented knowledge. Besides that, it is a W3C standard for representing ontologies, it is one of the most common languages for ontology sharing, and there is an up-to-date BFO implementation available.

Before starting our implementation, we had to import the BFO Ontology, which has an OWL implementation available on its Github repository [2]. After importing BFO, we created the classes and the axioms of the ontology according to the formal definitions that we have presented in section 7.1.

For consistency checking, we used the Protégé plugin OntoDebug, an interactive ontology debugging plugin (SCHEKOTIHIN; RODLER; SCHMID, 2018). This plugin implements many black-box algorithms for ontology debugging and combines modern reasoners such as Pellet (SIRIN et al., 2007) and Hermit (MOTIK; SHEARER; HORROCKS, 2009) for ontology consistency checking. We only used the consistency checking feature from this plugin. Figure 8.1 shows that the result of OntoDebug's consistency checking of GeoCore Ontology is that the ontology is consistent and coherent.

Figure 8.1: The result of GeoCore Ontology consistency checking using the OntoDebug plugin. Source: the author.



Now that we have shown that GeoCore Ontology itself is consistent, we may present a real-world use case. In the next section, we demonstrate how we can use GeoCore Ontology to represent an integrated geological study.

8.3 A use case from the Campos Basin

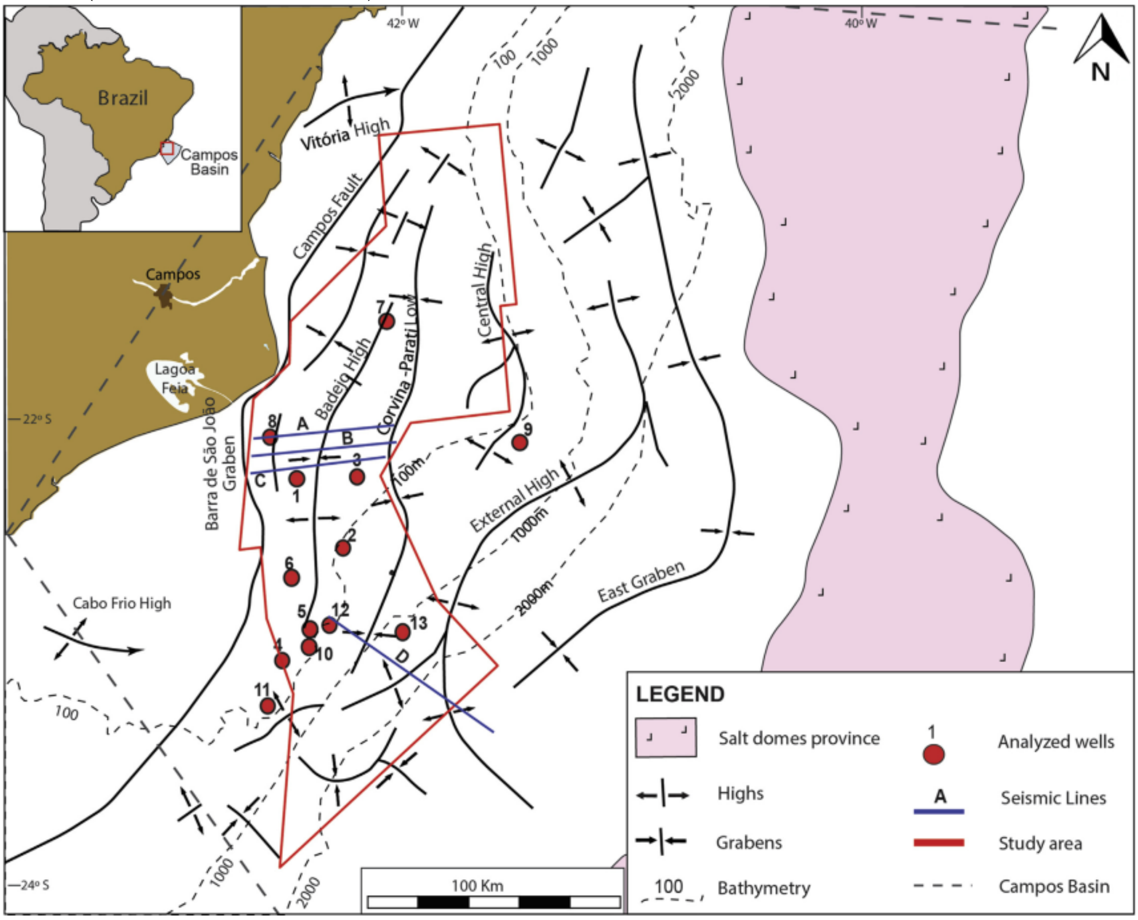
In this section, we present a real use case to show GeoCore's adequacy to real-world data. The use case that we give here is described in (GOLDBERG et al., 2017). We

[2]https://github.com/BFO-ontology/BFO

carefully selected this study with the support of domain experts. It is relevant to us because the authors constructed this study integrating seismic, sedimentologic/stratigraphic, and petrologic data. As one might expect, such a study is highly complex, and its complete understanding is outside of the scope of this thesis. For that reason, we only describe it partially. However, we consider that the portion presented is enough to demonstrate our ontology adequacy.

We focus in this section on the stratigraphic/sedimentologic and petrologic data of the Lagoa Feia group located on the Campos Basin at the eastern Brazilian coast. Figure 8.2 presents the Campos Basin location map; the study area is delimited within the red lines. The red dots are the locations where they drilled wells and retrieved well cores and thin sections.

Figure 8.2: Location map of the Campos Basin at the eastern Brazilian coast. Extracted from: (WINTER et al., 2007).



In the following, we first analyze the stratigraphical part of the study. For that, we define a basic stratigraphical framework to distinguish the relevant geological entities, referring to the definitions of the North American Stratigraphic Code (NASC) (NOMENCLATURE, 2005). After that, we analyze the sedimentological and petrological part of the study. Since the authors used the lithofacies notion to integrate this part of the data, we also define a basic framework for representing lithofacies to properly represent it.

In the NASC, the authors say that lithostratigraphic units are the basic units of general geologic work and serve as the foundation for delineating strata, local and regional structures, economic resources, and geologic history in regions of stratified rocks. They define a lithostratigraphic unit as a body of sedimentary, extrusive, igneous, metasedimentary, or metavolcanic strata that is distinguished and delimited on the basis of lithic characteristics and stratigraphic position. Lithics characteristics include chemical and mineralogical composition, texture, and features as color, structures, fossils (view as rock-forming particles), or other organic content. Lithostratigraphic units may be ranked in a supergroup, group, formation, member, and bed.

From the definition above, we may conclude that a lithostratigraphic unit is a geological object constituted by sedimentary, extrusive, igneous, metasedimentary, or metavolcanic amounts of rock.

According to the NASC, the most important unit is the formation. It is a body of rock identified by lithic characteristics and stratigraphic position. The content of a formation should possess some degree of internal lithic heterogeneity, and it should contain between its limits rock of one lithic type, repetitions of two or more lithic types, extreme lithic heterogeneity that in itself may constitute a form of a unit when compared to adjacent rocks. Formations may have a multiplicity of members as parts. Considering formation's definition, we may conclude that it is a specific type of lithostratigraphic unit.

The unit in the rank below is the member. A member is a lithostratigraphic unit that is a part of some formation, and it is recognized because it possesses characteristics distinguishing it from adjacent parts of a formation. Thus, a member is also a specific type of lithostratigraphic unit that is always a proper part of some formation.

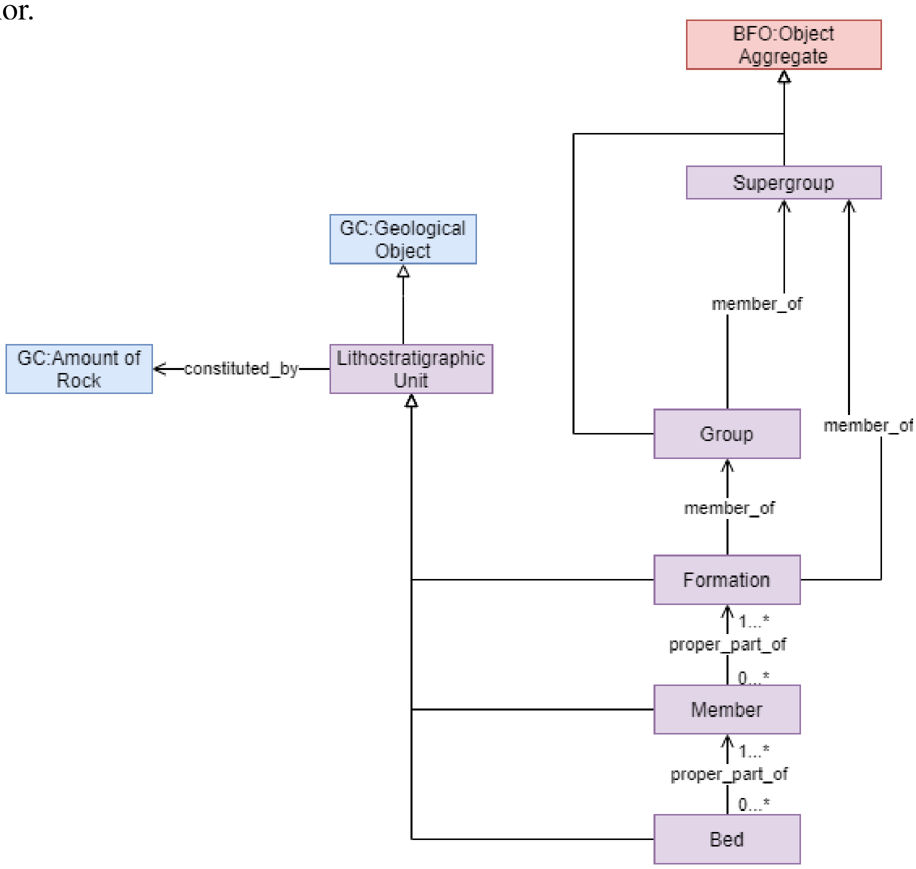
The smallest unit in the rank is the bed. It is a lithostratigraphic unit of sedimentary rock that is part of a member. Analogous to the member, it is a specific type of lithostratigraphic unit that is always a proper part of a member.

A problem arises when we analyze the definition of the term group. This is because the NASC defines the group as the unit higher in rank to formations and that it may consist entirely of named formations or, alternatively, need not to be composed entirely by named formations. The problem only becomes evident when they propose that groups should be defined to express the natural relations of associated formations. This leads us to believe that groups aren't objects but rather object aggregates. Consequently, it is a mistake to say that a group is a lithostratigraphic unit because these units are objects, while groups are aggregates, entities of distinct ontological nature.

The same problem occurs with the definition of supergroup. According to the NASC, a supergroup is a formal assemblage of related or superposed groups or of groups and formations. Supergroups are aggregates of lithostratigraphic units but shouldn't be lithostratigraphic units themselves.

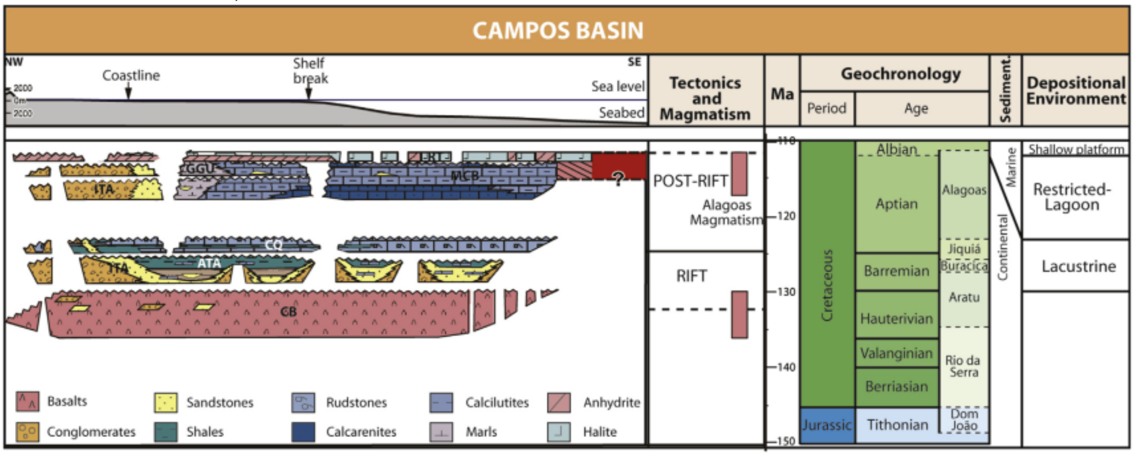
We summarize a basic stratigraphic framework in figure 8.3. A lithostratigraphic unit is a GeoCore geological object that is constituted by some GeoCore amount of rock. Formation, member, and bed are lithostratigraphic units, while supergroup and group are BFO object aggregates. The parthood relations between member and formation, and bed and member, are proper parthood relations. On the other hand, the parthood relations between supergroups, groups, and formations are memberhood relations.

Figure 8.3: Basic ontological framework for representing lithostratigraphic units. Source: the author.



Now that we have defined a basic stratigraphic framework, we may continue to analyze the relevant entities present in the use case. We start with the Campos Basin, which is an instance of a Basin. According to (HYNE, 2012), a Basin is a large area with a thick accumulation of sedimentary rocks. In our view, a Basin is a 3d immaterial entity, thus a BFO Site. The next relevant entity is the Lagoa Feia group. Lagoa Feia is an instance of the class group defined as a BFO object aggregate of formations. The members of Lagoa Feia are all instances of the formation class. They are: Itapaboana Formation (ITA); Atafona Formation (ATA); Coqueiros Formation (CQ); Gargau Formation (GGU); Macabu Formation (MCB); and Retiro Formation (RT). In figure 8.4, we have a stratigraphic chart that shows all the formations that are members of the Lagoa Feia group. On the left side of the figure, we have a visual representation of the succession of formations. In this stack of formations, newer formations are placed on top of older formations. The chart also presents the Cabiúnas Formation (CB), which is not a member of the Lagoa Group but is directly below it. The colored patterns filling the formations represent the rock types that constitute each one of the formations. On the right side of the figure, we have information on tectonic and magmatic processes that occurred, the geological time scale, information of sediment provenance, and interpretation of the depositional environment of the formations.

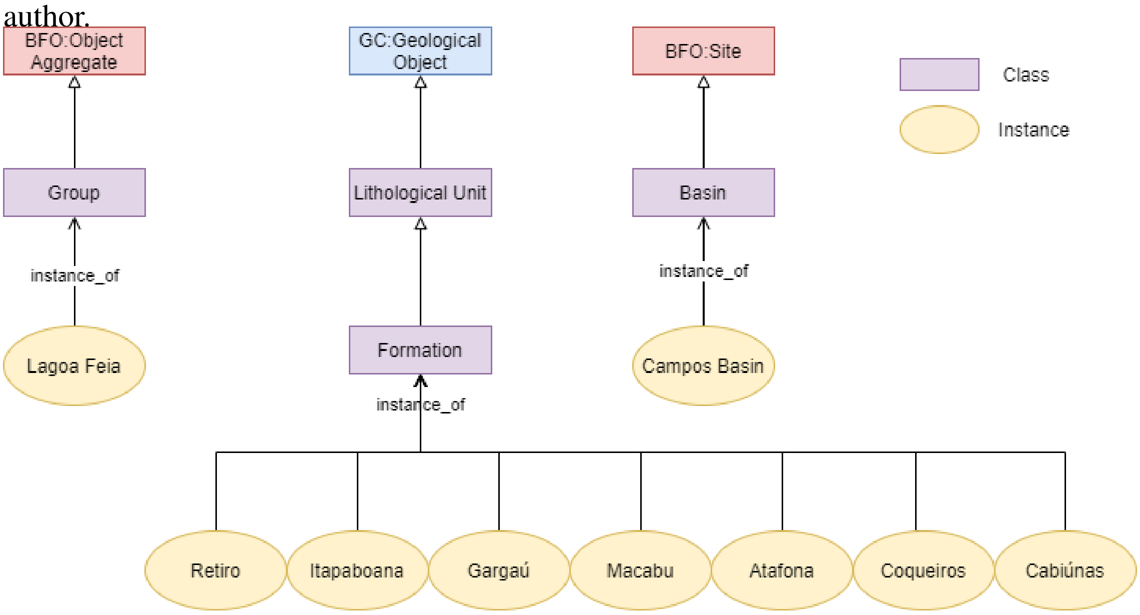
Figure 8.4: Stratigraphic Chart of a part of the Campos Basin. Extracted from: (GOLDBERG et al., 2017).



In figure 8.5, we show the instantiation relations of the entities presented so far. Rectangles represent classes, and ellipses represent instances. The relations between rectangles are subsumption relations, while relations between classes and ellipses are instantiation relations.

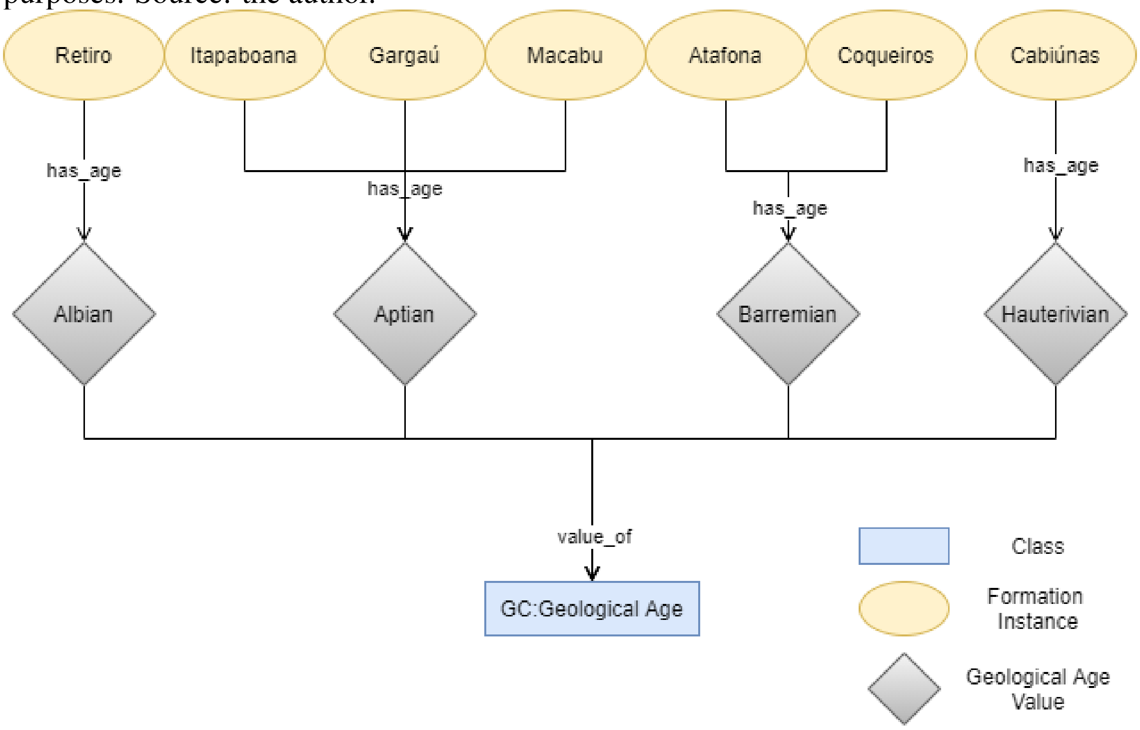
In our use case, a formation has a GeoCore geological age which is related to the geological time interval that the geological process which generated the formation occupies. Thus, for each instance formation, there is an instance of a geological age associated with it. In figure 8.6, we collapse for ease of reading the representation of the instances of geological age and their value in a single gray diamond. The Retiro Formation has an Albian age, the Itapaboana, Gargaú, and Macabu Formations have an Aptian age, the

Figure 8.5: The instantiation relations of the relevant geological entities. Source: the author.



Atafona and Coqueiros Formations have a Barremian age and the Cabiúnas Formation has a Hauterivian age.

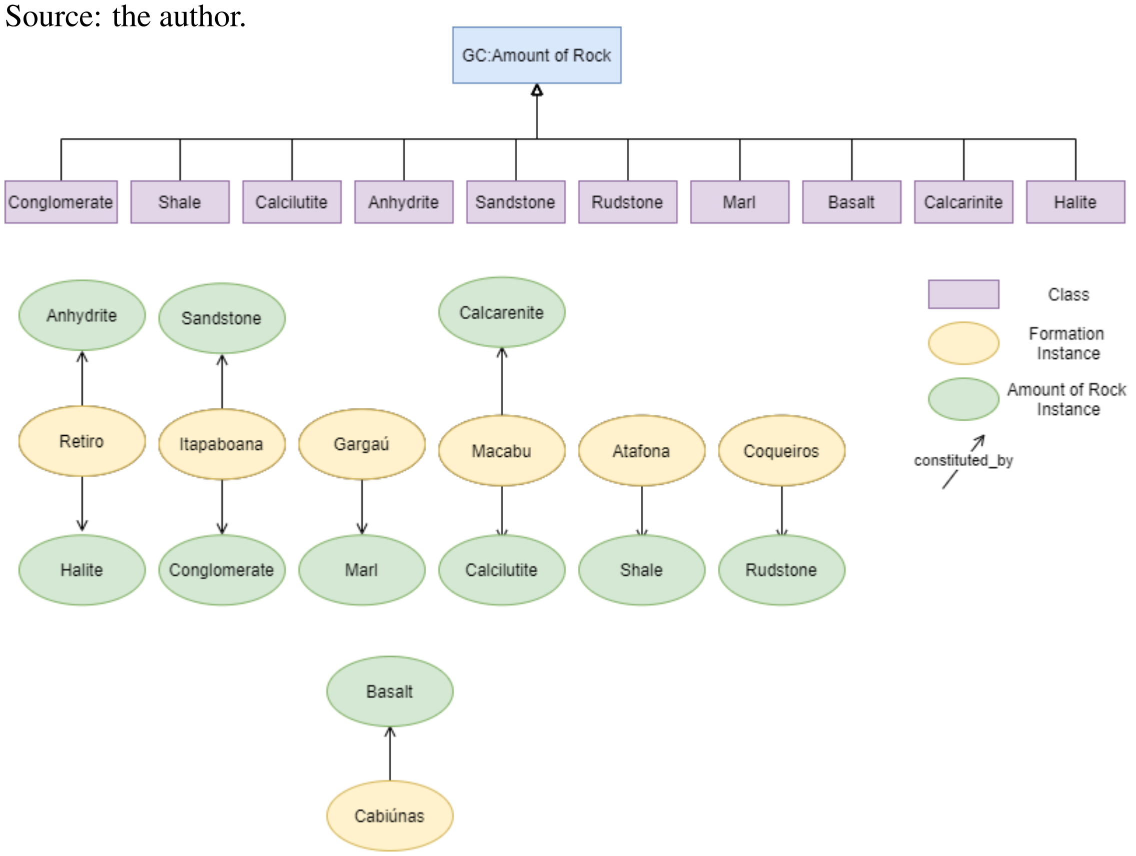
Figure 8.6: The geological age of each Formation instance. The instances of the geological ages and their value are collapsed for on the gray diamonds for representation purposes. Source: the author.



Because a formation is a lithostratigraphic unit, every formation is constituted by some amount of rock. In our use case, there are ten distinct types of amounts of rock.

However, due to the resolution of the data in the geological study, it is not possible to identify specific instances of rock. For that reason, we represent the instances of amounts of rock using the same name of the rock type that they instantiate in the figure 8.7. Green ellipses are amount of rock instances, yellow ellipses are formation instances, and purple rectangles are specific classes of amounts of rock. The relation between the rectangles is one of subsumption, while the relation between ellipses is one of constitution.

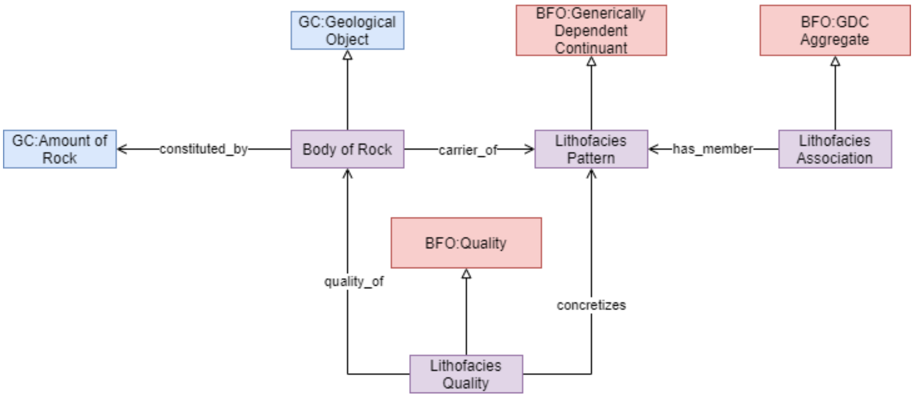
Figure 8.7: The formation instances and their constitution relations with amounts of rock. Source: the author.



The sedimentological and petrological analysis included the description of 340 meters of well cores from 13 wells, and the petrographic analysis of 197 thin sections extracted from 9 wells. The authors integrated this data defining lithofacies and facies associations. A lithofacies is a group of characteristics of a body of rock that reflects some specific process or depositional environment. In this case, they considered the grain size, grain composition, sorting, fossil content, and sedimentary structures. A facies association in this context is a grouping of lithofacies regarding some criteria defined by the geoscientist. The facies notion is helpful because it encapsulates a group of common characteristics to different instances of bodies of rock. Thus, the geoscientist may analyze rock data from a small group of cores and thin sections and then extrapolate to other regions where direct rock data is not available.

To represent lithofacies, we have to take into account both the specific and general aspects of it. In our view, there are two distinct entities collapsed on the facies notion, the general lithofacies pattern and the specific lithofacies of a body of rock. The lithofacies pattern is a BFO generically dependent continuant comprising a group of rock body's characteristics that are repeatable in different body instances. The specific lithofacies is a complex BFO quality that concretizes a lithofacies pattern in a particular body of rock. In figure 8.8, we have this basic facies framework. The body of rock is a GeoCore geological object that is constituted by some GeoCore amount of rock. The body of rock is the carrier of a generic lithofacies pattern because it is the bearer of a specific lithofacies quality.

Figure 8.8: A basic ontological framework for representing lithofacies. Source: the author.

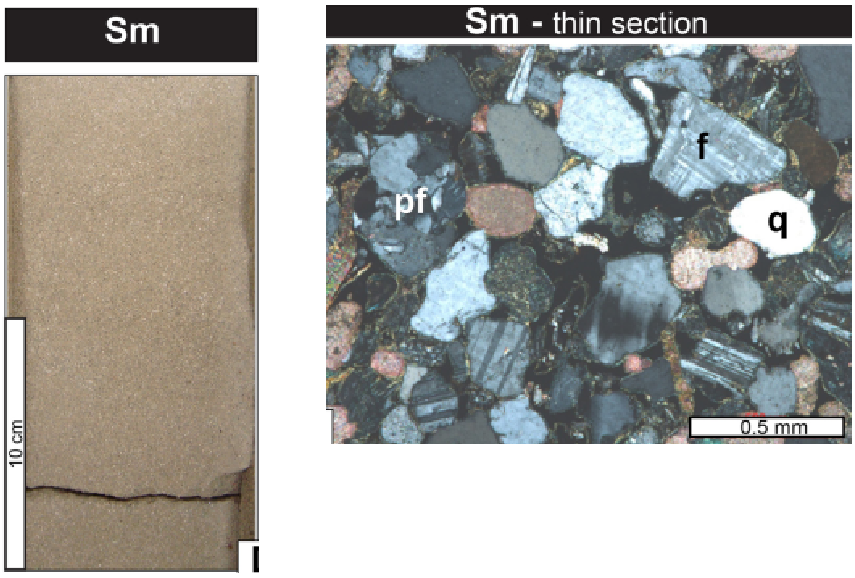


In the Lagoa Feia study, the authors identified seventeen distinct lithofacies patterns and grouped them into four lithofacies associations. In our context, the lithostratigraphic units are the carriers of the patterns. However, due to the lack of better data resolution, it is not possible to precisely identify the specific unit instances that are the carriers of the lithofacies patterns. Nevertheless, since we have only seven formation instances for seventeen distinct lithofacies patterns, we may conclude that at least ten other lithostratigraphic units exist and carry some lithofacies pattern. An example of a lithofacies pattern instance is the Sm[3] lithofacies. The authors define it as fine-grained to coarse-grained (grain size quality) sandstones (amount of rock constitution), with stevensite ooids/peloids, carbonate bioclasts, and mud intraclasts (composition quality). The geological process associated with carriers of this lithofacies pattern is the high-density turbidity current. It is important to note that the same term "Sm" can designate a pattern having a different group of characteristics in other contexts.

[3]According to the authors, the names of the lithofacies followed a code for facies naming, where the first capital letter represents the amount of rock type and the second lower-case letter represents the geological structure. Thus, Sm means a Sandstone with massive structure facies.

Figure 8.9 shows two samples carrying the Sm lithofacies pattern. On the left, we have a well core photo on the hand sample scale, where it is possible to identify the rock type of the amount of rock, and the average grain size of the sample. On the right, we have a photomicrograph of a thin section on the microscopic scale where it is possible to identify the sample composition and the minerals that constitute the grains and other particles. Ideally, we would like to know if the instance of the amounts of rock that constitute both samples are part of the same original amount of rock or not. However, due to data limitation, this is not possible for this study.

Figure 8.9: Two samples carrying the Sm lithofacies pattern. On the left part of the figure we have a photography of a well core. On the right part of the figure we have a microphotography of a thin section containing grains constituted by feldspar (f), grains constituted by quartz (q), and plutonic fragments (pf). Adapted from (GOLDBERG et al., 2017).

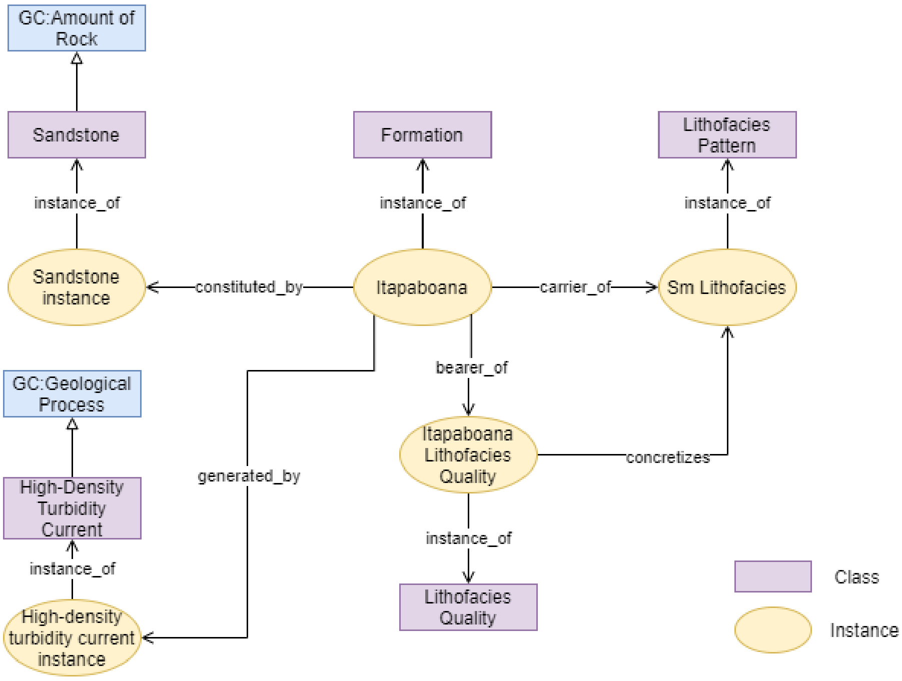


In figure 8.10, we represent the relations between the Sm lithofacies pattern and the Itapaboana Formation. The formation is the bearer of a specific instance of the lithofacies quality. This instance of quality concretizes the Sm lithofacies pattern on the formation, which carries this specific lithofacies pattern because a process of high-density turbidity current generated it. We omit the representation of other lithofacies patterns because it is analogous to Sm's representation.

Finally, representing lithofacies associations is trivial. Instances of lithofacies associations are related to instances of lithofacies patterns by a member of relation. For instance, the Sm lithofacies pattern is a member of the Re-sedimented coarse-grain dominated deposit facies association.

In this section, we used GeoCore Ontology to represent a real-world use case of

Figure 8.10: An example of representation of the Sm lithofacies pattern. Source: the author.



a geological study that integrates various Geology disciplines. It is not GeoCore's goal to define all the knowledge related to Geology. Thus the objective of this section isn't to show that GeoCore Ontology alone can represent all the geological data that exists. Still, we have shown that using GeoCore facilitates the development of domain ontologies suitable to represent domain knowledge. It facilitates because GeoCore provides a set of well-defined classes and relations that form the core of various geological disciplines, allowing the ontologist to focus on domain-specific terms.

In the next section, we present our final remarks regarding the validation of the GeoCore Ontology.

8.4 Final remarks on GeoCore Ontology validation

In this chapter, we presented our three-step plan for validating GeoCore Ontology:

- We showed that our ontology meets the experts' requirements of the competency questions.
- We showed that our ontology is consistent according to the Protégé plugin OntoDebug.
- We showed that our ontology could facilitate data representation of a complex real-

world use case.

Considering that we were successful in all the validation steps we proposed, we conclude that GeoCore Ontology is validated according to our standards. However, this is not to say that the ontology is perfect or even complete. We are aware that the life-cycle of an ontology is iterative, and further additions or modifications might be needed. Furthermore, the best possible validation for an ontology occurs when community members reuse it. For that reason, we are happy to see that there are already works that successfully used GeoCore, such as the work of Cicconeto described in (CICCONETO, 2021).

In the next chapter, we present the conclusions of this thesis and possible future works down the road.

9 CONCLUSION

The semantic interoperability problem is related to integrating resources that have different perspectives on the data. At the same time that advances in technology increase data acquisition and availability, the necessity of making diverse data readily integrable also increases. The problem is that heterogeneities in the data make it very hard to integrate.

In this work, we focused on semantic interoperability in the geological domain. During the geological interpretation, the geologist gathers data that ranges from a scale of millimeters to thousands of kilometers. Various professionals produce this data, resting many times on conflicting views of the domain and persisting it on incompatible data formats or databases. Thus, data integration in the geological domain heavily depends on the geologist's ability to integrate all this data manually.

In our view, to achieve interoperability for geological data, we need to ensure that geoscientists are talking and describing geological entities using a uniform view and a shared vocabulary over the domain.

Our work is an effort towards vocabulary uniformization in the geological domain. We base our approach on offering a core ontology for Geology. This ontology contains a limited set of classes with formal (computable) and natural language definitions representing geological entities shared in various geological subdomains.

For developing our core ontology, with the support of domain experts, we considered the disciplines of Structural Geology, Sedimentology, Stratigraphy, and Petrology. We identified the most relevant entities shared among these disciplines and formalized them following ontology engineering principles.

During the development of GeoCore, we adopted a view where we separated the amount of rock from the geological object that it constitutes. We identified that this approach helps to separate essential aspects of each one and understand how entities from different scales are related in Geology.

We can divide the contributions of this work into two levels: 1) in the conceptual level, with an ontological analysis of the geological domain that supports a clearer understanding of the domain; 2) in the application level, with a computational artifact that supports formal representation and automated reasoning of geological data.

In the following, we present our conclusions and possible future works.

9.1 Relevant aspects of the geological domain

Scales of Analysis. Geoscientists have to deal with entities ranging from the scale of a microscope to the scale of a continental basin. Geological interpretation is the result of gathering data from different scales and disciplines. The ontology we proposed takes into consideration this scale variability clarifying the relations of constitution and parthood. We also identified that the entity rock is vital for correctly understanding properties from different scales.

Description of interpretations. Geology is a science concerned with describing rocks, objects, and processes that exist on the Earth's surface and subsurface. However, due to the difficulty in observing the subsurface, geoscientists usually use indirect methods, such as seismic or well logging, for describing it. The result is that the geoscientists often describe their interpretation of this indirect data, leading to data discrepancy when we consider descriptions of different geoscientists. Our goal was to provide an ontological model that focuses on describing the geological reality objectively with a uniform view, independently of anyone's interpretation. Ideally, geological descriptions should capture directly observable characteristics of the geological entities. Our ontology supports this kind of description.

Domain entities and representations. Understanding that geological representations are different from the actual geological entities that they represent is crucial. Representations are informational entities - abstract entities-, while the entities they represent are usually material entities, such as amounts of rock and lithostratigraphic units. Our ontology supports the description of the real geological entities that the geological representations are referring to.

Geological qualities and the scales of analysis. Depending on the scale of analysis, geoscientists may focus on the material that constitutes an object or in the object itself. Separating the material from the object allows us to represent properties of the material and properties of the object independently, which facilitates data integration and extrapolation.

Different professionals. Due to the vast amount of data required for geological interpretation, data comes from professionals from different backgrounds, such as geologists, geophysicists, and engineers. An ontological model with a unified view of the domain ensures that the vocabulary used by these different professionals is explicit, eliminating possible ambiguities and data heterogeneity.

Vagueness. The subjects of study in Geology are crisp material entities or areas dependent on these entities. An object has definite and unique boundaries. The problem of different instances of material entities having the same term to designate them is a semantic problem, not a problem of object vagueness.

9.2 Philosophical choices

Realism. We assumed a philosophical view of realism in this work. We have this view because the goal of a science such as Geology is to describe the material entities that exist in reality. Conceptual models not grounded on reality may have their role in geological interpretation, but the task of geological description should focus on the real entities that may be equally observable by different people using the right tools.

The Constitution View. For developing the GeoCore Ontology, we assume a view where we treat an object and the matter that constitutes it as two distinct instances. This view enables us to represent an important characteristic of geological interpretation that is considering rocks and rock portions as distinct entities, each one having its characteristics, such as in the example of the rocks and rock portions we presented in chapter 5. The Constitution View also clarifies why geoscientists consider rocks as homogeneous on one scale and granular on another scale.

Continuants vs. Occurrents. We assumed a dichotomy of continuant vs. occurrent entities that many top-level ontologies adopt, including BFO. In this view, occurrents are entities that unfold in time, such as processes or events, and continuants are entities that persist in time and participate in these processes, such as material objects, sites, and regions. This view fits well the way geoscientists describe Geology.

Independent vs. Dependent Entities. In our work, we acknowledge that some dependent entities may only exist while other independent entities exist. It is the case of the age of some formation, the color of some amount of rock, or some pattern of matter concretizing a geological structure.

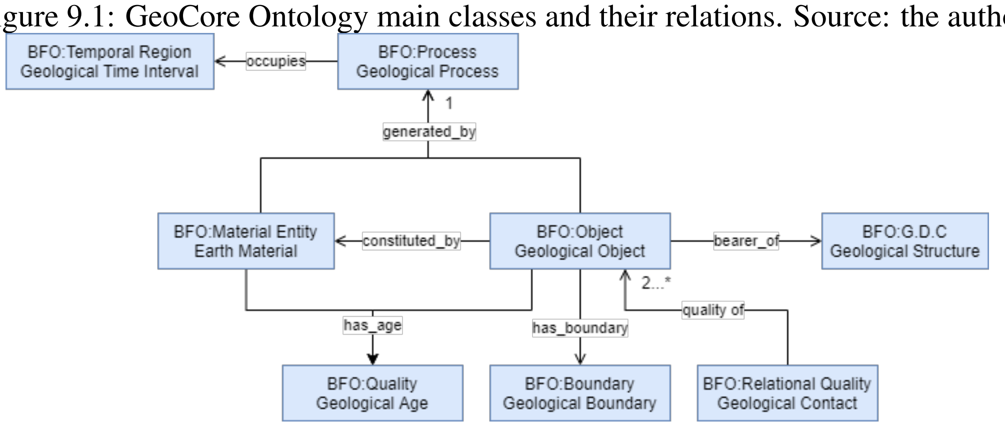
9.3 Ontological model

Competency Questions. The domain experts provided seven competency questions that the ontology should fulfill. Analyzing the competency questions and with the support of the domain experts, we defined that the scope of the ontology should englobe the disciplines of Petrology, Sedimentology, Stratigraphy, and Structural Geology. These are the disciplines that form the core of Geology and study the material entities objectively observed by geoscientists.

Terms present in the ontology. The ontology is composed of 11 classes representing geological terms shared in different disciplines of Geology. We identified two main entities whose instances are ontologically rigid - earth material and geological object -, three ontologically rigid entities that are specializations of earth material - amount of matter, amount of mineral, and earth fluid -, four dependent entities - geological age, geological structure, geological contact, and geological boundary, and two occurrents - geological process and geological time interval.

Ontological Model. The heart of GeoCore's ontological model is based on the idea that a geological object is constituted by some earth material. Both are generated by some geological process that occupies some geological time interval. The age of earth material or geological object is related to the time interval in which the process generated it occupied. Objects have physical boundaries and are possible bearers of geological structures. Finally, when two objects are physically adjacents, they are in geological contact. The figure 9.1 summarizes this model. This model can represent geological knowledge in different scales of Della Fávera's scales. Domain ontologies should specialize this basic framework in order to guarantee easier integration and reusability.

Figure 9.1: GeoCore Ontology main classes and their relations. Source: the author.



Basic Formal Ontology. We developed the GeoCore Ontology based on the BFO. BFO is a top-level ontology to support the development of ontologies which goal is to describe scientific domains. It is an ontology that is conformant with a realistic approach and has a well-defined set of primitives to represent material entities. All these characteristics suites an ontology that intends to deal with the representation of geological knowledge. Furthermore, it is an ontology with a vast user-base and has extensive documentation, which increases the likelihood of reuse of the ontologies developed based on it.

Integrating ontologies to GeoCore Ontology. We developed GeoCore in a modular way. We do not claim that GeoCore is complete and readily available for describing any domain of Geology. Nevertheless, GeoCore offers the possibility of extending it to specific domains of Geology. The fragment of an ontology for Stratigraphy that we provided in section 8.3 is an example of how to do that. Furthermore, all the classes we defined are consistent specializations of BFO classes. It is possible to integrate GeoCore with any BFO-based ontology or align with other ontologies that are not conflicting with it.

Validation. We validated GeoCore Ontology by analyzing the competency questions, checking our formal implementation, and representing a real-world use case that integrated different disciplines of Geology. However, in Computer Science, ontologies are knowledge artifacts developed for knowledge sharing and knowledge reuse. Full validation of GeoCore will only come with time when people use it to develop their domain ontologies and integrate them under GeoCore's umbrella.

9.4 Contributions to the state of the art

We contribute to state-of-the-art both at the application level and the conceptual level.

At the application level, we developed a formal core ontology for Geology. We implemented the ontology using the OWL language and based it on the BFO ontology. To our knowledge, GeoCore Ontology is the only core ontology in the domain developed considering ontology engineering principles and methodologies based on a top-level ontology such as BFO. Ontology developers can reuse our ontology for developing and integrating formal ontologies in the geological domain. We developed GeoCore in a modular way to facilitate its reusability.

At the conceptual level, we proposed an ontological model that captures a core structure of the geological domain. We created this model with the support of domain experts and considered a philosophical realist view over the domain. This model helps to understand how the main entities in the domain are related to each other. It also may help to evaluate legacy models in the light of an ontologically defined model.

The ontological pattern of the material constitution for Geology is also a novel contribution at the conceptual level. We analyzed a theory for material constitution from

Philosophy and proposed a computable ontological pattern for dealing with the relationship between objects and the matter that constitute them.

9.5 Future works

For developing our ontology, we considered the disciplines of Petrology, Sedimentology, Stratigraphy, and Structural Geology. Certainly, other disciplines, such as Mineralogy and Geochemistry, are relevant to Geology and deserve a proper ontological analysis and a possible extension to GeoCore.

We developed GeoCore Ontology to work as a central hub for the development of ontologies for Geology. Thus, a straightforward goal is to create an ontology network for Geology, composed of various domain ontologies defining their specific domains in a modular and integrated way.

The possible future extensions of GeoCore are many. Certainly, one of the most important integrations to GeoCore should be an ontology describing the major geological processes. A full formalization of the geological time scale integrated within the BFO/GeoCore framework is also an important future work.

In this work, we only considered the relevant material entities of the domain. However, information artifacts, the different representations of these material entities, such as maps, cross-sections, seismic images, well logs, are also important. Investigating the artifacts used in geological interpretation is crucial for maintaining data traceability. The Information Artifact Ontology, a BFO-based ontology developed for dealing with informational entities, is an excellent option for developing an ontology of geological information artifacts. These different information artifacts can refer to the material entities (or their specializations) defined in GeoCore Ontology.

The ontological pattern for material constitution showed an interesting pattern of aggregates constituting amounts constituting objects in the geological domain. Further investigations for checking if this pattern is valid for other natural domains are exciting future works.

REFERENCES

ABEL, M. et al. The multiple applications of a mature domain ontology. In: **ONTOBRAS**. [S.l.: s.n.], 2015.

ABEL, M. et al. A knowledge organization system for image classification and retrieval in petroleum exploration domain. In: **Proceedings of the XII Seminar on Ontology Research in Brazil and III Doctoral and Masters Consortium on Ontologies**. [S.l.: s.n.], 2019.

ABEL, M. et al. Lithologic logs in the tablet through ontology-based facies description. **AAPG Annual Convention and Exhibition**, 2012.

ABEL, M. et al. Ontologies and data models: essential properties and data modeling for petroleum exploration. **Foundations, the Journal of the Professional Petroleum Data Management Association - Volume 3, Issue 1**, 2016.

ARP, R.; SMITH, B.; SPEAR, A. D. **Building ontologies with basic formal ontology**. [S.l.]: Mit Press, 2015.

BABAIE, H. A. Ontological relations and spatial reasoning in earth science ontologies. **Societal Challenges and Geoinformatics, Geological Societa of America (GSA)**, Special Paper 482, 13-27, 2011.

BABAIE, H. A.; DAVARPANAH, A. Semantic modeling of plastic deformation of polycrystalline rock. **Computers & Geosciences**, Elsevier, v. 111, p. 213–222, 2018.

BABAIE, H. A. et al. Designing a modular architecture for the structural geology ontology. **Sepcial Papers - Geological Society of America**, Boulder, Colo.; Geological Society of America; 1999, v. 397, p. 269, 2006.

BAKER, L. R. **Persons and bodies: A constitution view**. [S.l.]: Cambridge University Press, 2000.

BAKER, L. R. **The metaphysics of everyday life**. [S.l.]: Cambridge: Cambridge University Press, 2007.

BORST, W. N. **Construction of Engineering Ontologies for knowledge sharing and reuse**. Thesis (PhD) — University of Twente, Enschede, The Netherlands, 1997.

BOUTEILLER, P. L. et al. A new conceptual methodology for interpretation of mass transport processes from seismic data. **Marine and Petroleum Geology**, Elsevier, 2019.

BOYD, T. **Ontology of Geological Mapping**. Dissertation (Master) — College of Arts and Sciences, Georgia State University, 2016.

BRITANNICA, T. E. o. E. **Stratigraphy**. 2014. Encyclopedia Britannica. Accessed 10 May 2021. Available from Internet: <https://www.britannica.com/science/stratigraphy-geology>.

BRITANNICA, T. E. o. E. **Petrology**. 2018. Encyclopedia Britannica. Accessed 10 May 2021. Available from Internet: <https://www.britannica.com/science/petrology>.

BRITANNICA, T. E. o. E. **Sedimentology**. 2018. Encyclopedia Britannica. Accessed 10 May 2021. Available from Internet: <https://www.britannica.com/science/sedimentology>.

BRITANNICA, T. E. o. E. **Structural Geology**. 2018. Encyclopedia Britannica. Accessed 10 May 2021. Available from Internet: <https://www.britannica.com/science/structural-geology>.

CARBONERA, J. L.; ABEL, M.; SCHERER, C. M. Visual interpretation of events in petroleum exploration: An approach supported by well-founded ontologies. **Expert Systems with applications**, Elsevier, v. 42, n. 5, p. 2749–2763, 2015.

CHANDRASEKARAN, B.; JOSEPHSON, J. R.; BENJAMINS, V. R. What are ontologies, and why do we need them? **IEEE Intelligent systems**, IEEE, n. 1, p. 20–26, 1999.

CICCONETO, F. **GeoReservoir: An ontology for deep-marine depositional system description**. Dissertation (Master's Disseration) — Federal University of Rio Grande do Sul, 2021.

CICCONETO, F. et al. A spatial relation ontology for deep-water depositional system description in geology. **Proceedings of the XIII Seminar on Ontology Research in Brazil and IV Doctoral and Masters Consortium on Ontologies**, 2020.

COHEN, K. M. et al. The ics international chronostratigraphic chart. **Episodes**, International Union of Geological Sciences, v. 36, n. 3, p. 199–204, 2013.

COMMITTEE, N. A. G. M. D. M. S. et al. Nadm conceptual model 1.0 - a conceptual model for geologic map information. **US Geological Survey Open-File Report**, v. 1334, n. 2004, p. 60, 2004.

COX, S. J.; RICHARD, S. A geologic timescale ontology and service. **Earth Science Informatics**, Springer, v. 8, n. 1, p. 5–19, 2015.

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**, Geological Society of America, v. 1, n. 3, p. 119–137, 2005.

EDISON, L. S.; BRANTLEY, J. D.; EDWARDS, S. The value of smarter oil and gas fields. **IBM Center for Applied Insights**, 2011.

EVNINE, S. J. Constitution and composition: Three approaches to their relation. **ProtoSociology**, v. 27, p. 212–235, 2011.

FÁVERA, J. C. D. **Fundamentos de Estratigrafia Moderna**. [S.l.]: EdUERJ, 2001.

FERNÁNDEZ-LÓPEZ, M.; GÓMEZ-PÉREZ, A.; JURISTO, N. Methontology: from ontological art towards ontological engineering. American Asociation for Artificial Intelligence, 1997.

FINE, K. Things and their parts. **Midwest Studies in Philosophy**, Wiley Online Library, v. 23, n. 1, p. 61–74, 1999.

FINE, K. et al. Acts, events and things. In: **Sixth International Wittgenstein Symposium, Kirchberg-Wechsel (Austria)**. [S.l.: s.n.], 1982. p. 97–105.

FORREST, P. The identity of indiscernibles. In: ZALTA, E. N. (Ed.). **The Stanford Encyclopedia of Philosophy**. Winter 2016. [S.l.]: Metaphysics Research Lab, Stanford University, 2016.

GANGEMI, A. et al. Sweetening ontologies with dolce. In: **Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web**. London, UK, UK: Springer-Verlag, 2002. (EKAW '02), p. 166–181. ISBN 3-540-44268-5. Available from Internet: <http://dl.acm.org/citation.cfm?id=645362.650863>.

GARCIA, L. F. et al. A conceptual framework for rock data integration in reservoir models based on ontologies. **International Journal of Monitoring and Surveillance Technologies Research (IJMSTR)**, IGI Global, v. 5, n. 1, p. 71–82, 2017.

GOLDBERG, K. et al. Re-sedimented deposits in the rift section of the campos basin. **Marine and Petroleum Geology**, Elsevier, v. 80, p. 412–431, 2017.

GÓMES-PÉREZ, A.; FERNÁNDEZ-LÓPEZ, M.; CORCHO, O. **Ontological Engineering: with examples from the areas of knowledge management, e-commerce and the semantic web**. Londres: Springer, 2004. 403 p.

GRUBER, T. R. The role of common ontology in achieving sharable, reusable knowledge bases. **KR**, v. 91, p. 601–602, 1991.

GRUBER, T. R. A translation approach to portable ontology specifications. **Knowledge Acquisition**, v. 5, p. 199–220, 1993.

GRÜNINGER, M.; FOX, M. S. Methodology for the design and evaluation of ontologies. In: **International Joint Conference on Artificial Inteligence (IJCAI95), Workshop on Basic Ontological Issues in Knowledge Sharing**. [S.l.: s.n.], 1995.

GUARINO, N. The ontological level: Revisiting 30 years of knowledge representation. In: BORGIDA, A. et al. (Ed.). **Conceptual Modeling: Foundations and Applications**. [S.l.]: Springer, 2009. p. 52–67.

GUARINO, N.; WELTY, C. A formal ontology of properties. In: SPRINGER. **International Conference on Knowledge Engineering and Knowledge Management**. [S.l.], 2000. p. 97–112.

GUIZZARDI, G. **Ontological Foundations for Structural Conceptual Models**. Enschede, The Netherlands: Universal Press, 2005. 410 p. (CTIT PhD Thesis Series, v. 05-74).

GUIZZARDI, G.; WAGNER, G. Ontologies and business systems analysis. In: ____. [S.l.]: IDEA Publisher, 2005. chp. Some Applications of a Unified Foundational Ontology in Business Modeling, p. 345–367.

GUTMAN, J.; RIBEIRO, M. J. G. **Plano de Dados Abertos**. [S.l.], 2019.

HEFLIN, J.; HENDLER, J. **Semantic interoperability on the web**. [S.l.], 2000.

HERRE, H. General formal ontology (gfo): A foundational ontology for conceptual modelling. In: **Theory and Applications of Ontology: Computer Applications**. [S.l.]: Springer, 2010. p. 297–345.

HYNE, N. J. **Nontechnical guide to petroleum geology, exploration, drilling, and production**. [S.l.]: PennWell Books, 2012.

JOHNSTON, M. Constitution and identity. **The Oxford Handbook of Contemporary Philosophy.**, 2005.

JOHNSTON, M. Hylomorphism. **The Journal of Philosophy**, v. 103, n. 12, p. 652–698, 2006.

KEET, C. M. **An introduction to ontology engineering**. [S.l.]: Maria Keet, 2018.

KOP, C. Domain expert centered ontology reuse for conceptual models. In: SPRINGER. **OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"**. [S.l.], 2011. p. 747–762.

KOSLICKI, K. **The structure of objects**. [S.l.]: Oxford University Press on Demand, 2008.

LORENZATTI, A. et al. Ontological primitives for visual knowledge. In: **Proceedings of the 20th Brazilian conference on Advances in artificial intelligence (2010)**. Sao Bernardo do Campo: Springer Berlin / Heidelberg, 2010. (Lectures Notes in Artificial Intelligence, v. 6404), p. 1–10.

MA, X. **Ontology Spectrum For Geological Data Interoperability**. Thesis (PhD) — University of Twnte, Faculty of Geo-Information Science and Earth Observation, 2011.

MANTOVANI, A.; LOMBARDO, V.; PIANA, F. Ontology-driven representation of knowledge for geological maps. **Computers & Geosciences**, Elsevier, p. 104446, 2020.

MILLER, A. Realism. In: ZALTA, E. N. (Ed.). **The Stanford Encyclopedia of Philosophy**. Winter 2016. [S.l.]: Metaphysics Research Lab, Stanford University, 2016.

MOTIK, B.; SHEARER, R.; HORROCKS, I. Hypertableau reasoning for description logics. **Journal of Artificial Intelligence Research**, v. 36, p. 165–228, 2009.

NEUHAUS, F. et al. Creating the ontologists of the future. **Applied Ontology**, IOS Press, v. 6, n. 1, p. 91–98, 2011.

NILES, I.; PEASE, A. Towards a standard upper ontology. In: ACM. **Proceedings of the international conference on Formal Ontology in Information Systems-Volume 2001**. [S.l.], 2001. p. 2–9.

NOMENCLATURE, N. A. C. on S. North american stratigraphic code. **AAPG Bulletin**, American Association of Petroleum Geologists (AAPG), v. 89, n. 11, p. 1547–1591, 2005.

NOY, N. F.; MCGUINNESS, D. L. et al. **Ontology development 101: A guide to creating your first ontology**. [S.l.]: Stanford knowledge systems laboratory technical report KSL-01-05, 2001.

OBERLE, D. **Semantic management of middleware**. [S.l.]: Springer Science & Business Media, 2006.

OBRST, L. Ontologies for semantically interoperable systems. In: ACM. **Proceedings of the twelfth international conference on Information and knowledge management**. [S.l.], 2003. p. 366–369.

PARTRIDGE, C. The role of ontology in semantic integration. In: **Second International Workshop on Semantics of Enterprise Integration at OOPSLA**. [S.l.: s.n.], 2002.

PARTRIDGE, C.; STEFANOVA, M. Building a foundation for ontologies of organisation. **The ontology and modelling of real estate transactions: European jurisdictions (international land management series)**, p. 141–149, 2003.

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

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

PRESTES, E. et al. Towards a core ontology for robotics and automation. **Robotics and Autonomous Systems**, Elsevier, v. 61, n. 11, p. 1193–1204, 2013.

RADEMAKER, A. et al. Extending sumo to geological times. **ONTOBRAS**, 2019.

RAINAUD, J. A short history of the last 15 year's quest for it interoperability in the petroleum e&p industry. **Oil & gas science and technology**, EDP Sciences, v. 60, n. 4, p. 597–605, 2005.

RASKIN, R. G.; PAN, M. J. Knowledge representation in the semantic web for earth and environmental terminology (sweet). **Computers & geosciences**, Elsevier, v. 31, n. 9, p. 1119–1125, 2005.

RICHARD, S. M. Geoscience concept models. **Special Papers-Geological Society of America**, Boulder, Colo.; Geological Society of America; 1999, v. 397, p. 81, 2006.

RODRIGUES, F. H. et al. An ontological model for turbidite channel migration. In: **Proceedings of the XII Seminar on Ontology Research in Brazil and III Doctoral and Masters Consortium on Ontologies**. [S.l.: s.n.], 2019.

SCHEKOTIHIN, K.; RODLER, P.; SCHMID, W. Ontodebug: Interactive ontology debugging plug-in for protégé. In: SPRINGER. **International Symposium on Foundations of Information and Knowledge Systems**. [S.l.], 2018. p. 340–359.

SCHERP, A. et al. Designing core ontologies. **Applied Ontology**, IOS Press, v. 6, n. 3, p. 177–221, 2011.

SINHA, A. K. et al. Towards a reference plate tectonics and volcano ontology for semantic scientific data integration. **Proceedings of Geoinformatics**, p. 43–46, 2007.

SIRIN, E. et al. Pellet: A practical owl-dl reasoner. **Journal of Web Semantics**, Elsevier, v. 5, n. 2, p. 51–53, 2007.

SMITH, B. Beyond concepts: ontology as reality representation. **Proceedings of FOIS, International Conference on Formal Ontology and Information Systems**, nov. 2004.

SMITH, B. On classifying material entities in basic formal ontology. **Proceedings of the Third Interdisciplinary Ontology Meeting**, 2012.

STUDER, R.; BENJAMINS, V. R.; FENSEL, D. Knowledge engineering: Principles and methods. **Data and Knowledge Engineering**, v. 25, n. 1-2, p. 161–197, March 1998.

SUÁREZ-FIGUEROA, M. C. **NeOn Methodology for building ontology networks: specification, scheduling and reuse**. Thesis (PhD) — Informatica, 2010.

SUÁREZ-FIGUEROA, M. C. et al. **Ontology engineering in a networked world**. [S.l.]: Springer Science & Business Media, 2012.

SURE, Y.; STAAB, S.; STUDER, R. Ontology engineering methodology. In: **Handbook on ontologies**. [S.l.]: Springer, 2009. p. 135–152.

THOMSON, J. J. The statue and the clay. **Noûs**, JSTOR, v. 32, n. 2, p. 149–173, 1998.

USCHOLD, M.; GRUNINGER, M. Ontologies and semantics for seamless connectivity. **ACM SIGMod Record**, ACM, v. 33, n. 4, p. 58–64, 2004.

VARZI, A. C. Vagueness. **Encyclopedia of cognitive science**, Wiley Online Library, 2006.

WASSERMAN, R. Material constitution. In: ZALTA, E. N. (Ed.). **The Stanford Encyclopedia of Philosophy**. Fall 2017. [S.l.]: Metaphysics Research Lab, Stanford University, 2017.

WINTER, W. et al. Bacia de campos. **Boletim de Geociencias da PETROBRAS**, v. 15, n. 2, p. 511–529, 2007.

ZHONG, J.; AYDINA, A.; MCGUINNESS, D. L. Ontology of fractures. **Journal of Structural Geology**, Elsevier, v. 31, n. 3, p. 251–259, 2009.

ZIMMERMAN, D. Theories of masses and problems of constitution. **The philosophical review**, JSTOR, v. 104, n. 1, p. 53–110, 1995.

APPENDIX A — RESUMO ESTENDIDO

Neste trabalho, propomos a ontologia GeoCore, uma ontologia core para Geologia que representa os termos utilizados que se referem as principais entidades ao longo de todo domínio. O objetivo da GeoCore é facilitar o desenvolvimento e integração de ontologias de domínio e modelos conceituais que dão suporte a sistemas de informação no domínio geológico.

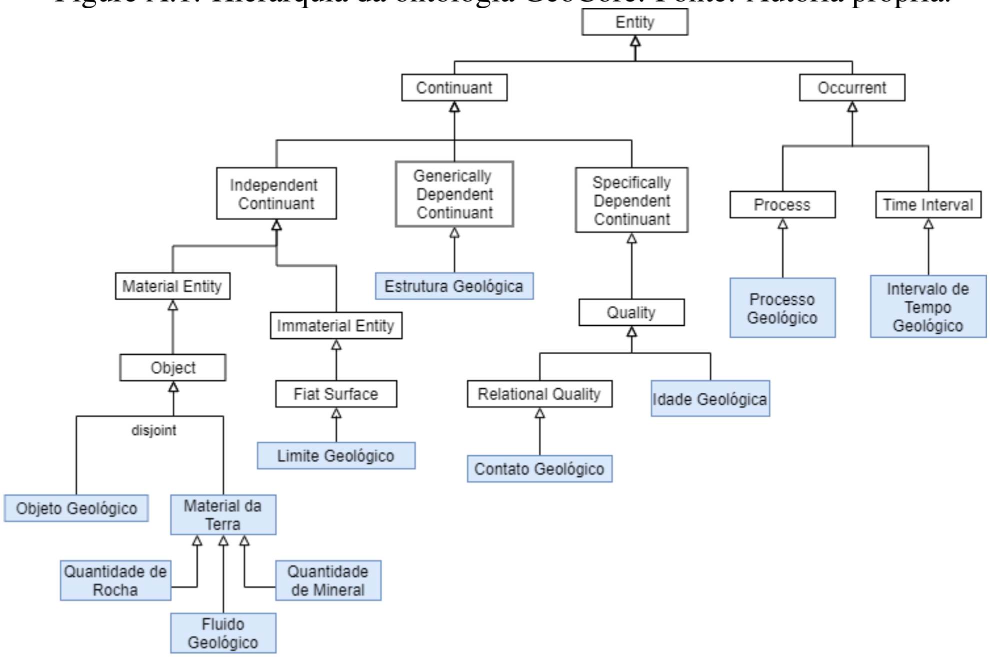
A GeoCore é o resultado de uma análise ontológica baseada em uma visão filosófica realista. Sendo assim, assumimos que as entidades observadas pelos geocientistas existem independentemente da visão ou crença de qualquer pessoa. Além disso, durante o desenvolvimento da ontologia GeoCore, consideramos também uma visão onde um objeto e a matéria que o constitui são entidades distintas. Para isto, nos baseamos em uma teoria filosófica de constituição material. Esta visão assume que entidades materiais podem co-existir no mesmo espaço-tempo quando estão em uma relação de constituição material. Esta visão nos permite tratar propriedades do objeto de propriedades da matéria de forma separada.

Utilizamos como ontologia de topo para o desenvolvimento da GeoCore a *Basic Formal Ontology* (BFO). Consideramos a BFO uma boa escolha para representar conhecimento do domínio geológico porque foi desenvolvida com o objetivo de descrever domínios científicos. Além disso, a BFO possui um conjunto de classes e axiomas para representação de entidades materiais que pode ser aproveitada em um domínio com Geologia.

Na GeoCore, definimos onze classes e duas relações - uma de geração, e uma de constituição - para representar conhecimento geológico. A figura A.1 apresenta a hierarquia de nossa ontologia com as classes da BFO que são especializadas.

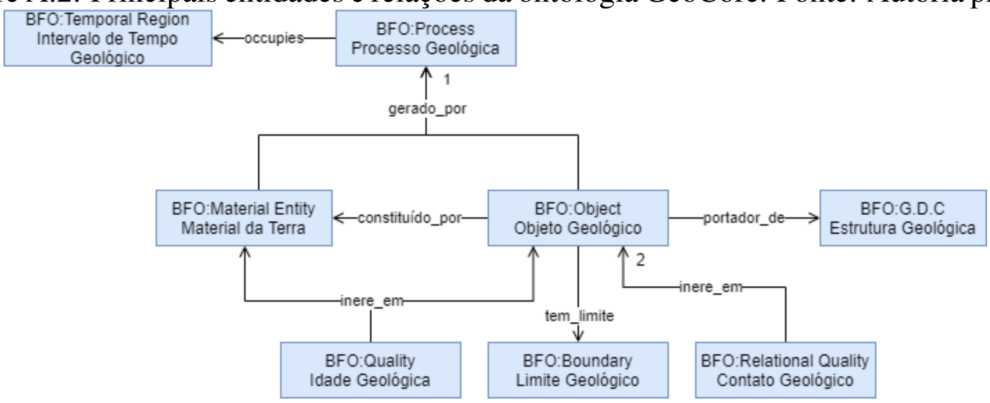
A figura A.2 resume as principais entidades e relações da ontologia GeoCore. A ideia central é de que em Geologia os geocientistas lidam com dois tipos de entidades materiais: materiais da terra e objetos geológicos. São entidades distintas, porém relacionadas. Objetos Geológicos são entidades materiais maximamente conectados, delimitados por descontinuidades físicas de matéria. Por outro lado, matérias da terra não necessariamente são maximamente conectados, pois não necessariamente possuem unicidade. Materiais da terra possuem uma relação de constituição com objetos geológicos. Tanto materiais da terra, quanto objetos geológicos são gerados por processos geológicos. A idade de um material da terra ou de um objeto geológico corresponde ao intervalo de

Figure A.1: Hierarquia da ontologia GeoCore. Fonte: Autoria própria.



tempo em que o processo que os gerou ocorreu. Além disso, objetos geológicos podem ainda ser portadores de estruturas geológicas e podem estar em contato físico com outros objetos.

Figure A.2: Principais entidades e relações da ontologia GeoCore. Fonte: Autoria própria.



Validamos nossa ontologia analisando as questões de competência criadas por especialistas, verificando aa consistência de nossa implementação com raciocinadores automáticos, e representando um caso de uso geológico com dados reais do grupo Lagoa Feia, na Bacia de Campos.

Este trabalho possui contribuições para o estado-da-arte tanto no nível de aplicação quanto no nível conceitual. No nível conceitual, desenvolvemos um artefato computacional que contém um conjunto de classes e axiomas para representação de conhecimento geológico. No nível conceitual, fizemos uma análise ontológica do domínio geológico com base em visões filosóficas e metodologias de engenharia de ontologias. Também propomos um padrão ontológico para a questão da constituição material no domínio da Geologia baseado em uma teoria filosófica de constituição.