Pontifícia Universidade Católica
do Rio de Janeiro

Patricia Ferreira da Silva

ResRiskOnto: an application ontology for risks in the petroleum reservoir domain

Dissertação de Mestrado

Dissertation presented to the Programa de Pós–graduação em Informática of PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Informática.

Advisor : Prof. Hélio Côrtes Vieira Lopes
Co-advisor: Dr. Rafael Jesus de Moraes

Rio de Janeiro
April 2022

Pontifícia Universidade Católica
do Rio de Janeiro

Patricia Ferreira da Silva

ResRiskOnto: an application ontology for risks in the petroleum reservoir domain

Dissertation presented to the Programa de Pós–graduação em Informática of PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Informática. Approved by the Examination Committee:

Prof. Hélio Côrtes Vieira Lopes
Advisor
Departamento de Informática – PUC-Rio

Dr. Rafael Jesus de Moraes
Co-advisor
CENPES

Profª. Simone Diniz Junqueira Barbosa
PUC-Rio

Dr. Régis Kruel Romeu
UFRGS

Rio de Janeiro, April 1st, 2022

Patricia Ferreira da Silva

Completed the B.Sc. in Electrical Engineering from Universidade Estadual de Campinas (UNICAMP) with double degree from École Centrale Paris in 2010. Attended the Petroleum Engineering Training from PETROBRAS in 2012, where she has worked since then as a reservoir engineer.

Bibliographic data

Silva, Patricia Ferreira da

ResRiskOnto: an application ontology for risks in the petroleum reservoir domain / Patricia Ferreira da Silva; advisor: Hélio Côrtes Vieira Lopes; co-advisor: Rafael Jesus de Moraes. – 2022.

109 f: il. color. ; 30 cm

Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática, 2022.

Inclui bibliografia

1. Informática – Teses. 2. Processamento em Linguagem Natural. 3. Modelagem Conceitual. 4. Ontologia. 5. Gerenciamento de riscos de projetos. 6. Reservatórios de Petróleo. I. Lopes, Hélio. II. Moraes, Rafael. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. IV. Título.

CDD: 004

To Juliano.

Acknowledgments

To all those people that, either by paving the way or by guiding me through shortcuts, made it possible for me to get closer to my *ikigai*.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

I would like to thank Professor Hélio for all the excitement with my work and for gathering, guiding and making available the potential of an amazing group of people: Fernando, William, Marisa, Flavia, Dayson, Rômulo, Pedro, João Vitor, Marcelo and many others. Professor Hélio and my colleagues, even during COVID times, were able to share their knowledge and point possibilities to the conduction of this work.

I am grateful to Rafael Jesus, for the generous guidance and for always asking the right questions. I will miss our Tuesday discussions.

A very special thanks to Régis, without whom this work would not have been possible. I extend this acknowledgement to all the people that I contacted through Régis: Professor Mara Abel, Luan Garcia, Fabrício Rodrigues and Guylerme Velasco.

I also thank Petrobras for providing the means necessary for me to accomplish this work. I would also like to thank Gabriel Serrão for the patience and support during almost two years, and to thank the aid of Cadu, Erico and Marcos Aurelio.

Finally, I send all my love to Zezé and Bela, my true cheerleaders. Thank you for showing interest even when I was repeating myself.

Abstract

Silva, Patricia Ferreira da; Lopes, Hélio (Advisor); Moraes, Rafael (Co-Advisor). **ResRiskOnto: an application ontology for risks in the petroleum reservoir domain**. Rio de Janeiro, 2022. 109p. Dissertação de Mestrado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

This work proposes the Reservoir Risks Ontology (ResRiskOnto), an application ontology for risks in the oil & gas industry associated with the petroleum reservoir domain. ResRiskOnto's building blocks are terms dominated by reservoir professionals, so that it can be easily adopted in future risk documentation.

ResRiskOnto is developed having at its center the concept of Risk Events. Each event has a set of possible Participants, that have its Characteristics manifested by the event. The ontology provides a total a set of 97 terms, 29 of which are derived from the Risk Event class.

To develop the ResRiskOnto, we conducted a semantic analysis of documents that contain over 2500 reservoir-related risks described in natural language. This repository is the result of hundreds of risk assessment workshops in oil & gas projects, conducted in over ten years in Petrobras.

This ontology is founded on the principles of the Basic Formal Ontology (BFO), a top-level ontology designed to describe scientific domains. One of BFO's most distinct characteristic is its commitment to Realism, a philosophical view of reality in which its constituents exist independently of our representations. On the domain-level, reservoir entities are described under the principles of the GeoCore Ontology, a core ontology for Geology.

To validate the ResRiskOnto we annotate our risk documents repository with the ontology's entities and relations, developing a model that recognizes named entities and extracts the relations among them.

Our contribution is an application ontology that allows semantic reasoning over the risk documents. We also expect to provide (i) a basis for data modelling in the case of reservoir-related risks; and (ii) a standard for future risk documentation in the reservoir domain.

Keywords

Natural Language Processing; Conceptual Modeling; Ontology; Project risk management; Petroleum Reservoir.

Resumo

Silva, Patricia Ferreira da; Lopes, Hélio; Moraes, Rafael. **ResRiskOnto: uma ontologia de aplicação para riscos no domínio de reservatórios de petróleo**. Rio de Janeiro, 2022. 109p. Dissertação de Mestrado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Este trabalho apresenta a Reservoir Risks Ontology (ResRiskOnto), uma ontologia aplicada aos riscos na indústria de óleo e gás associados ao domínio de reservatórios. Os componentes da ResRiskOnto são termos do domínio de trabalho de profissinais de reservatório, de forma a facilitar sua adoção na documentação futura de riscos.

A ResRiskOnto tem como ideia central o conceito de Evento de Risco. Cada evento tem um conjunto de possíveis Participantes, que por sua vez possuem Características manifestadas pelo evento. A ontologia dispõe de um total de 97 termos, 29 dos quais derivados da classe Evento de Risco.

Para desenvolver a ResRiskOnto, foi feita uma análise semântica em aproximadamente 2500 riscos de reservatórios documentados em linguagem natural. Este repositório é fruto de centenas de workshops de avaliação de riscos em projetos de óleo & gás, conduzidos na Petrobras durante uma década.

A ontologia proposta fundamenta-se nos princípios da Basic Formal Ontology (BFO), uma ontologia de topo projetada para descrever domínios científicos. A BFO baseia-se no Realismo, uma visão filosófica segundo a qual os entes que constituem a realidade existem independentemente da nossa representação. No nível de domínio definimos os entes de reservatório usando os conceitos da GeoCore Ontology, uma ontologia para a Geologia.

Para validar a ResRiskOnto os documentos do repositório foram anotados utilizando os entes e relações definidos na ontologia, e desenvolvido um modelo capaz de reconhecer entidades nomeadas e extrair as relações entre elas.

Nossa contribuição é uma ontologia aplicada que permite o raciocínio semântico no repositório de documentos de risco. Esperamos que ela forneça (i) as bases para modelagem de dados de riscos relacionados a reservatórios; e (ii) um padrão para futura documentação de riscos no domínio de reservatório.

Palavras-chave

Processamento em Linguagem Natural; Modelagem Conceitual; Ontologia; Gerenciamento de riscos de projetos; Reservatórios de Petróleo.

Table of contents

List of figures

List of tables

ADI – Análise Digital de Imagens
BIF – *Banded Iron Formation*

Miguilim, Miguilim, eu vou ensinar o que agorinha eu sei, demais: é que a gente pode ficar sempre alegre, alegre, mesmo com toda coisa ruim que acontece acontecendo. A gente deve de poder então ficar mais alegre, mais alegre, por dentro!

Guimarães Rosa, *Corpo de Baile.*

1
Introduction

"Get down the stairs!" — would a mother shout to her child, implicitly warning the danger of a fall. This is an example of a risk statement in natural language. While it is clear to the mother the possible outcomes of the scene she witnesses, those are not explicitly stated: the child's foot could miss a step, in which case she could end up by falling and bruising herself.

The expression of a *risk* in natural language can assume multiple forms, often concealing detailed information about the way an agent perceives the risk: what are the elements that could lead to the occurrence of a risk event? What are the possible consequences of the risk to the agent?

Risk constitutes a polysemous term, referencing concepts such as a triggering event to an undesired consequence — as for example, the uncertainties around an event, the dimension of the consequence, and so on.

In the industrial domain, risk management is often conducted along with project management workflows. Risk management practices can comprise activities for risk assessment, including (i) risk identification; (ii) risk analysis and evaluation; (iii) risk treatment and mitigation actions; and (iv) risk monitoring.

Such practices, however, can end up by producing many documents and reports elaborated by experts of a technical domain, who may not necessarily have been provided with adequate qualification on the *risk* domain. These documents containing text in natural language are unstructured data (as are video, audio and images). When this data is organized with marks that enforce some structure or hierarchies among elements, it becomes semi-structured. As opposed to unstructured, structured data is the one organized in a table format.

Unstructured and semi-structured data correspond roughly to 80% of the data in oil & gas industry (CHELMIS et al., 2013). Even though this type of data is inherently challenging in terms of information extraction and analysis, it is already well known that the information it retains can be of enormous economic potential (ITTOO; BOSCH et al., 2016).

1.1
Mining value from risk documentation

Over the last decade, Petrobras adopted the *Front-End-Loading* (FEL) methodology (SAPUTELLI et al., 2013) in the deployment of capital investment projects of the Exploration & Development (*upstream*) segment (MOTTA

IDENTIFICAÇÃO DE RISCOS DO PROJETO

Nome do Projeto:	[redacted]
GER/COORD:	[redacted]
CATEGORIA:	[redacted]
FASE:	[redacted]
Tipo de GR:	GR DE RESERVATÓRIOS
Disciplina:	Reservatórios
Coord. Revisor:	[redacted]

Identificação (*)	Título (*)	Tipo de Risco (*)	Impacto (*)	Probabilidade (*)	Severidade do Risco	Recomendação (*)
Risco	Produção de água nos novos produtores maior do que a prevista pelo caso "P50" devido à saturação de água acima da esperada.	Ameaça	Muito Alto	Média		Adquirir, processar, interpretar e incorporar nos modelos a Sísmica 4D. Identificar com os atributos sísmicos a posição espacial das janelas de comunicação entre as subzonas.
Risco	Comportamento da produção abaixo do previsto, principalmente nos médio e longo prazos, devido a algumas inconsistências nos modelos.	Ameaça	Baixo	Média		Rever os modelos geológicos e de fluxo: 1. Aplicar nos dados sísmicos 3D (2018) e 4D as novas técnicas de processamento, tais como como FWI, RTM e outras; 2. Investigar as anomalias 4D relacionadas ao aumento de saturação de água; 3. Definir a posição espacial das janelas de fluxo entre as subzonas do reservatório com base na análise de atributos; 4. Preservar no simulador de fluxo as relações Kv / Kh do modelo geológico; 5. Verificar a possibilidade de não realizar o "upscaling". Caso seja necessário fazer o "upscaling", preservar as heterogeidades verticais na área do projeto.
Risco	Projeto não estar otimizado em relação ao tipo e quantidade de poços.	Oportunidade	Muito Alto	Média		Avaliar o número e posicionamento dos produtores e injetor para garantir a otimização do projeto.
Risco	Fraturamento hidráulico da capeadora próximo aos injetores.	Ameaça	Muito Alto	Muito Baixa		Realizar microfrac no folhelho capeador do reservatório ou em regiões vizinhas (em poços a serem perfurados na região) para melhor definir as pressões limites de injeção.
Risco	Produção de água acima da prevista causada pelo posicionamento de algum produtor próximo à zona de falhas.	Ameaça	Alto	Baixa		Posicionar os produtores distante das falhas.
Recomendação	Reavaliar o modelo petrofísico.	N/A	N/A	N/A		a. Determinar a saturação de água por pressão capilar. Isso evita problemas com poços que foram perfurados após o início da produção do campo e possibilita a aplicação direta no modelo geológico. Utilizar dados do campo e, se necessário, análogos; b. Investigar as curvas de porosidade dos poços com RMN à luz dos conhecimentos atuais sobre essa ferramenta. Observou-se que, em pelo menos um poço, o perfil de porosidade do RMN apresentava valores maiores que os valores de pulgue da petrofísica básica. Essa investigação poderá resultar em novo modelo de porosidade (com possibilidade de maior volume poroso) ou, ao menos, na incorporação de incertezas na porosidade; c. Determinar as curvas de permeabilidade isentas da saturação de água. Investigar também as diferenças de permeabilidade entre as zonas de água e óleo.
Recomendação	Durante a implantação do projeto, perfurar os poços sequencialmente para avaliar as oportunidades de otimização.	N/A	N/A	N/A		Avaliar o desempenho inicial dos dois produtores antes de perfurar o injetor. Caso o desempenho inicial de um deles seja insatisfatório (alto BSW), avaliar sua conversão em injetor.

Figure 1.1: Risk Report (for the reservoir domain) in an oil& gas project

et al., 2014). Project approval is conducted in phases with defined objectives, and each phase contains workflows and processes in which requirements are fulfilled, providing elements to measure the project maturity and to deliver the necessary information for decision support.

Entangled with project management rituals are the assurance procedures, including workshops for project risk identification. In the last ten years, petroleum reservoir specialists in Petrobras have conducted hundreds of workshops of project-phase risk identification, documenting over 2500 risks in *upstream* projects associated with their domain of knowledge.

Figure 1.1 shows an example of a piece generated during a workshop for reservoir risk identification. Each risk is described in a sentence, and is subject to an evaluation of its probability and impact. For each risk, the experts provide a set of recommendations for the reservoir team. Those recommendations are part of the risk management strategy for the project.

The probability of risk occurrence and the impact that it represents are evaluated by expert using a categorical scale (with values "very high", "high", "medium", "low" and "very low"), and are used to compute risk severity, which is the parameter used to quantify the risk. More detail on risk definitions is provided in Section 2.2.

From workshops that took place between 2011 and 2020, a set of approximately 2500 sentences describing risks was collected. This repository[1] poses a challenge to domain specialists, that wish to identify similarities among risks from different projects, so that they can evaluate and address the main challenges in oil & gas projects that are associated with the reservoir domain.

When expressed in natural and non-standardized language, the idea of risk in a given domain can be documented in a multitude of ways, often concealing implicit definitions. For this reason, we believe that a proper conceptual model over risk events can enable semantic reasoning, enhancing natural language processing algorithms applied to risk documentation.

This work proposes ResRiskOnto, an application ontology that provides proper conceptualization of risks in oil and gas projects documented by specialists of the Petroleum Reservoir domain. It offers domain experts a predefined set of concepts that resonate with their area of expertise. We expect that the result, an applied risk ontology composed of the words dominated by reservoir professionals, can be easily adopted in future risk documentation.

ResRiskOnto is developed as an extension of the BFO and GeoCore ontologies, and uses the Common Ontology of Value and Risk as conceptual basis for risk descriptions. It also reuses concepts developed in the GeoReservoir Ontology and Information Artifacts Ontology. ResRiskOnto characteristics have the advantage of facilitating ontology management and reuse.

The proposed ontology is freely available. To the best of our knowledge, there are no public relevant available conceptual models targeting risk documentation in natural language, which usually remain as corporate intellectual property.

The main goal of our work is to clarify and model concepts through a common understanding of reservoir risks in projects of the oil & gas industry, allowing the development of algorithms to risk processing, including semantic reasoning. We also aim to make available named entity recognition models that support such natural language processing tasks.

1.2 Text Structure

This document is structured as follows:

1. In Chapter 2 we present the Risk domain, providing an analysis of Risk Management techniques and the effort documented in the literature in defining and standardizing the concept of *risk*.

[1] In this work, the expression "risk corpus" is used to refer to the aforementioned repository of risk sentences.

2. In Chapter 3 an overview on the domain of Petroleum Reservoir is provided.

3. Chapter 4 brings up Ontology definitions and previous related work.

4. Chapter 5 describes the development and documentation of ResRiskOnto, the ontology that is the main contribution of this work and that is freely available in a public repository [2].

5. In Chapter 6 the experiments conducted for ontology validation are described.

6. In chapter 7, we present our conclusions, taking into consideration the competency questions elaborated to define the scope of ResRiskOnto. Possible improvements and future works are also appreciated.

[2] https://github.com/patriciaferreiradasilva/resriskonto

2
Risk Management and Risk Definition

According to the Cambridge Dictionary[1], risk is *the possibility of something bad happening*. This definition can evoke a wide range of meaning, according to the context in which it is applied: in the health domain, risk could be the possibility of a disease affecting the well-being of a patient; in the environmental domain, the possibility of a forest burning as a result of lightning can be evaluated as a risk.

In the industrial domain, efforts have been directed towards narrowing the definition of risk, making explicit what elements organizations should measure and control in order to manage their risks. In this chapter, we discuss some of the proposed definitions of risk, as well as common practices of industrial risk management.

2.1
Risk Management

Risk research began after World War II (DIONNE, 2013), with the first academic books in risk management being published in the 1960s, along with the development of systematic studies and technological models on the subject. The adoption of risk management as a corporate activity began in the 1990s, mainly in the financial sector.

In the industrial domain, important guidelines for risk management were established in the late 2000s (ISO Central Secretary, 2018). Organizations then became able to integrate risk management into significant activities, including decision-making.

Risk management deploys technological control measures in order to avoid the occurrence of a damage, or to mitigate its effects (MORAES, 2013). It aims to protect the people, the environment, and the assets and economical results of a human activities (AVEN, 2010b). Its process can be customized to better fit different applications within the organizations.

Figure 2.1 show a generic view on the main elements of the risk management process. Its core activities constitute a goal-oriented process with a sequence of iterative steps: risk identification, analysis, evaluation and treatment — the former three steps being referred to as *risk assessment*.

The solutions proposed in our work are directed towards the challenges that reservoir experts in Petrobras face when contributing to the risk assess-

[1]https://dictionary.cambridge.org/

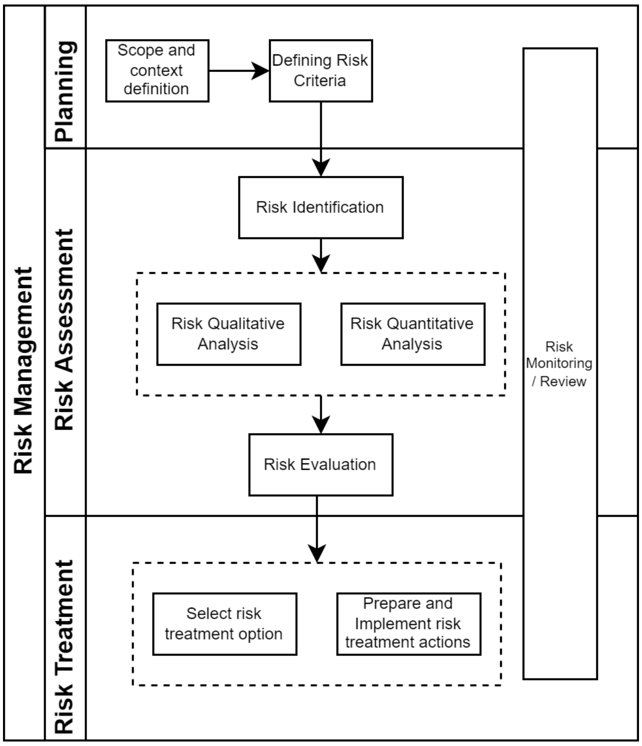


Figure 2.1: Elements of Risk Management Process, adapted from (ISO Central Secretary, 2018)

ment of oil & gas projects. For this reason, in the following subsection we present risk assessment and provide more detail on each of the steps that make up this activity in the risk management process.

2.1.1 Risk Assessment

Risk assessment is "the scientific process of defining the components of risk in precise, usually quantitative terms" (RENN, 1998). This activity is conducted collaboratively, merging technical knowledge with the managerial view of the organization's goals. The steps of risk assessment, briefly presented in this work, are detailed in (ISO Central Secretary, 2018):

Risk Identification

Comprises the description of risks, taking into consideration its causes and events, sources (either tangible or intangible), threats and opportunities, vulnerabilities and capabilities, context, consequences and the harm they pose to assets.

Risk Analysis

Activity of detailing risk uncertainties and the likelihood of multiple sources in different scenarios that could trigger events causing multiple consequences that affect objectives.

Risk analysis can be quantitative and qualitative, depending on the context and objectives of the risk management process. It takes into consideration the likelihood of events, the magnitude of consequences, context complexity, time-related factors, confidence levels and the pre-existence and effectiveness of control and mitigation actions.

This step on risk assessment is highly subject to human biases, and it usually demands a collaborative effort to converge different opinions on the subject aspects of risk into a well-defined set of assumptions and parameters evaluated using established techniques. Because of multiple opinions and perceptions on risk, combined with different levels of aversion to risk of the agents involved in risk analysis, this activity can be very demanding and time consuming. Consensus is reached after a detailed evaluation of the technical aspects of the activity being executed, and a broad debate on each analyst's judgement of risk.

Risk analysis serves as input to risk evaluation, and to the definition of the appropriate control techniques to be deployed as risk treatment strategies. It is a crucial activity that unfolds the elements that will be taken into account in decision making.

Risk Evaluation

Risk evaluation is the comparison between risk analysis and pre-determined risk criteria, supporting decisions that can affect or even interrupt projects and activities. Possible risk-related decisions are: accepting the risk, taking risk treatment actions, mitigating risk consequences, deepening risk analysis to a better understanding, or reconsidering objectives.

Decisions usually take into account the utility of the organization's objectives, or the threshold between its benefits and possible harms. This step often involves higher organization levels in decision making and communication.

Due to the complexity of oil & gas projects, each area of technical expertise may conduct its own risk management activities. In the last decade, petroleum reservoir specialists in Petrobras have conducted hundreds of workshops of project-phase risk assessment.

As shown in Chapter 1, our document repository is composed of one of the documents that are produced in risk assessment workshops of the reservoir domain in Petrobras. We collected hundreds of documents, that have in common the following information:

- risk description;

– risk impact;
– risk probability;
– suggested actions for risk treatment (or "recommendations").

The first item, the risk description, is a sentence in natural language that describes the risk. It satisfies the activity of *risk identification* described above. If we consider that each sentence constitutes one risk, reservoir specialists have documented over 2500 risks in *upstream* projects associated with their domain of knowledge over the last decade. Those risk sentences, as will become clear, will be subject to algorithms of natural language processing to validate our ontology.

The items *impact* and *probability* compose the quantitative analysis of the risk, satisfying the *risk analysis* activity. The analysis serves as input for decision-making over the treatment options for each risk.

The *recommendation* is yet another text in natural language, in which some suggested actions for risk treatment are described. Recommendations satisfy the *risk evaluation* step.

Because of its size and characteristics, our risk repository is rarely analysed as a whole. The knowledge it holds, however, is considered to be of much value by reservoir technicians in Petrobras. This is due to the fact that the content of the documents reflects the effort of some of the most skilled experts into analysing technical aspects of oil & gas projects and converging to a set of risk descriptions for each project.

A common complaint among those technicians is the lack of standardization over the risk concepts and guidelines on how to properly write risk descriptions. Because risk is a polysemous and somewhat unclear concept (PATT; SCHRAG, 2003), the description of risks in natural language may carry implicit knowledge on the causes and uncertainties that lead to the hazards posed by a specific risk experience. For that reason, the documents that result from risk management practices often express a project's risks in an incomplete fashion.

We believe that the challenge to unlock this knowledge and make it available for Petrobras' reservoir community can be overcome with the organization of the risk sentences according to a proper definition. For this reason, in the next subsection we present a literature review on risk definitions.

2.2 On risk definitions

Historical attempts to elaborate sound risk definitions walk hand-in-hand with risk management studies, being a topic of research for at least the last fifty years. Successful applications of the risk management framework depend on an explicit and accepted definition of the term "risk" (FISCHHOFF; WATSON; HOPE, 1984) (AVEN; RENN, 2009b). Our work, one of proposing proper ontological risk definitions, hopefully can contribute to enhance risk management on the oil & gas domain

Numerous discussions over the ontological and epistemological status of risk lead to definitions, that can be either quantitative or qualitative. (AVEN; RENN; ROSA, 2011) synthesizes eleven interpretations of risk, divided into three categories, namely (a) risk as a concept based on events, consequences and uncertainties, (b) risk as a modelled, quantitative concept and (c) risk measurements.

Qualitative definitions enlighten the polysemous aspect of the word risk, that can refer to events, consequences and their uncertainties — or the first of the three categories above. Certain definitions can emphasize one of these semantic dimensions of risk, but it usually gravitates around the idea of risk as an event or as the event's consequences, provided that events and consequences are subject to uncertainties.

Among those definitions are:

- Risk is the effect of uncertainty on objectives (ISO Central Secretary, 2009);
- Risk is a situation or event where something of human value is at stake and the outcome is uncertain (ROSA, 1998);
- Risk comes from a situated cognition relating a *risk object* and an *object at risk*, the former subject to certain circumstances (BOHOLM; CORVELLEC, 2011);
- Risk is an event where the outcome is uncertain (AVEN; RENN, 2009a).

Quantitative definitions cover the aspect of the uncertainty that is inherent to a risk, usually making use of probability as a measure of uncertainty. These are definitions that try somehow to quantify events consequences and their probabilities.

(AVEN, 2010a) proposes a formulation according to which the risk is represented as the following tuple Risk = (A, C, P), where A are the events (scenarios, cause events), C represents the events' consequences and P are the associated probabilities, defined to an ensemble of case-scenarios according to confidence intervals, frequency of occurrence or even a factor defined by a specialist.

Still in the quantitative views on risk, (KAPLAN; GARRICK, 1981) suggests the formalization of risk as a tuple $Risk = < s_i, p_i, x_i >$, in which s_i is a scenario and p_i its probability, while x_i measures the consequences.

The most commonly adopted quantitative definition of risk is one the combines the magnitude of the damage with the probability of the event (RENN, 1998).

Some of the above mentioned definitions emphasize the *consequence* dimension of the risk. Risk outcomes are well known to be potentially negative or positive — case in which risks are usually referred to as "opportunities". Companies are as interested in maximising opportunities for gain as in minimizing risks (LOOSEMORE et al., 2012). However, because of the more common association of risks to negative events, the concept of opportunities will not be explored in our work.

Despite all these definitions we may still face risks that would be considered to be "impossible", "unthinkable", or even "irrelevant". Take as an example the Fukushuma Daiichi nuclear disaster in Japan, the Macondo accident and, more recently, the COVID-19 global pandemic — all of which were considered impossible prior to the moment they occurred. A proper risk conceptualisation, one through which we name and define risk events, may present itself as an ideal perspective to assess and manage risk events (AVEN; KROHN, 2014).

Beyond the quantitative and qualitative approaches cited above, recent works focus on modelling the concept of risk. Conceptual models try to handle the polysemous and somewhat confusing aspect of risk, most of the times relying on proper tools designed for the task of conceptualisation.

Offering a customised language for risk modelling, (LUND; SOLHAUG; STØLEN, 2010) proposes a model driven method aimed at asset protection. On the software development domain, (ASNAR; GIORGINI; MYLOPOULOS, 2011) proposes a goal-oriented risk modeling and reasoning framework, in a three-layered fashion (asset, event, and treatment). Also in software engineering, (SIENA; MORANDINI; SUSI, 2014) presents a framework for risk modelling and risk evaluation tailored for open source software adoption.

Describing risk in terms of events and their causes, (SALES et al., 2018) proposes an ontology for the risk domain according to an experiential perspective and a relational perspective. In Section 4.2.3, the experiential perspective of risk brought by this work was the main guide to instantiate an applied ontology to the petroleum reservoir risk scenario — since our work will be centered in defining *risk objects* and *risk events*.

It is important to point out that instead of seeking an universal definition for reservoir-related risk, the goal of this work is to express reservoir-related risk events in a proper fashion. In order to do so, first we will analyse the intricate relationship between uncertainty and risk.

2.2.1
Uncertainty, complexity and risk

From the previous sections, we note that the key element distinguishing a harmful event from a risk is the uncertainty that characterizes risk events.

The adoption of probabilistic approaches to deal with risk uncertainty was proposed in the nuclear industry since the late 70's (APOSTOLAKSI, 1978). In the 90's, the National Research Council in the United States pointed out to the lack of scientific data quantifying health exposure phenomena (COUNCIL et al., 1994). This led to the Food Quality Protection Act, widely spreading the use of probabilistic tools for risk assessment in the food industry.

Specifically for the oil and gas industry, the activity of oil production forecasting constitutes an inherently risky activity. Usually conducted by petroleum reservoir professionals, production forecasts have as input uncertain parameters of the geological conditions that generated the reservoir, with the ultimate goal of subsidizing investment and operational decisions. The adoption of probabilistic approaches for uncertainty assessment is a recurrent practice in this domain as well.

It is important to notice the peculiarities of each application domain when evaluating the suitability of a defined set of risk management strategies. (AVEN, 2010b) introduces four categories of risk problems, having as characteristics their *simplicity*, *complexity*, *uncertainty* an *ambiguity*.

Simply put, simple risk problems have low complexity and few uncertainties and ambiguities (which does not imply that they have low impact). Complex risk problems are those of which the multitude of causal agents and effects lead to difficulties in identifying cause-effect links. This can be true either due to synergies of the agents, long periods between cause and effect, or even intervening variables. Uncertain risk problems present low predictability, sometimes as a result of modelling inaccuracies or inadequate reduction of complexities. And finally, ambiguous risk problems are subject to different interpretative and normative views.

Reservoir-related risks are highly complex and concern a great range of uncertainty, for reasons that will be discussed in the next chapter, in which we approach the nature of the petroleum reservoir activity.

In this work, we believe that a proper conceptual model is a powerful tool for dealing with the *complex* aspect of reservoir-related risks, because it clarifies the causal agents and risk effects. A better understanding of risk and risk complexity can improve uncertainties, especially those related to complexity reduction in risk modelling — thus enhancing risk predictability. Ultimately, proper conceptualization could help enhancing risk management responses.

In the next chapter we present the Reservoir domain, highlighting aspects of reservoir modelling and production forecasting that show how this is a rather complex activity.

3
An overview on the Petroleum Reservoir domain

In this chapter, we present an overview on the Petroleum Reservoir domain. In the first section we explore what is a Petroleum Reservoir and its role in the oil & gas exploitation. In the second section we present reservoir characterization, which is one of the main activities of reservoir specialists.

3.1
Petroleum Reservoir and Oil Exploitation

In this section we introduce the Petroleum Reservoir Domain (BAKER; YARRANTON; JENSEN, 2015) (WHEATON, 2016). Most of the definitions expressed are consonant with the adopted terminology for our ontology's classes (Section5.4). The terms that compose our ontology had their meanings documented by an expert, using definitions based on Petrobras' Glossary (PETROBRAS, 2007), Schlumberger Oilfield Glossary (SCHLUMBERGER, 2019), the Open Wordnet (TESSAROLLO; RADEMAKER, 2020) and Wikipedia.

The substance commercially referred to as petroleum is in fact a hydrocarbon (or a mixture of hydrocarbons), which is an organic chemical compound constituted mainly by molecules composed by carbon and hydrogen atoms, that can eventually contain oxygen, nitrogen and sulfur atoms. Petroleum is a naturally occurring fluid generated by some geological process and derived from organic material. It persists in nature either as a solid, liquid (as in Figure 3.1) or a gas — the two latter, known commonly as "oil" and "natural gas", respectively, are the focus of our work.

Contrary to popular belief, oil is not found in underground caverns. In

Figure 3.1: A sample of petroleum in liquid form (oil) (PETROBRAS, 2022)

fact, oil and gas are contained within microscopic pores of rock formations. Rocks are a solid constituted by an aggregate of particles, either made of mineral matter or material of biological origin. Accumulations of particles are created by the action of wind and water, and when those deposits are buried to a certain depth for enough time, they consolidate into rocks. Because of the gradual accumulation of unconsolidated material, buried deposits are called sedimentary layers.

A layer is composed of a rock as a distinct segment in a vertical stack of formation sequences, often with areal extent. After buried, a layer can participate in many geological processes (including processes generated by tectonic forces). This may result in tilting, folding, fracturing and to the generation of traps.

Hydrocarbons migrating from more deeply buried rocks migrate upwards with time, displacing water, and can eventually be trapped in a porous layer. One or more adjacent porous layers that store hydrocarbon and water within its pores constitute a reservoir. Overlying the reservoir, a non-permeable rock layer known as caprock acts as a top seal to fluid flow.

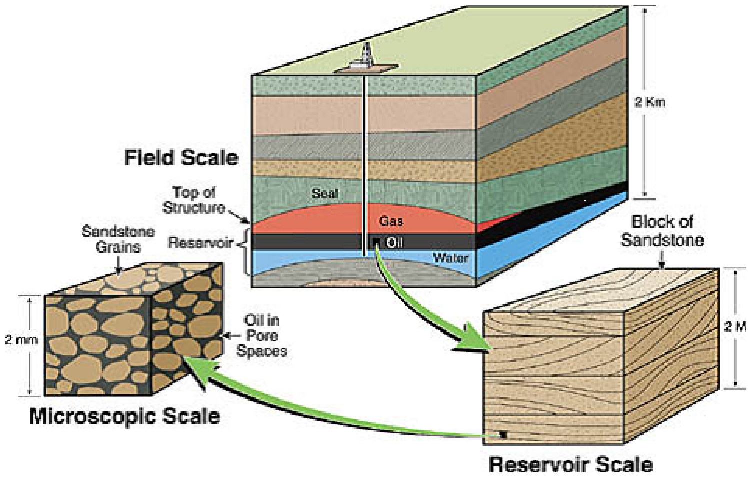


Figure 3.2: Schematic view of a petroleum reservoir in multiple scales (ZITHA et al., 2011)

Figure 3.2 shows a schematic view of a reservoir in different scales: in the petroleum field scale, we see the reservoir and its fluids; in a reservoir scale, the image illustrates the layered fashion in which it is formed; in a microscopic scale, we see an schematic view of the mineral grains that compose the rock, along with the oil stored in its pores.

A reservoir with enough porosity and permeability can have its hydrocarbons extracted. Porosity is the ratio between the volume of the empty spaces (or pores) and the total volume of reservoir rocks. Permeability is the in-

	Gases			Oils		
	Dry gas	**Wet gas**	**Gas condensate**	**Volatile oil**	**Black oil**	**Heavy oil**
Characteristics	-			very low viscosity	nearly constant viscosity and compressibility	high viscosity
Phase behavior (reservoir)	gas	gas	gas with liquid dropout as pressure decreases	initially oil, with significant phase variations as pressure changes	initially oil, can form a gas phase as pressure decreases	oil
Phase behavior (wellbore to surface)	gas	gas with liquid dropout as pressure decreases	mixture of gas and liquid	mixture of gas and liquid	mixture of gas and liquid	in some cases, is initially immobile
API	-	>45	45 - 60	42 - 55	15 - 45	<15
Solution GOR (scf/stb)	-	>30,000	3,500 - 30,000	900 - 3,500	200 - 900	<200

Table 3.1: Summary of fluid types in a reservoir. Modified from (BAKER; YARRANTON; JENSEN, 2015).

terconnectivity between the rock's void spaces, or pores, and determines the capability of the fluid to flow through the reservoir.

To produce oil, wells are drilled through the reservoir and properly equipped to extract fluids. Fluid flow happen in the wellbore because its pressure is lower than the pressure in the reservoir. According to the hydrocarbon mixture at reservoir conditions, the fluid type can be classified according to six different types: dry gas, wet gas, gas condensate, volatile oil, black oil and heavy oil.

Table 3.1 summarizes some of the characteristics of each fluid type in a reservoir — including API gravity and solution Gas/Oil Ratio (GOR). API gravity is, a measure of how heavy a petroleum is compared to water. The solution GOR expresses the amount of gas (in cubic feet) dissolved per barrel of oil, that comes out of solution at surface conditions.

As the fluids are drained and reservoir pressure declines, natural drive mechanisms may provide the necessary energy to the production. The mechanism of *undersaturated oil expansion* happens when the fluid is in oil phase, and is a combination of rock compaction and fluid expansion as the pressure decreases. When the oil reaches its bubble point, gas starts to be liberated and the main drive mechanism is *solution gas drive*, a result of gas expansion. If enough gas is liberated to segregate in a gas cap, the *gas cap expansion* becomes the drive mechanism. Other possibilities arise when the reservoir contains an aquifer, that is an underlying portion of the formation fully saturated with water. If there's significant water influx, *water drive* mechanism acts.

Natural drive mechanisms usually account for up to 10-15% of the total volume of oil, in oil reservoirs. Except in the cases of strong gas cap and water mechanisms, production rapidly declines and it is necessary to implement production strategies to obtain more oil from the reservoir. Techniques for additional recovery comprise the injection of fluid of low value, in order to maintain the pressure in a reservoir.

Injection well are wells properly equipped to inject fluids in a reservoir.

Waterflooding is the injection of water, positioning injection wells in order sweep oil towards production wells. *Immiscible and miscible gas flooding* are the injection of a gas forming, respectively, a multi-phase or a single phase fluid along with the oil. The injection of chemicals, steam, polymers and other types of fluids is also possible.

In the surface, fluids drained from a reservoir are separated into oil, water and gas streams. The oil is then destined treated in order to have its water and impurities content under optimal conditions, before being destined to a refinery. Part of the gas can be used in the facilities as fuel to generate energy, and it can also be re-injected in the reservoir or be used in artificial fluid elevation (gas lift). The remaining gas may be compressed and exported through a pipeline to a processing plant. The water is treated to have its oils droplets removed, and can be either reinjected in the reservoir or discarded.

Production facilities, along with the set of wells, constitute the main elements of the exploitation strategy of a reservoir (that also comprise operational parameters such as the volume of injected fluids, well geometry, borehole pressure, etc). A petroleum company will invest its capital in those assets, provided that it will increase the company's value for its stakeholders. Reservoir engineers define some characteristics of a development strategy, such as number, type and position of wells, the spacing between wells, the fluids to be injected into the reservoir and its rates, artificial elevation strategies. Figure 3.3 shows a schematic view of a field exploitation strategy.

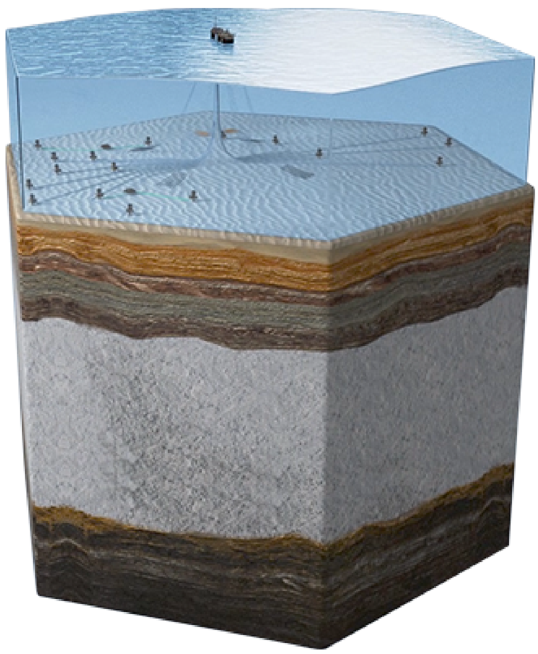
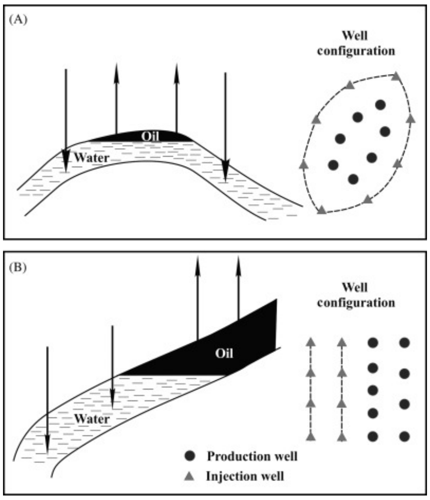


Figure 3.3: Schematic view of production facilities, well heads and pipes (PETROBRAS, 2022)

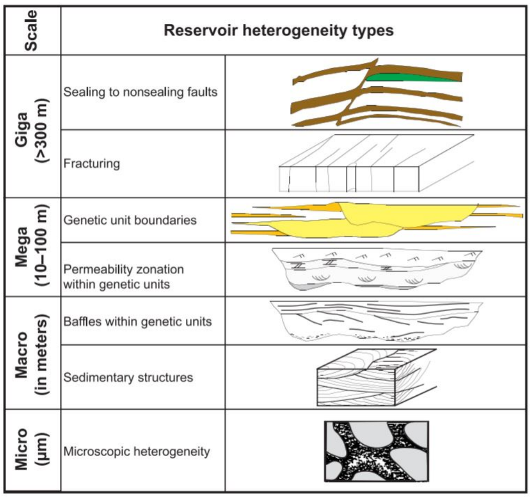
To maximize the value associated to an exploitation strategy, a reservoir engineer take into account the amount of investments and the corresponding volume of oil that can be produced in a given period of time. The reservoir's drive mechanism, along with the fluid type and reservoir architecture, are the main parameters taken into account in the selection of an exploitation strategy,

Many factors determine what we call the reservoir architecture. Materialized in the form of a rocky bed in the subsurface, the reservoir is known through indirect data (seismic and well logs) and well samples. Further discussion on reservoir data and the associated uncertainties in conducted Subsection 3.2. Geologists infer from the very sparse data available the geometry and architecture of the reservoir.

The external geometry of the reservoir, whether it is a flat, dome-shaped, or tilted body of rock, is a major factor to determine well positioning. This effect is illustrated in Figure 3.4a. Fluid flow behavior, however, will be very strongly influenced by reservoir heterogeneities.



(a) Production-injection well arrangements for different reservoir geometries



(b) Heterogeneity types according to the scale in a sandstone reservoir

Figure 3.4: Reservoir architecture aspects: (a) external geometry (VISHNYAKOV et al., 2019); (b) heterogeneities (MORAD et al., 2010).

Physical and textural variations of rocks within a reservoir come out as vertical and lateral variations in reservoir's porosity, permeability and capilarity (ALPAY, 1972). Since fluid will flow preferably through the path of least resistance, variation of characteristics within a reservoir may lead to a nonuniform oil recovery. Reservoir heterogeneities play a major role on fluid displacement, oil recovery and, consequently, in the economic success of an exploitation strategy (WEBER, 1986).

Figure 3.4b show possible heterogeneities ranging from micrometers to hundreds of meters in a sandstone reservoir. Heterogeneities are a result of geological conditions overcome by a reservoir, and reflect variations in:

– Depositional facies: facies is the pattern that becomes concrete in the internal arrangement and properties of the rock. Different environmental conditions during mineral deposition result in different depositional facies within the reservoir;

– Diagenetic evolution pathways: diagenesis are physical and chemical processes that may alter the sediments, after they have been deposited to the moment of their consolidation;

– Structural features: tectonic forces can lead to rock displacements, resulting in a topological mechanical discontinuity in the reservoir.

To select the ideal exploitation strategy, reservoir professionals invest their efforts in reservoir characterization. *Reservoir characterization is the mental or mathematical model representation of the reservoir based on reservoir data* (BAKER; YARRANTON; JENSEN, 2015).

A reservoir specialist will use the reservoir characterization as a tool that, combined with his applied knowledge, allows the exploration of possibilities for well locations, fluid rates and other parameters that make up for the optimal reservoir exploitation.

Proper reservoir characterizations will reveal the best positioning of wells, help select advanced recovery techniques, indicate the flow behavior within the reservoir. At the very last, good characterizations generate accurate production forecasts, within an acceptable range of uncertainty.

The activity of reservoir characterization will be described in the next subsection.

3.2
Reservoir Characterization and Uncertainty

To maximize hydrocarbon recovery at a minimum cost, reservoir professionals seek the best comprehension of the reservoir's rocks and heterogeneities, fluid distribution and dynamics. Reservoir comprehension is the result of the engagement of geologists, geophysicists and engineers, who gather and interpret data, building conceptual, analytical and mathematical models to characterize the reservoir.

In this chapter we introduce reservoir characterization in a nutshell (BAKER; YARRANTON; JENSEN, 2015). Our goal is to illustrate how uncertainties interfere with the objectives of reservoir professionals (MA; POINTE et al., 2011), i.e., optimize hydrocarbon resource extraction.

After the discovery of a hydrocarbon accumulation, mental and conceptual models of geologic aspects of the reservoir are confronted and complemented with data, which are at the beginning of reservoir characterization. Data acquisition occurs at many stages of reservoir exploration and development, and are crucial to reservoir characterization. Main reservoir data sources are:

- Seismic surveys;
- Core samples;
- Well logs;
- Fluid samples;
- Pressure data;
- Flow data.

Seismic acquisition involves subjecting the area to a seismic energy source, usually percussion waves. This energy travels along geologic layers and are reflected back to seismic receivers (or geophones). The seismic data represents the response of the wavefield density contrasts in the interfaces of geologic layers.

Seismic data, combined with previous geological conceptual modeling, subsidizes the decision to drill exploration wells, that aim to confirm or not the reservoir existence. Wells are the source of direct measures of reservoir properties, through the extraction of rock samples (core samples), fluid samples, and production profile (pressure and flow) data. Indirect measures can also be realized via well logging.

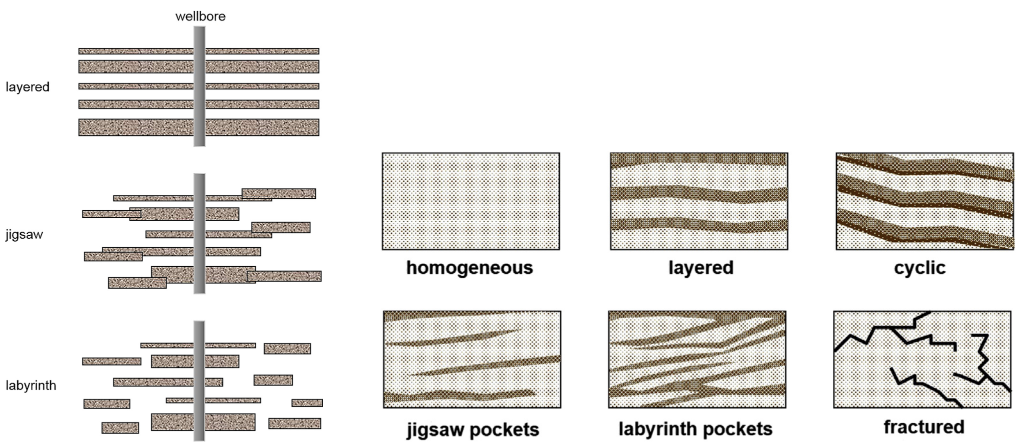
Core is a cylindrical section of rock extracted using special perforation tools in a well. The core is sampled and enables laboratory analysis of porosity, permeability and fluid saturation measures, as well as rock description (mineralogy and facies). Core samples are scarce due to the high costs of extraction. Few wells are sampled, while most (if not all) wells are logged.

Well logging is an activity conducted after well perforation, and consists in measuring physical quantities at different reservoir depths. Gamma ray, electrical and electromagnetic, acoustic, radioactive and magnetic responses are measured using tools attached to a wireline inserted into the well. These properties are used as indirect measures of rock and fluid properties. Pressure can also be measured along various reservoir depths with adequate logging tools.

Once the well is drilled and properly equipped, it is possible to measure the reservoir response to fluid flow. Fluid flow happens due to pressure contrast between the wellbore and the formation, and continuous rates reflect on pressure drawdown within a fixed depth in the well formation. When the flow is interrupted, pressure builds up gradually, due to fluid and rock compressibility and reservoir dynamics. The opposite happens if we inject fluid in the reservoir: flow interruption will then result in a gradual pressure fall-of.

Pressure Transient Analysis (PTA) provides information on reservoir permeability and continuity, which are calculated from drawdown, build-up and/or fall-of behavior. Flow rate tests help evaluate reservoir response under different pressure drawdowns. Reservoir characterization comprises combining and interpreting the available data, to build representations of the reservoir. The scarcely available data is not enough to fully describe subsurface complexity, and reservoir experts then combine interpretation and conceptual descriptions of the reservoir with probabilistic approaches.

The exact rock properties and geological structures cannot be measured directly, and thus the reservoir response to a given field exploitation strategy is uncertain. Figure 3.5a illustrate how different distributions on rock properties can reflect on the same pattern in a wellbore, and Figure 3.5b shows schematically the main types of porosity and permeability distribution — all of them indirectly assessed through core data, well logs and flow behavior.



(a) Wellbore and property distribution

(b) Major types of rock property distribution

Figure 3.5: Schematic view of reservoir complexity and uncertainty (BAKER; YARRANTON; JENSEN, 2015): (a) Different possible porosity and permeability distributions to the same wellbore; (b) Major types of permeability and porosity distribution.

Reservoir uncertainty can be either related to measurement uncertainty or inference uncertainty (MA; POINTE et al., 2011). The former is related to inherent properties of measurement tools and data handling, which can lead to measurement errors, while the latter is the conceptual, interpretational and methodological choices that reservoir specialists make to characterize its complexity.

Take as an example a logging tool used to measure formation porosity at wellbore. Beyond environmental variables and tool parameters that can influence the measurement (measurement uncertainty), logging data does not provide a direct measurement and has to be calibrated and integrated with core and fluid data and flow tests (inference uncertainty). The true porosity is then estimated within an uncertain range.

As exposed in section 2.2.1, probability is adopted as a measure of uncertainty at least since the advent of quantum mechanics in the late 1960's (MA; POINTE et al., 2011). As with risk assessment, reservoir modelling also deploys probabilistic methods to deal with its uncertainties.

Uncertain events concerning the reservoir dynamics can lead to consequences that impact the reservoir professional goal to optimize hydrocarbon exploitation. Oil flow estimates, used to calculated the expected economic value associated to a exploitation strategy, is also subject to uncertainties.

If we define risk as the consequences associated to an uncertain event, it becomes clear why predicting reservoir behavior is an inherently risky activity. Uncertain events associated to the reservoir dynamics may lead to negative impacts in oil & gas projects. Economic value of exploitation strategies is at stake, and the long term results of a selected strategy is uncertain.

In our work, we use conceptual modelling to clarify uncertain events in the reservoir activity. We seek to translate complex reservoir behavior to a set of events, agents and agents' characteristics, and the relations among them. We expect that this model will help reservoir specialists in the identification of reservoir-related risks in oil & gas projects. By providing a tool to deal with risk complexity, we expect that some risk uncertainties may be better managed — those related to inadequate representation of complexities.

In the next chapter we present what are ontologies, and observe how their capacity to explicitly define our risk domain can be useful in dealing with risk complexity. We also show how ontologies are applied to enhance natural language processing tasks.

4
Ontology and Information Systems

In Philosophy, Ontology corresponds to the science of *being*, and studies concepts that apply to everything that exists. Defined by Aristotle as the science of "being *qua* being", Ontology describes the nature and structure of things *per se*, independently of any further considerations — and even of their existence (GUARINO; OBERLE; STAAB, 2009).

For knowledge-based systems in the computer science domain, what “exists” is what can be represented (GRUBER, 1993). This leads one of the most broadly adopted definitions of ontology, proposed by (STUDER; BENJAMINS; FENSEL, 1998) as an adaptation of Gruber’s and Borst’s (BORST, 1999) definition:

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

Conceptualisation is an abstract model of some knowledge domain: the objects, concepts, and other entities that are assumed to exist in some area of interest and the relationships that hold among them (GENESERETH; NILSSON, 2012).

Explicit and formal in the sense that an ontology is expressed according to a machine-readable language, containing an explicitly defined vocabulary expressing entities of the knowledge domain (such as classes, relations, functions), and the formal axioms that constrain the interpretation of such entities, intentionally and explicitly specifying the conceptualization.

Shared because an ontology represents a commitment of a community towards the proposed conceptualisation. This agreement is important for practical reasons, because according to Guarino *"the ontology may turn out useless if it is used in a way that runs counter to the shared ontological commitment"*.

Ontology is a representational artifact, designed to refer to the given knowledge domain (ARP; SMITH; SPEAR, 2015). It defines a common vocabulary for knowledge processing, sharing and reuse in a given domain — which includes machine-interpretable definitions of basic concepts in the domain and relations among them (NOY; MCGUINNESS et al., 2001).

Because ontology aims to represent intended states of the domain, choosing an adequate vocabulary and domain of discourse is a key factor to ontology construction. A non-sufficient vocabulary will fail to represent some intended state of affairs, while the excess of terms render reasoning intractable, or even allow non-intended states of affairs (GUARINO; OBERLE; STAAB, 2009).

The degree to which the vocabulary is formalized ranges from simple lists of concepts in a domain, up to shareable formal structures with rich semantics. This view can be organized in an ontology-continuum, a spectrum mentioned by many researches of the domain (KHAZRAEE; LIN, 2011), and illustrated in Figure 4.1.

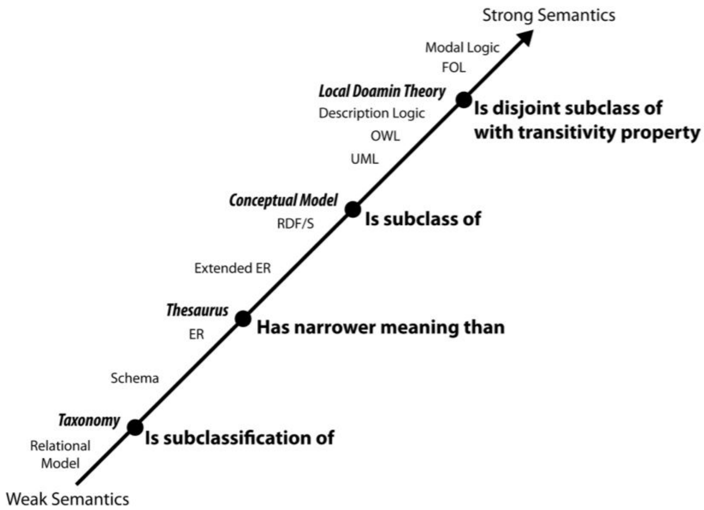


Figure 4.1: Ontology spectrum based on formal semantics adopted (DACONTA; OBRST; SMITH, 2003)

Another way of classifying ontologies is according to their level of generality, as in Figure 4.2 . In this sense, *top-level ontologies* (or foundational ontologies) are context-independent and describe general entities such as *space*, *time*, *matter*, *events* and so on. *Domain ontologies* and *task ontologies* are related to a generic domain or task, such as Medicine or Management. *Application ontologies* are domain-specific, describing concepts that are usually entities performing certain tasks in a given domain.

Other classifications based on ontology's generality establish *core ontologies*, that define terms of a domain (Geology, for example) and are located in a layer between top-level ontologies and domain ontologies, linking the general concepts of the former to specific terms carried by the latter that refer to a sub-field (Statigraphy and Sedimentology for example) (OBERLE, 2004).

Top-level ontologies were made available and are well adopted in the last 20 years, such as DOLCE (GANGEMI et al., 2002), SUMO (NILES; PEASE, 2001), UFO (GUIZZARDI, 2005a) and BFO (ARP; SMITH; SPEAR, 2015). In the Geology domain, it is worth noticing the development of the GeoCore, (GARCIA et al., 2020a) a core ontology that elucidate many of necessary concepts for the development of this work.

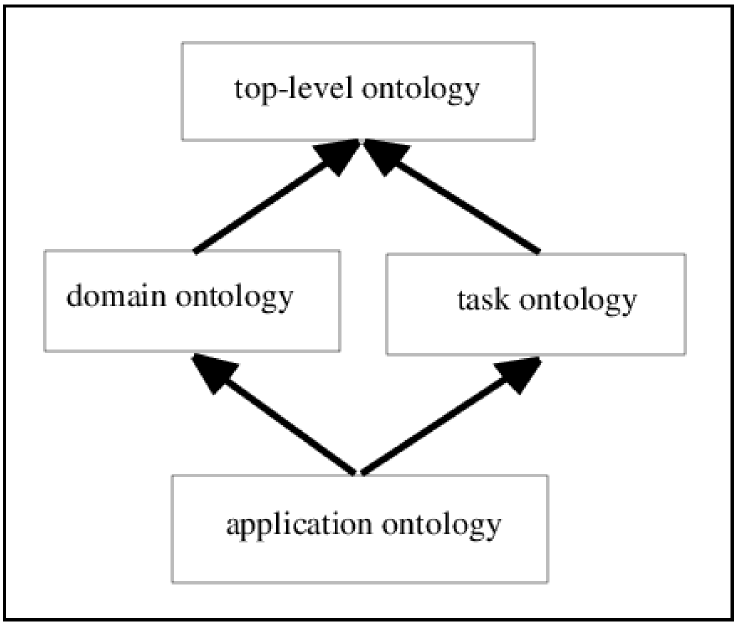


Figure 4.2: Ontology classification based on generality (GUARINO, 1998)

Ontologies are useful in various tasks where automated knowledge interpretation and / or sharing is required. They allow communication among various knowledge-based systems, by sharing a common vocabulary and axioms. (GRUBER, 1993).

In the late 1990s, the development of the Resource Description Framework (BRICKLEY; GUHA; LAYMAN, 1999) provided the necessary language to encode the structured content on the Web for knowledge processing tasks. RDF language, along with ontologies, are the basic components of the Semantic Web, providing specification of classes and inference rules that boost the results of search engines, relate content that are relevant, disambiguate meaning, and so on (BERNERS-LEE; HENDLER; LASSILA, 2001).

More recently, domain experts develop ontologies for specific tasks, annotating and sharing information in their area of expertise. Besides providing the means for making explicit domain assumptions (NOY; MCGUINNESS et al., 2001), these tasks may involve natural language processing, such as question & answering and automated reasoning.

In the next subsections, we present an overview of application of ontologies in natural language processing tasks, and we provide an overview on the top-level, core and domain ontologies that structured our work, namely the Basic Foundation Ontology (BFO), the GeoCore Ontology and the Common Ontology of Value and Risk. (SALES et al., 2018).

4.1
Ontologies and Natural Language Processing tasks

Although the use neural-network-based methods and word embeddings are pointed out as the most promising techniques for the future challenges of natural language processing tasks (ITTOO; BOSCH et al., 2016), ontologies applied to such tasks have also proven to deliver interesting results, especially in cases where the amount of textual data is limited. The application of ontologies has the potential of transforming texts, which are a non-structured type of data, in machine-readable knowledge bases. In this section we discuss briefly some ontologies applied to Natural Language Processing (NLP) tasks in Portuguese language.

(ALBUQUERQUE et al., 2016) propose a methodology to interpret *tweets* containing descriptions of traffic-related events, using the TEDO (Traffic Event Domain Ontology) as model to define the event's entities (*e.g.* actor, location) and their relations. The *tweets* are then subject to *Named Entity Recognition* and *Relation Extraction* tasks, and then expressed as RDF triples — enabling the interpretation of traffic events reported in selected *Twitter* accounts.

Following a similar methodology, (FURTADO, 2017) proposed an ontology aiming at the interpretation of operational reports in oil & gas platforms. This work used the ISO 14224 (*Petroleum, petrochemical and natural gas industries—Collection and exchange of reliability and maintenance data for equipment*) as the basis for the formal representation of the operational events (as for an example, the failure of a pump).

Once a knowledge base composed of texts in natural language is structured in RDF format according to a specific vocabulary (or ontology), it becomes possible to query this base by asking question in natural language, a task known as Semantic Question Answering (SQA). After investigating 72 publications about 62 SQA systems developed from 2010 to 2015, (HÖFFNER et al., 2017) concludes that aspects of natural language such as ambiguity, generality and multiple possible languages of the queries still pose major challenges to the performance of SQA tasks.

Aiming at delivering the best fit results to a question in a SQA tasks, (SOUSA, 2019) uses domain ontologies, NLP techniques, and knowledge bases built from these ontologies in a methodology to structure and interpret questions in natural language.

Finally, (SANT'ANNA et al., 2020) structure knowledge about product compatibility based on pairs of questions and answers already answered by human attendants of an e-commerce customer service. To handle meaning, an ontology is develop for product compatibility representation. In a brief period of evaluation, the system was able to automatically answer 3.9% of 2,667 questions posed to SQA system.

In the oil & gas domain, the most remarkable effort of standardization is the ISO 15926 standard (ISO Central Secretary, 2003). It specifies an ontology for asset planning for process plants, including oil and gas production facilities, supporting data exchange and interoperability among multiple computer systems. Production planning and risk analysis, however, are out of the scope of ISO 15926.

In our work, we believe that a proper conceptualisation can clarify the concept of risk, thus allowing semantic reasoning over risk documentation. To validate this idea, we will perform natural language processing tasks — namely Named Entity Recognition (NER) — as the means of evaluating the proposed ontology. We analyse the completeness and soundness of the ontology by its ability to structure risk description sentences in natural language in triples RDF.

One important aspect of ontologies is that they allow interoperability across different tools and information systems. To guarantee an ontology that is prone to be reused by the reservoir community, we extend top-level and core ontologies that are already consolidated and have been validated. In the next section we present ontologies that provide founding concepts to the development of ResRiskOnto.

4.2 Reused Ontologies

This section presents an overview of the ontologies that were reused or served as conceptual foundations to the development of ResRiskOnto. The first two subsections describe BFO an GeoCore, ontologies that were fully extended into the ResRiskOnto. The third subsection shows how the Common Ontology of Value and Risk provided the basis to the definition of the risk concept in ResRiskOnto. Finally, in the last subsection we present other work that provided previous definitions for reused entities during our ontology development.

4.2.1 Basic Formal Ontology

The Basic Formal Ontology is a small ontology "developed to support integration of data obtained through scientific research" (ARP; SMITH; SPEAR, 2015). Because of the generality that characterizes its entities (material enti-

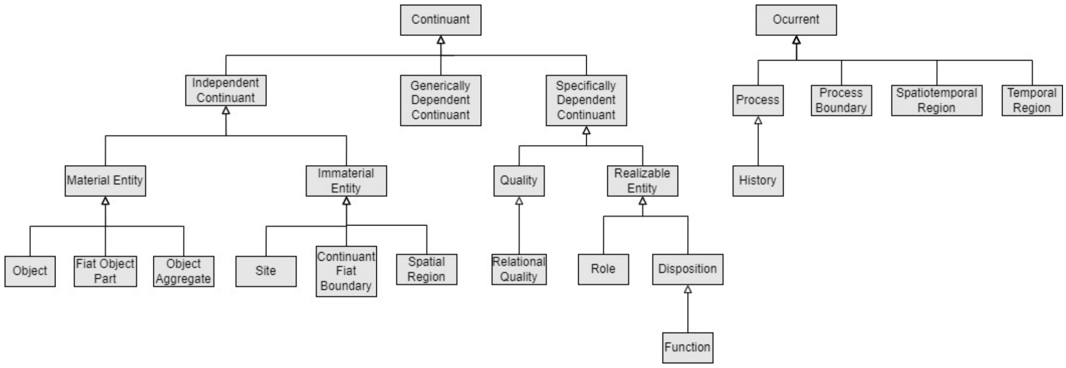


Figure 4.3: Hierarchy tree of BFO entities (ARP; SMITH; SPEAR, 2015)

ties, qualities, processes, etc.), BFO provides a common top-level structure that serves as departing point for the development of domain ontologies — thus supporting interoperability of the multiple domain ontologies created in its terms.

The main categories are depicted in Figure 4.3, revealing two distinct main categories: *Continuants*, entities that persist through time, and *Occurrents*, entites that occur in time. These categories represent distinct and complementary perspectives of reality, often expressed simultaneously, as in "*an amount of air* (continuant) *in its natural movement* (occurrent) *is what we call wind*". Continuants are divided in *Independent Continuants*, *Generically Dependent Continuants* and *Specifically Dependent Continuants*, while Ocurrents are divided in *Processes* and *Temporal Regions*.

Independent Continuants are those that maintain their identity and existence despite gain and loss of parts or changes in their qualities. They are also the bearer of qualities, that are said to *inhere in* them — so that the **color green** of my **zamioculca plant** is *inherent* to the plant, i.e. it exists as long as the plant exists. Independent Continuants can be either *Material* or *Immaterial Entities*.

Material Entities have some portion of matter as part, and are divided in *Object*, *Fiat Object Part* and *Object Aggregate*. Objects have their parts causally unified, extend spatially in three dimensions and are maximally self-connected. Examples of Objects are a **dog**, a **plant**, a **person**. An Object Aggregate is a collection of Objects, whose parts are exactly exhausted by the Objects of the collection. An example is a **dog pack**, **the woods**, a **symphony orchestra** (in this last case specific objects of the aggregate play different roles). A Fiat Object Part is a proper part of an Object but is not demarcated from the remainder of this Object by any physical discontinuities — so that it is not an Object itself. Examples would be a dog's **snout**, a tree's **branch**, a person's **arm**.

Immaterial Entities are Independent Continuants that contain no matter as a proper part. They are divided in *Continuant Fiat Boundaries*, *Sites*, and *Spatial Regions*. A Continuant Fiat Boundary is a boundary of a Material Entity that exists exactly where that object meets its surroundings. An example it **the surface of the Earth**. 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 cavity of a snout**, **the inside of a person's mouth**, or **the trunk of a car**. A Spatial Region is a part of space, that can be occupied both by Material or Immaterial Entities. Spatial Regions are defined relative to some frame of reference (e.g. a Geographic Coordinate System, an Astronomical Coordinate System).

A Generically Dependent Continuant can be thought of as a complex Continuant pattern that exists only if it is *concretized* in some counterpart Specifically Dependent Continuant. They are Dependent Continuants that seem to be capable to migrate from one bearer to another, as multiple copies. Examples are a **pdf file**, the **plaid pattern** of my shirt, the **Apple logo**. An Independent Continuant is the *bearer of* a Generically-Dependent Continuant.

Specifically Dependent Continuant is a Continuant that depends on an Independent Continuant that is its bearer, existing only as long as the Independent Continuant exists. They are divided in *Qualities* and *Realizable Entities*. If a Quality inheres in an entity at all, it is fully exhibited or manifested or realized in that entity. On the other hand, a Realizable Entity can inhere in a Continuant without being realized, or sometimes only being realized through a process. A *Relational Quality* is a Quality with more than one Independent Continuant as bearer — such as a **marriage bond**. A *Role* is a Realizable Entity that is possessed by its bearer because of some external circumstances, an if it ceases to exist there's no change in the physical make-up of the bearer — as a **professor** or a **bodyguard**, for example. A *Disposition* is a Realizable Entity that is possessed by its bearer because of internal circumstances, an if it ceases to exist the bearer is physically changed — as a **disease**, for example.

Processes are Ocurrents that exist in time by occurring or happening, have temporal parts and depend on some Material Entity. Examples are **person's life** or the **blossoming of a flower**. Material Entities *participates in* the Process. Dispositions are also *manifested by* Processes (such as the disposition of a bud of blossoming into a flower).

A *Process Boundary* is an Occurrent entity that is the instantaneous temporal boundary of a process. Finally, *Temporal Region* is an Occurrent entity that is a part of time.

4.2.2 GeoCore Ontology

GeoCore Ontology is a core ontology for the description of geological knowledge (GARCIA et al., 2020a). It was developed using the BFO top-level ontology (ARP; SMITH; SPEAR, 2015) with the purpose of facilitating the development and integration of domain ontologies in the geological domain. In the following, we describe the entities from GeoCore depicted in Figure 4.4 that we used in this work.

The central idea in the GeoCore Ontology is that geoscientists have to deal with two distinct kinds of material entities: *Earth Materials* and *Geological Objects*. The relation between Earth Materials and Geological Objects is one of **constitution** — the relation between something and what it is made of —, which, in the view adopted by GeoCore, isn't a parthood relation.

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 and aren't necessarily maximally connected, but they are ontologically rigid and provide an identity criteria. On the other hand, a *Geological Object* is a naturally occurring entity *constituted by* some Earth Material that is maximally connected, thus, provides an unifying criteria.

The external surface delimiting a Geological Object is named a *Geological Boundary*. Geological Objects may *bear* some *Geological Structure*, which is the pattern of the internal arrangement of the object.

Finally, when two distinct Geological Objects are in physical contact (i.e., are physically adjacent), they are in a special relation of *Geological Contact*.

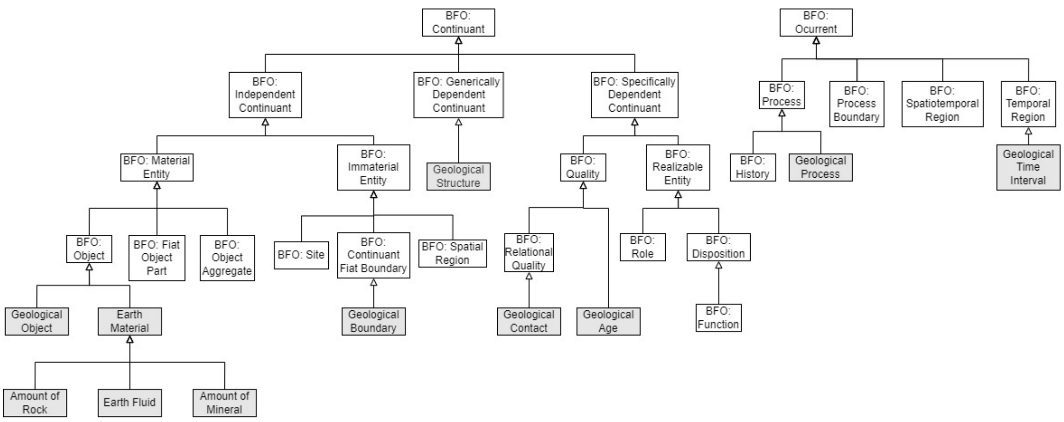


Figure 4.4: GeoCore overview and subsumption relations (GARCIA et al., 2020a)

4.2.3
Common Ontology of Value and Risk

Conducted under the principles of the Unified Foundational Ontology (UFO) (GUIZZARDI, 2005b), the Common Ontology of Value and Risk (SALES et al., 2018) provides a well-founded ontology which describes value and risk in terms of events and their causes.

This ontological analysis proposes a relationship between *risk* and *value* notions, and concludes that the process of assessing risk is a particular case of that of ascribing value. By formalizing those two concepts, the authors seek to disentangle three perspectives: (i) an experiential perspective (value and risk in terms of events and their causes), (ii) a relational perspective (the subjective nature of value and risk), and (iii) a quantitative perspective.

The work presents the similarities between value and risk as both having goal dependency, context dependency, uncertainty and impact — thus, Value and Risk are commonly decomposed into "smaller" events.

The experiential perspective, depicted in Figure 4.5, shows that risk is ascribed to objects and events, and was the main guide to instantiate an applied ontology to the petroleum reservoir risk scenario — since our work will be centered in defining such *objects* and *events*.

Risk Experiences are decomposed into unwanted events that have the potential of causing losses. *Threat Events* might be intentional (such as a cyber attack) or unintentional (slipping on a wet floor, for example). *Loss Event* is the event that directly impacts (negatively) the *Intentions* of a *Risk Subject* — i.e., the subject that perceives the Risk Experiences as a damaging one.

The Common Ontology of Value and Risk also differentiates the roles played by objects in the Risk Experiences, so that *Threat Objects* cause the threat, *Objects at Risk* are exposed to potential damage and *Risk Enablers* play an ancillary role in the Risk Experience. The dispositions of those objects are manifested through the Risk Experience, are *Threat Capabilities* (in the case of Threat Objects), and *Vulnerability* (for Objects at Risk or Risk Enablers).

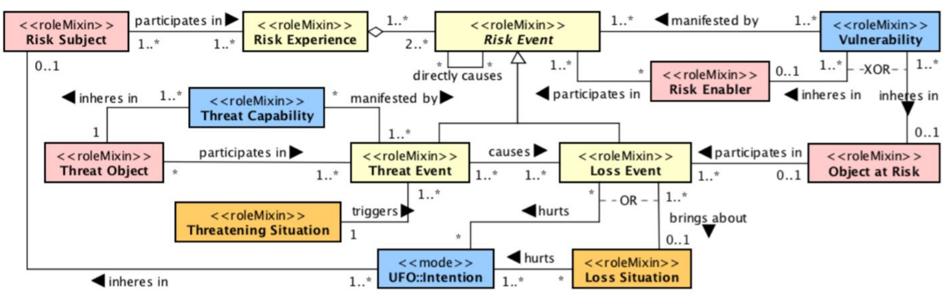


Figure 4.5: Risk experiences, their parts and participants (SALES et al., 2018)

4.2.4
Other ontologies

Some concepts already defined in previous work were reused during the development of the ResRiskOnto. In this subsection, we briefly describe other reused terms and their sources.

GeoReservoir Ontology The GeoReservoir Ontology (CICCONETO, 2021) is an artifact that supports the description of deep-marine depositional system geological occurrences, that constitute the main type of petroleum reservoir in the world.

In its foundations is an effort conducted along with professional reservoir geologists to collect and disambiguate the Geology terminology to describe reservoir occurrences in deep-marine depositional systems and build a domain ontology for reservoir description.

Entities in GeoReservoir Ontology are distinguished between *substantials*, or existentially independent entities (e.g. a rock), and *moments*, existentially dependent entities (the dimensions of the rock).

In our work, we are interested in the way the GeoReservoir Ontology constructs the concept of *Depositional Unit*. Depositional Units are Sedimentary Geological Objects (which are Geological Objects constituted by some Sedimentary Rock or Sediment), recognizable in a mapping scale of at least 1:1000 m.

The concept of Depositional Unit will be used to identify instances of the word "layer". Because deposition is a process governed by gravitational forces, depositional units stack on each other as geological layers. GeoReservoir defines this entity in such a way that depositional bodies in all scales are instances of Depositional Unit. Because layer is a loosely-defined word that is broadly adopted by experts in our domain, we consider that this scale-blind aspect of this entity better corresponds to the concept that is expressed in natural language.

Moments characterizing geological objects are also of our interest, and we import the concepts of *facies*, *dimension* and *geometry*. Of special interest is the definition for facies, a combination of features that might repeat in several geological objects as a pattern.

Information Artifacts Ontology "Model" is among the fifteen most frequent words in our risk sentence corpus (after the removal of stop words). It stands for a simplified representation or interpretation of reality, or the interpretation of the fragment of a system according to a structure of concepts.

In Chapter 3 we describe the reservoir characterization activity, in which professionals gather rather scarce data, composed of direct or indirect measurements of reservoir properties, and combined the data with their field knowledge to build conceptual, analytical and mathematical models that characterize the reservoir.

ResRiskOnto intends to clarify events in reservoir-related risk experiences, and we believe a thorough discussion on reservoir modelling should be conducted in order to clarify relations between models, reservoirs and possible states of reality. For this reason, we offer a very limited set of concepts to deal with frequent words related to reservoir characterization.

We base those definitions in the Information Artifact Ontology (IAO), which is an attempt to give a realism-based account of the essence of information entities and how components of such entities relate to each other and to that what they are information about (CEUSTERS, 2012).

IAO defines an *Information Content Entitiy* (ICE) as generically dependent continuant that is about some thing. We specify a reservoir model as an ICE, referring to the software content generated by reservoir modellers, that can be transferred and copied from one computer to another. However, a model *is about* a limited portion of reality, which are the material entities that provide actual data from the reservoir. The unknown parts of the real reservoir are expressed in the model according to the knowledge of the modeller.

To deal with cognitive aspects of representation artifacts, we also adopt the in-depth analysis on "aboutness" conducted in (CEUSTERS; SMITH, 2015), that attempts to launch the necessary effort to carefully treat the aboutness relations between ICEs and associated cognitive representations and their targets in reality. Representation, in this sense, is a Quality which is about or is intended to be about a PORTION OF REALITY (POR).

In the next Chapter we present the methods applied to create ResRiskOnto, and provide the definitions and axioms of our ontology for reservoir-related risk events in oil & gas projects.

5
Ontology Development

In this section, we present the means adopted to support the development of the ResRiskOnto Ontology, an ontology that provides the proper conceptualization of reservoir-related risks in the oil & gas domain. The ResRiskOnto is developed as an extension of the BFO and GeoCore ontologies, and uses the Common Ontology of Value and Risk as conceptual basis for risk descriptions. BFO, GeoCore and Common Ontology of Value and Risk are briefly describe in Sections 4.2.1, 4.2.2 and 4.2.3 respectively. ResRiskOnto also reuses concepts developed in the GeoReservoir Ontology and Information Artifacts Ontology.

To support the development of the ontology, we conducted a careful semantic analysis of the risk documents provided by Petrobras. Contributions of domain experts were crucial at every step of the ontology development — through the participation at workshops that elucidated the community's needs, in discussions over the technical aspects of risk events, providing agreed-upon definitions used in the ontology's entities, annotating documents and sharing insights in valuable conversations.

The adoption of the concepts proposed in this ontology is still a challenge to the community, one which we expect to overcome through the application of visualization techniques that render the ontology more easy to understand and manipulate, and by providing the necessary training in the main concepts an entities in the ontology.

In the first sections, we describe the tools and methods applied in the ontology development. Then we present the ResRiskOnto, providing definitions for the entities and relations covered by the ontology.

5.1
Building the Ontology

Building an ontology is an iterative process conducted under certain methodological guidelines. As a structured methodology for ontology engineering, we adopted concepts of METHONTOLOGY (CORCHO et al., 2005) and NeON (SUÁREZ-FIGUEROA et al., 2012). We also applied methodological best practices indicated in the works of Noy and McGuiness (NOY; MCGUINNESS et al., 2001) — that enumerate a series of steps to be executed during the ontology engineering — and of Arp, Smith and Spear (ARP; SMITH; SPEAR, 2015) — in their proposed guidelines of how to evolve from a set of domain terms to an ontology by developing Aristotelian definitions for the ontology's terms.

METHONTOLOGY defines five activities in the iterative cycle of ontology building:

1. *Specification*: this step consists in stating why the ontology is being built, and comprises the definition of its intended uses and end-users;

2. *Conceptualization*: activity in which a perceived view of the domain is organized into a semi-formal specification with the use of taxonomical structure and graph notations;

3. *Formalization*: during the formalization we transform the semi-formal ontology into a formal conceptual model, using formal definitions form the terms and establishing the necessary formal relations;

4. *Implementation*: activity that builds computable models using an ontology language (such as RDF Schema or OWL, for example);

5. *Maintenance*: is the activity of updating, complementing and correcting the ontology when needed.

Our main source of domain knowledge to organize the concepts of the reservoir-related risk realm was the repository with approximately 2500 sentences described in Chapter 1. The first iteration of the ontology building process was conducted using a subset of the original corpus, containing roughly 340 sentences describing those risks associated with fouling and geomechanical aspects of the petroleum reservoir. This iteration was performed as an initial evaluation of our framework for the development of the conceptualization of risks in oil and gas projects documented by specialists of the Petroleum Reservoir domain, and is described in (SILVA et al., 2021).

The *conceptualization* and *formalization* steps in ontology engineering are those during which we acquire the necessary knowledge and conduct an ontological analysis of domain terms. Here, our goal was to define consistent subsumption relations between domain terms and the BFO and GeoCore entities. *Subsumption* is a taxonomical relation between two entities, p e q, where if p subsumes q, all instances of q are also instances of p (GUARINO; WELTY, 2002).

The ontological analysis relies on the philosophical notions of identity, unity, essence and existential dependence — or "ontological metaproperties" — to identify the entities' ontological nature and relations (GUARINO; WELTY, 2000):

- *Identity*: identity is related to the problem of distinguishing a specific instance of a certain class from other instances of that class by means of a characteristic property, which is unique for it (that whole instance).
- *Unity*: differently from identity, 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 them together (not involving anything else).
- *Essence*: a property holding for a certain individual in a certain state of affairs at time t is said to be essential if it necessarily holds for this individual at every possible time in every possible world, i.e. it must hold for every possible instance of this individual.
- *Rigidity*: when dealing with properties, one wants to identify which ones can change and which must not, or even reidentify an instance of a certain property after some time. This leads to a special type of essenciality. Rigidity is a property that is essential to all its instances. *Person*, for example, is a rigid property since no person instance can cease to be an instance of person and continue to exist.
- *Anti-rigidity*: contrary to rigid properties, anti-rigid properties are not essential to all its instances. *Student* is an anti-rigid property, since a student could cease to exist by interrupting his studies, but would still exist as a *Person*.
- *Existential dependence*: it relates to the case when all instances of a concept depend on other entities' existence to exist. For example, the height of a person depends on the existence of that person to exist.

Supported by the ontological analysis, coherent subsumption relations are established between the terms and BFO and GeoCore. During this process, Aristotelian definitions were created in order to formalize the ontology, following principles proposed in (ARP; SMITH; SPEAR, 2015):

1. Provide all nonroot terms with definitions.
2. Use Aristotelian definitions. An Aristotelian definition is in the form A =def is a B that C, where A is the term we are defining, B is a class of GeoCore or BFO, and C is the set of properties that makes A a specialization of B.
3. Use essential features in defining terms. 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 the domain. Using Aristotelian definitions, we start by defining the most general terms (the ones in the upper parts of the taxonomical tree), and move downwards to more specific terms.

5. Avoid circularity in defining terms. Circularity occurs when the term is present in it's own definition, e.g. "hydrogen = def. anything having the same atomic composition as hydrogen".

6. Use simpler terms than the term you are defining.

7. Do not create terms for universals through logical combinations.

8. Definitions should be unpackable. This means that 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,".

Ontologies also comprise the definition of relations for entities. Take, for example, the relation of constitution defined in the GeoCore Ontology: "*Geological Object* is <u>*constituted by*</u> only *Earth Material*", meaning that all instances of Geological Object can only be constituted by instances of Earth Material. In the case of Geological Age, a "*Geological Object* <u>*has age*</u> some *Geological Age*", we conclude that every Geological Object, have at least one Geological Age.

During ontology *implementation*, the available relations constraints are:

– some: existential quantifier (at least one instance);
– only: universal quantifier (all instances);
– min n: existential quantifier with minimum cardinality;
– max n: existential quantifier with maximum cardinality;
– exactly n: existential quantifier with exact cardinality;

The next subsections show how each of the steps in the ontology construction were applied in the case of ResRiskOnto.

5.2 Ontology Specification

To specify the ResRiskOnto, we developed an ontology requirements specification document (ORSD) as in the methodology proposed in (SUÁREZ-FIGUEROA et al., 2012), stating why the ontology is being built, its intended uses and end users, and the specific requirements it should fulfill are.

5.2.1 Purpose

The main goal of the ResRiskOnto is to provide a proper conceptual modelling of the risks in the petroleum reservoir domain. It should allow users of the oil & gas industry to (i) standardize risk assessment documentation; (ii) manage risk data; thus (iii) enhancing risk analysis in the reservoir domain.

5.2.2 Scope

ResRiskOnto should provide enough concepts to describe the events of the risk assessment documents obtained from workshops of project-phase risk identification of the last ten years in Petrobras.

5.2.3 Implementation Language

ResRiskOnto will be implemented using the Web Ontology Language (OWL).

5.2.4 Intended end users

The ontology intended end users are reservoir specialists, either geoscientists or engineers. Other possible users are IT professionals and data scientists.

5.2.5 Intended uses

Intended uses of this ontology are:

- Providing an adequate taxonomy for the annotation of preliminary risk documents;
- Allowing semantic reasoning over the risks documents, by providing enough classes and relations to the execution of Named Entity Recognition an Relation Extraction tasks;

– Serve as basis for data modelling in the case of risk data storage and analysis;

– A standard for future risk documentation in the reservoir domain.

5.2.6 Ontology requirements

Requirements are the general aspects the ontology should fulfill. These can be either non-related to the ontology content (non-functional requirements) or content-specific (functional) requirement in the for of group competency questions and their answers.

(a) Non-Functional requirements
All terms should present labels in English and Portuguese languages. Terms' definitions must be in English, following the good practices established in (ARP; SMITH; SPEAR, 2015) and described in 5.1.

(b) Functional requirements
During workshops with domain experts, we noted the main objective regarding the ResRiskOnto: to standardize the description of reservoir-related risk in oil & gas projects. Based on this standardization necessity, we formulated the following competency questions:

1. What types of events compose a Risk Experience in the Petroleum Reservoir domain?

2. Does the ontology properly describes at least 250 risk sentences?

3. Is the performance evaluation of the NER tasks in the original corpus annotated with the ontology satisfactory?

4. Do reservoir experts recognize the proposed risk events?

Given the ontology scope — providing enough concepts to describe the events on the risk assessment corpus — this so-called corpus is the main source of information for the conceptualization step, described in the next section.

5.3 Ontology Conceptualization

Because our main source of information regarding reservoir-related risks is the corpus of sentences, we followed a corpus-based methodology to select ontological seed proposed in (GARCIA et al., 2020b).

The work describes a Frequency-based Ontological Analysis of Petroleum Domain Terms. Applying statistical analysis of relevant terms in a corpus of the geology domain, it defines a set of terms as the first step towards developing a domain ontology for Petroleum Geology.

The proposed framework ranks the relevant terms present in a domain thesaurus according to their frequency in a selected domain corpus, which are then examined by domain experts, in order to identify the continuant entities relevant within the geological domain (i.e., those that are exclusively related to Geology).

After applying the basic raw text processing tasks (tokenization, removal of punctuation, stopwords and special characters), the application of statistical analysis provided a rank of the words appearing in the sentences of the reservoir risk corpus. Approximately 1400 words were then subject to two questions:

1. Is it a term from the Petroleum Reservoir domain?
2. Does it refer to a continuant (i.e., an object or a quality to describe an object)?

The remaining terms in Portuguese have their meanings documented, using definitions based on Petrobras' Glossary (PETROBRAS, 2007), Schlumberger Oilfield Glossary (SCHLUMBERGER, 2019), the Open Wordnet (TESSAROLLO; RADEMAKER, 2020) and Wikipedia. It is important to note the importance of those previous efforts into reaching a consensus among the community's experts, which made possible the existence of such glossaries.

The existing definitions, combined with the previous experience of this thesis author as a reservoir engineer, were then used as ontological seed to build a conceptual model of the phenomena in petroleum reservoir that lead to risk experiences in investment projects.

Once the continuants were organized, terms categorized as occurrents were selected and also had their definitions documented. In this case, for most of the terms there was a lack of previous definitions.

The selection of terms that constitute the ontology was in itself an iterative process. After capturing the domain knowledge, the ontological analysis of the terms was conducted to create the ontology's first conceptual model. We believe that an ontology constructed with the words dominated by domain experts can be easily adopted, since frequent terms would naturally represent concepts that are already agreed upon.

The next subsection contain important aspects of the ontological analysis of the domain that support the adopted definitions of Risk.

5.3.1 What type of event is a Risk Event in the Reservoir domain?

Our first analysis is related to the conceptualization of a *Risk Experience* according to (SALES et al., 2018). The experiential perspective of a Risk is composed by three main types of entities: *events*, *objects* and *qualities* (or *modes*), as in Figure 4.5.

The structure of the risk sentences in our corpus lead to the understanding that each sentence describes one *Risk Experience* as perceived by the Company itself. Petrobras deploys its oil & gas projects with the intention to achieve certain *Intentions*. Each risk related to a project, as stated in our corpus, jeopardizes the intentions of the Company as a whole. For this reason, we make two assumptions:

1. One sentence from the risk corpus is equivalent to one *Risk Experience*; and
2. The *Risk Subject* of each of these experiences is Petrobras itself.

These assumptions are important because they remain implicit in the sentences: one would never state a risk of "*gas canalization due to reservoir heterogeneity, leading to a reduction in oil recovery, which affects Petrobras' intention to produce a certain amount of oil*". We understand that the last part of the sentence remains implied, because of the context in which they are created. Experts will not likely state those are Petrobras' risk because the process itself is a Risk Asessment of Petrobras' projects.

Regarding the *Risk Events* that constitute a *Risk Experience*, they can be of two types: *Threat Events*, the ones with the potential of causing a loss (either intentionally or unintentionally), and *Loss Events*, that necessarily impact intentions in a negative way.

Looking at the proposed risk conceptualization in (SALES et al., 2018) and comparing to risk sentences, we notice that all those events are perceived through the behavior of the material objects that participate in them, such as the produced fluids, reservoir rock, the wells, etc. In this sense, we propose a definition for a Risk Event in the Reservoir domain that subsumes the concept of *Process* present in the BFO Ontology:

Risk Event = def. a *BFO:Process* that either (i) has the potential of causing a loss or (ii) impact intentions of a subject in a negative way.

And, consequently, Loss Event and Threat Event are defined as:

Loss Event = def. a *Risk Event* that impact intentions of a subject in a negative way.

Threat Event = def. a *Risk Event* that has the potential of causing a loss.

Specifically for applications aimed at the corpus of reservoir risks, we defined the risk subject as Petrobras itself. This risk repository was generated over years in a non-standardized fashion, in such a way that picking specific loss or threat events could be somewhat an heroic task. Luckily, the most recently updated guidelines to the Risk Management activity in Petrobras indicate a proper taxonomical structure according to which risks should be identified. Such sentences should be written in a way that emphasizes three elements of Risk:

Risk of [UNCERTAINTY] due to [CAUSE], that may lead to [CONSEQUENCE].

Consequence is *a result of a particular action or situation, often one that is bad or not convenient*[1]. So in our case, we consider impact as the result, or consequence, of the risk experience — which leads to the association of the concepts of *Loss Event* (the event describing an impact) to that of *consequence*.

Back to the proposed sentence structure, the exchange with domain experts lead to a very specific understanding of *cause* and *uncertainty* in the case of risk assessment. Uncertainty is inherent to reservoir characterization for reasons already discussed in Chapter 3, and uncertain parameters of the reservoir cannot be changed, but investigated. Those experts, however, understand that there is a number of decisions regarding the reservoir development strategy that can be at the source of a risk experience.

We then define two types of threat events:

Source Event = def. a *Threat Event* that has participants entities with characteristics that are controllable by human action.

Uncertain Event = def. a *Threat Event* that has participants entities of the Reservoir, with characteristics naturally defined and non-controllable by human action.

[1]https://dictionary.cambridge.org/

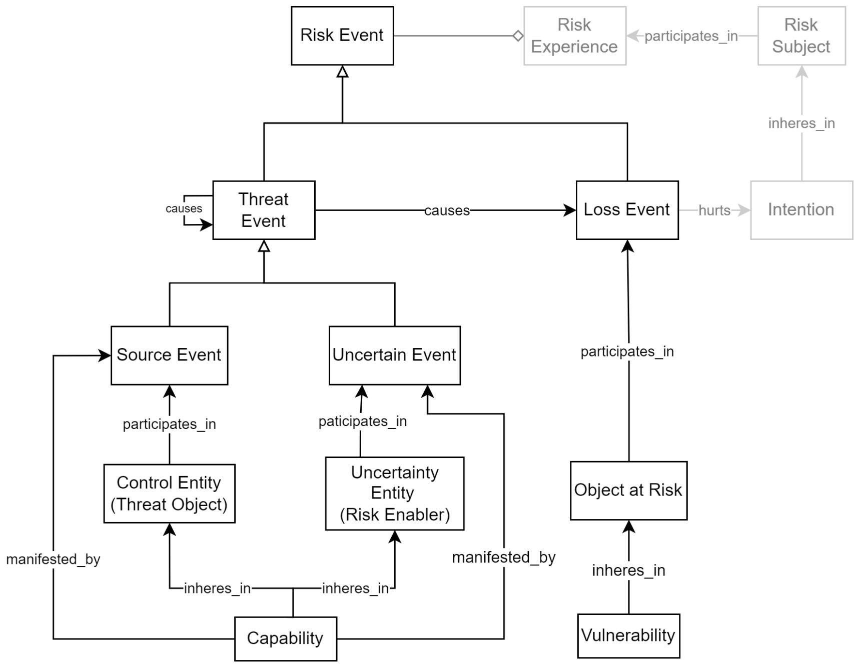


Figure 5.1: Experiential view of Risk in the Reservoir domain

The risk conceptualization in our domain, adapted from (SALES et al., 2018), can be seen in Figure 5.1. The elements in gray boxes are those that are not explicitly formalized in our work, since they remain implicit in the corpora, and should be subject of a wider discussion, one that involves other actors in oil & gas development projects than the reservoir experts.

In its guidelines, Petrobras recognizes the following impact dimensions for the risks in its projects:

– Non-financial impact: HSE (Health, Safety and Environment), legal and compliance, image and reputation;

– Financial impact: (oil) volume, investment, operational costs, deadlines and NPV (net present value).

In order to achieve the ontology's purpose, the one of providing a proper conceptual modelling of the risks in the petroleum reservoir domain, we want to reuse as much as possible the concepts previously agreed upon — common domain words, predefined glossaries, and, of course, risk guidelines. Those impact dimensions resonate with the Company's goals in an investment project, and for that reason we think of them as the implicit *Intentions* of our Risk Subject that are hurt during the Risk Experience.

Having in mind our subject's intentions, we look critically at the corpus to identify those sentence fragments that carry the meaning of a *Loss Event*, and identified situations according to the impact dimension. We emphasize that the majority of risks in the reservoir domain are those that hurt the Company's intention to produce a certain volume of oil. This is easy to understand if we think of the reservoir activity, that aims at maximizing oil recovery. One could think, for example, that an operations engineer would have the goal to guarantee operations at a minimum cost, or the infrastructure team's objective would be to deploy the project at minimum investment cost — and for those domains the main risk impact would be on the operational and investment costs, respectively.

Idealizing loss events as the ones that necessarily impact one of the above mentioned dimensions is something that was well accepted in discussions with experts of the reservoir community. Here are some examples of potential instances of *Loss Events* in natural language, considering some of the established impact dimensions for oil & gas projects:

- Impact on Volume:
  - *decrease in produced oil ratio* (in relation to other produced fluids)
  - *lower well productivity than expected*
  - *gas breakthrough in oil wells*
  - *less oil in place than predicted*
- Impact on Investment Costs
  - *well loss/collapse*
- Impact on EHS
  - *oil exudation*

Despite the fact that the above events are perceived through the behavior of the material objects that participate in them (i.e. produced fluids, well, oil), the *Vulnerability* of each one of these objects manifested in a loss event is almost never explicit in the sentences — the disposition of the oil to leak, or the collapse potential of the well, for example. This will impact how the relations between these characteristics and the risk events are defined.

Having defined what types of instances may constitute a *Loss Event*, we observe frequent words that indicate other events in the risk experience. Here are some potential instances of *Threat Events* in natural language

- Source Events:
  - *well damage during perforation*
  - *fractures inducted by thermal effect*
- Uncertain Events

- *reactivation of the reservoir faults and fractures*
- *asphaltene precipitation inside the reservoir*
- *stratigraphic compartmentation of the reservoir*

Analogously to what happens in the description of *Loss Events* in the case of *Threat Events* the *Capability* of the objects is not always stated. However, while interacting with domain experts it was possible to notice that the terminology used for such capabilities coincides with the one used to name the events. For example, one would say that the "reactivation potential" of a reservoir fault is a disposition realized during a "reactivation" process.

In the next two subsections, we explore the peculiarities associated to vulnerabilities and capabilities manifested by risk events that are expressed in the risk corpus.

5.3.2 The problem in predicting

In chapter 2 we discussed an important aspect of risk events, which is the uncertainty that characterizes them. For this reason, and because a reservoir's characteristics are inherently uncertain (as discussed in chapter 3), the main activity of reservoir experts — the one of maximizing hydrocarbon recovery at a minimum cost — constitutes an inherently risky activity.

Analysing the sentences in the available corpora, we notice a profusion of sentences expressing risk as a forecasting problem. For example, one might state as a risk the "*rock permeability being lower than expected, resulting in smaller amounts of produced oil than predicted*". Sentences with similar construction are so commonly expressed that we felt the necessity to define two characteristics of objects: *unpredictability* and *inadequacy.*

In our work, we advocate the existence of an "optimal" development strategy, one that would maximize the return on investment of the oil & gas project, without compromising the environment nor the safety of the operations. The ultimate goal of reservoir professionals is to define a strategy for oil recovery that is the optimal. This optimal strategy, defined at a certain point in time, is optimal regardless of our perception about it.

Production forecasts have as input uncertain parameters about the reservoir, with the ultimate goal of subsidizing investment and operational decisions. Given the difficulty of having a precise full-picture of the reservoir, we consider that reservoir specialists deploy their best efforts in reservoir characterization, with the ultimate goal of achieving the so-called "optimal" development strategy.

The selected development strategy, however, can stray from the optimal if uncertain parameters of the reservoir turn out to be different than expected. This situation is common and recent practices in the reservoir community to address reservoir uncertainty is the adoption of probabilistic approaches — in such a way that the reservoir characteristics is expected to be within a range of possible values.

A situation in which the oil project reveals itself not to be optimal can itself turn out to be a risk experience for the Company, since it impacts its financial intentions, especially the ones expressed in the "volume" intention. Producing a certain amount of oil is directly related to the return on investment that an oil & gas project will present. *Unpredictability* and *inadequacy* are capabilities defined to express such situations.

Unpredictability refers to a characteristic of the entities related to the reservoir. Petroleum reservoirs and their constituents were generated under unknown geological circumstances, and lay buried kilometers underground. Samples and data on reservoirs are expensive, and thus scarcely accessible. Our knowledge about their size and productivity is incomplete and uncertain.

Inadequacy expresses the characteristic of certain entities in reservoir development of not having the best performance in terms of the strategy to produce a reservoir's hydrocarbon (whatever the best strategy may be). Whether it is the capacity of a pump (equipment) to pump fluids from a reservoir or the efficiency of a well in injecting fluids into the reservoir, in this work we will establish that an entity's suitability in our domain is regarding the best existent strategy to commercially produce hydrocarbon from a reservoir, regardless of our perception of what that strategy might be. So, if the pump or the well do not perform as "the ideal" pump or well would, it is considered to be inadequate.

These two characteristics correspond to BFO: Realizable Entity, in the sense that they are exhibited only through certain characteristic processes of realization. Risk Events are the processes through which certain material entities manifest their unpredictability and inadequacy.

Inadequacy = def. a *BFO: Realizable Entity* of Material Entities that participate in a Risk Event of not performing optimally in terms of the strategy to produce a reservoir's hydrocarbon (whatever the best strategy may be).

Unpredictability = def. a *BFO: Realizable Entity* of Material Entities that participate in a Risk Event of not behaving as expected.

It may seem controversial to conduct an ontological analysis in events and characteristics that might (or might not) be realized in the future, since they do not truly exist (they have not been observed in the past). But as stated in (SALES et al., 2018), "accounting for future events (...) seems to be unavoidable for any theory of risk, as uncertainty and possibility are core aspects of this concept".

We consider that the same is valid for the characteristics manifested during risk events. Even though they might not be considered as *real* properties of material entities — since they only exist as one's perceptions of these material entities — it seems unavoidable to account for expected states and characteristics. In the next subsection we analyse other characteristics of the geological entities that are realized during the reservoir development, affecting the fluid dynamics.

5.3.3 Capabilities of reservoir entities

In the case of the entities of the reservoir domain, in the course of the *Uncertain Events* they participate in, we notice that the characteristics manifested are internally grounded and inherent to those entities in such a way that if they cease to exist, the physical make up of the entities would change.

In most of the cases, especially the case for risks associated with geomechanic phenomena, the terminology of the event and the characteristic manifested is the same. In this case we define:

Collapse Potential = def. it is a *BFO:Disposition* of porous Rock or Karsts and is realized by a drastic reduction in its pores or void spaces.

Reactivation Potential = def. it is a *BFO:Disposition* that inheres in a Fracture and is realized by a change in its capability to allow fluids to pass through it under certain conditions.

Rupture Potential = def. it is a *BFO:Disposition* of Rocks and is realized by the emergence of Fractures within it under certain conditions.

In other cases, the characteristic refers to how the objects behave in respect to fluid flow. In this case we refer to (GARCIA et al., 2020b), and adopt the concept of Permeability as a disposition, since according to domain experts' analysis, it "may refer to the degree of interconnection between the void spaces in a rock or to the disposition based on such interconnectivity, which makes the rock able to allow the passage of fluids".

Similarly, the sentences of the corpus bring up the case in which geological entities behave as a seal to fluid flow, in which case we define:

Permeability = def. it is a *BFO:Disposition* of porous Earth Material and Geological Objects (or its parts) and realized by its cabability to allow fluids to pass through it.

Seal = def. it is a *BFO:Disposition* that inheres in a Geological Object (or its parts) that is its capacity to form a barrier, containing or isolating fluids from adjacent porous Geological Objects.

Having defined the ontological nature of events and characteristics, in the next section we make some considerations about the objects that participate in risk events manifesting capabilities.

5.3.4 Dealing with natural language

Now to the case of the *Objects* that participate in Risk Events, one important thing to distinguish is that we require them to be existentially independent entities. This is necessary because of the *Vulnerabilities* and *Capabilities* expressed by those objects during a risk event — which in turn are existentially dependent entities, needing a substantial entity as its bearer to exist.

To properly distinguish the metaproperties of the objects in a risk experience lead to some choices: one regarding how specialists in the reservoir domain may express the objects by the patterns (or generically dependent continuants) that occur in them, and other regarding a trait of natural language: to express objects using words that express the roles played by them. In this subsection we address this particular issue.

The GeoCore Ontology provides a sound ensemble of concepts to capture the entities of geological domain, and for this reason it serves as the base to the definition of objects that participate in risk events in the reservoir domain — and in most of the cases the subsumption relation is straight-forward.

We notice, however, that domain specialists refer to portions of Geological Objects in which specific patterns of arrangement occur. In (GARCIA et al., 2020a), those patterns are properly identified, considering that they are repeatable in nature, as different types of geological processes are repeatable along geological time. Geological Structures are these patterns of the internal arrangement of geological objects, defined as *BFO generically dependent continuants* concretized by some complex quality that inheres in the geological object that is its carrier.

In our case, we define entities that materialize specific Geological Structures, like faults and karsts. Such structures describe a pattern of internal arrangement of Geological Objects, defining its shape, permeability, and other qualities. Although ontologically speaking the Permeability Disposition itself inheres in the Rocks that constitute a Geological Object that carries some Fracture Structure, such situation it is most commonly expressed as the "fracture's permeability".

To define those specific objects is a decision that resonates with the experts' linguistic choices. To disambiguate colloquial usage of entities, we created the entities that express the Fiat Object Parts of some Geological Objects in which some Geological Structure is concretized.

Fracture = def. a *BFO:Fiat Object Part* of a *GeoCore: Geological Object* that is the carrier of some defined Fracture Structure, expressing internal arrangement qualities that concretize such Fracture Structure.

Karst = def. a *BFO:Fiat Object Part* of a *GeoCore: Geological Object* that is the carrier of some defined Karst Structure, expressing internal arrangement qualities that concretize such Karst Structure.

Facies Object = def. a *GeoCore:Rock* that constitutes a *GeoCore: Geological Object* that is the carrier of some defined *GeoReservoir:Facies*, and by extension expresses the Facies Qualities that concretize such Facies pattern.

In natural language, it is common to express an object by the role it performs. Take, for instance, a **student**. When we assign certain qualities to the student (e.g., the student's height), we know that formally the quality inheres in the *person* that performs a student role.

In our ontology, we define entities for the roles played by *GeoCore:Earth Materials* and *Wells*:

Produced Fluid = def. a *BFO:Role* that is possessed by some Amount of Fluid to be extracted from within a Reservoir.

Injected Fluid = def. a *BFO:Role* that is possessed by some Amount of Fluid to be inserted in a Reservoir.

Production Well = def. a *BFO:Role* that is possessed by a certain Well properly equipped to obtain Oil or Natural Gas from a Reservoir.

Injection Well = def. a *BFO:Role* that is possessed by a certain Well properly equipped insert fluids in a Reservoir.

This will have an impact on how the ontology is applied in the sentences that constitute our corpus. For pragmatic reasons, when one expresses an event of *decrease in produced oil ratio*, we will adopt an annotation scheme that identifies both the objects and the role performed by them.

To exemplify, an intuitive annotation scheme would be of the form:

*decrease*_[LOSS_EVENT] *in produced oil*_[ROLE] *ratio*.

but, in our case, the adopted annotation scheme will be of the form:

*decrease*_[LOSS_EVENT] *in produced*_[ROLE] *oil*_[OBJECT] *ratio*.

In the next section, we present the ResRiskOnto, in the form of taxonomical trees along with an extensive list of all the terms' definitions and their formalization.

5.4 Ontology Formalization

ResRiskOnto is an Application-level ontology that is the conjunction of fundamental concepts in Risk Analysis along with an evaluation of the main processes in the Reservoir Domain. What leaded this ontology development was the necessity to determine which are the risk events in the reservoir domain, and what roles reservoir entities play in those events. It is the result of a deep semantic analysis conducted in documents that synthesize a decade of Risk Assessment workshops in Petrobras' oil & gas projects. In our model, we adopt a experiential view of the risk, constituted by a chain of *Events*. A *Risk Event* *has participant* some *Object* and *manifests* the *Quality* that *inheres in* this Object.

Risk Events unfold through time and are perceived through the behavior of the entities that *participate in* them (Figure 5.2). ResRiskOnto specifies 9 different *Loss Events* (Figure 5.3), 6 *Source Events* (Figure 5.4) and 9 *Uncertain Events* (Figure 5.5)

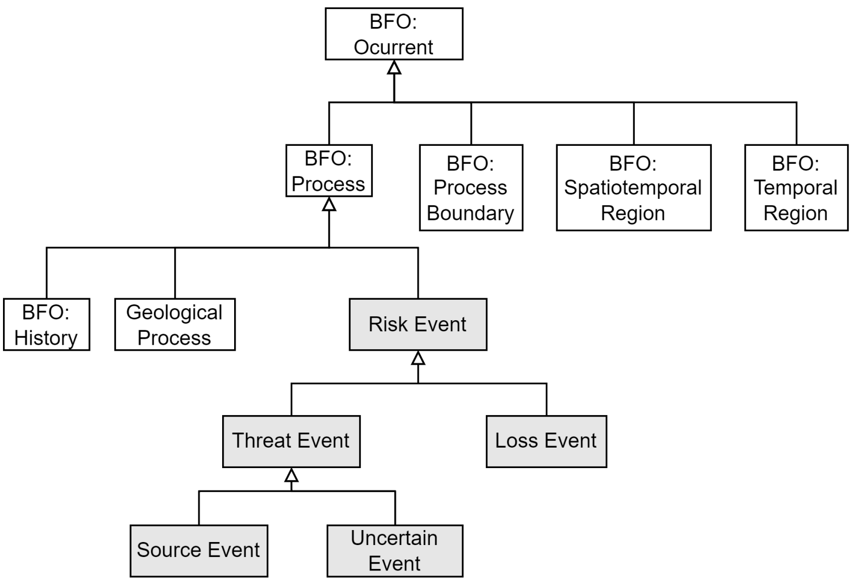


Figure 5.2: Risk Events

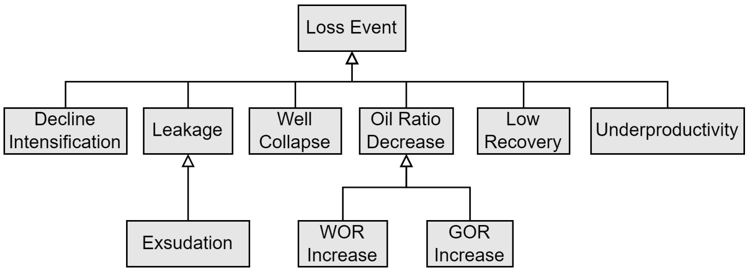


Figure 5.3: Loss Events in the Reservoir domain

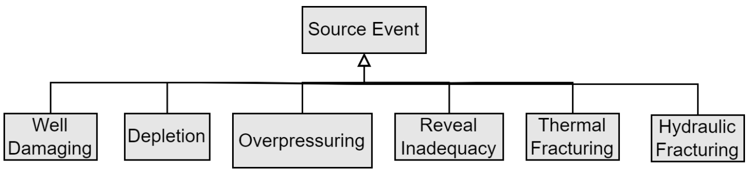


Figure 5.4: Source Events in the Reservoir domain

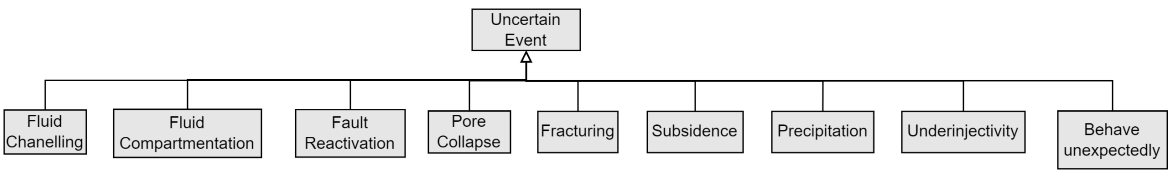


Figure 5.5: Uncertain Events in the Reservoir domain

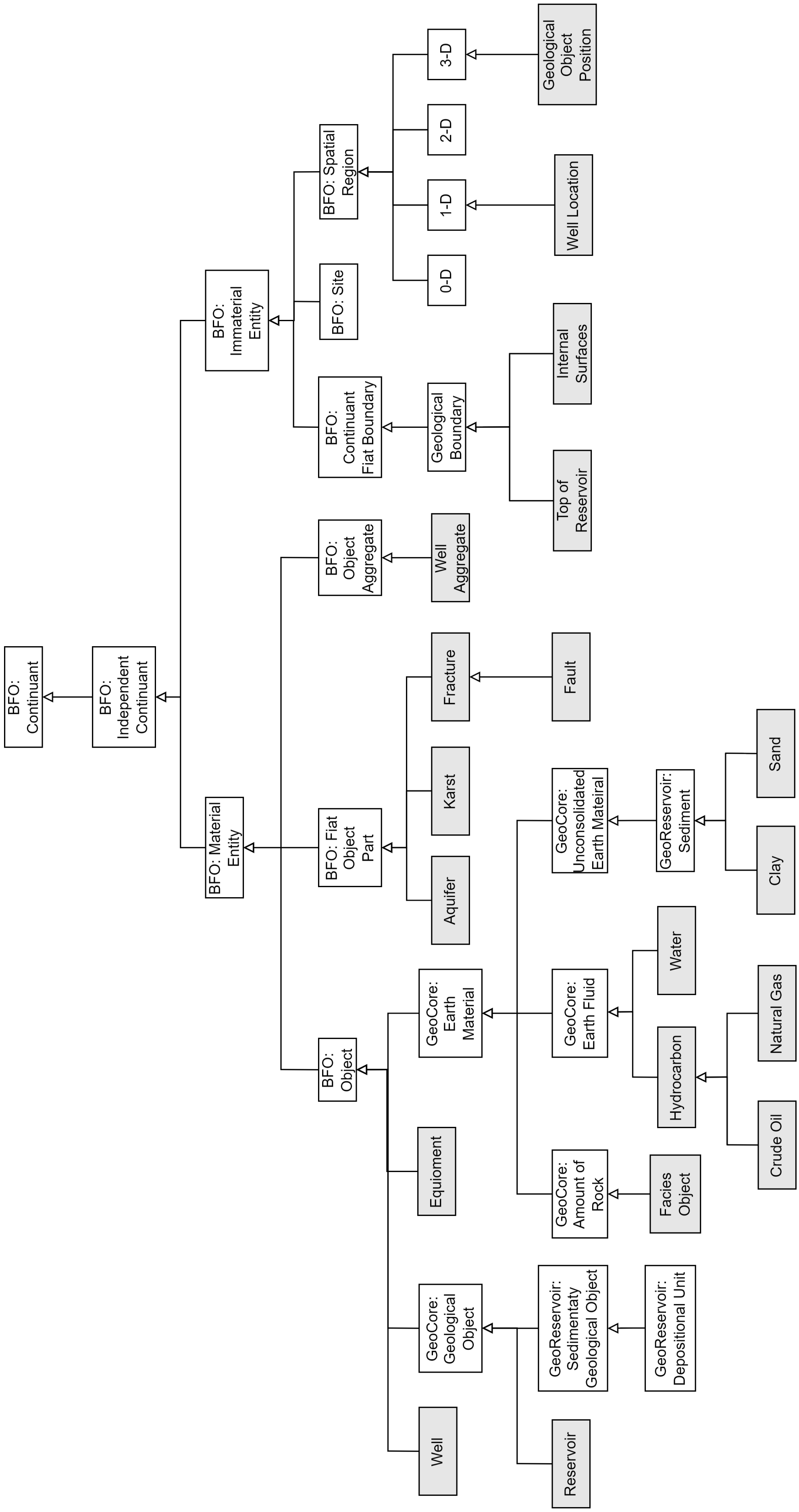


Figure 5.6: Objects in the Reservoir domain that participate in Risk Events

Objects participate *Risk Events*, manifesting either the capability associated to threats to or the vulnerabilities that relate to losses. ResRiskOnto specifies 19 different *Risk Objects* in the Reservoir Domain (Figure 5.6).

Objects are frequently expressed by the roles they perform. Taking natural language peculiarities into account, ResRiskOnto specifies 12 roles of earth materials and wells (Figure 5.7).

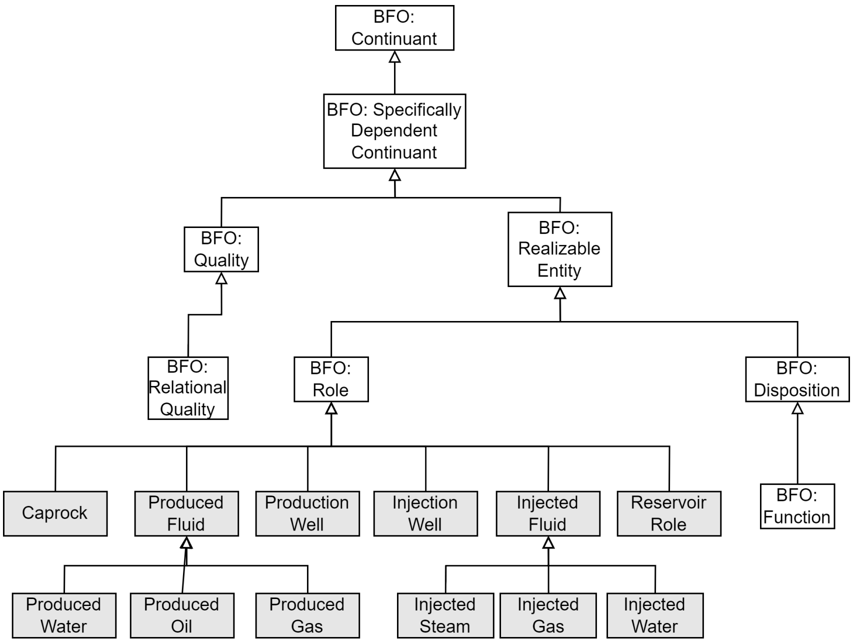


Figure 5.7: Roles in the Reservoir domain

Finally, certain *Qualities* of the objects are manifested as the Risk Event unfolds. In BFO, qualities that are exhibited in a particular manifestation, functioning or process that occurs under certain circumstances are realizable entities (Figure 5.8). Due to the approach based on a corpus of risks, not all the characteristics are defined in ResRiskOnto. Some characteristics, especially in the case of risk *Vulnerabilities*, remain implicit in the sentences.

For the purposes of correctly interpreting risk sentences, two other types of entities are present in ResRiskOnto: the ones related to reservoir modelling (Figure 5.9) and the qualities of reservoir objects those that, in contrast to roles and dispositions, do not require any further process in order to be realized (Figure 5.10).

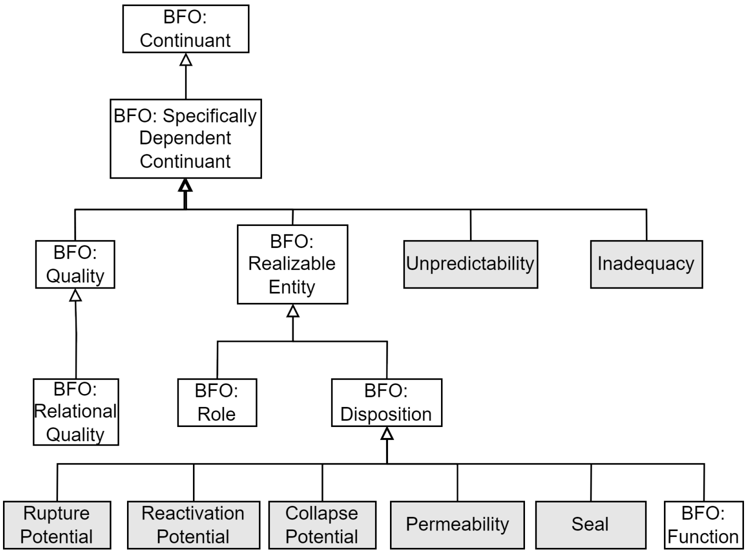


Figure 5.8: Characteristics expressed during Risk Events in the Reservoir domain

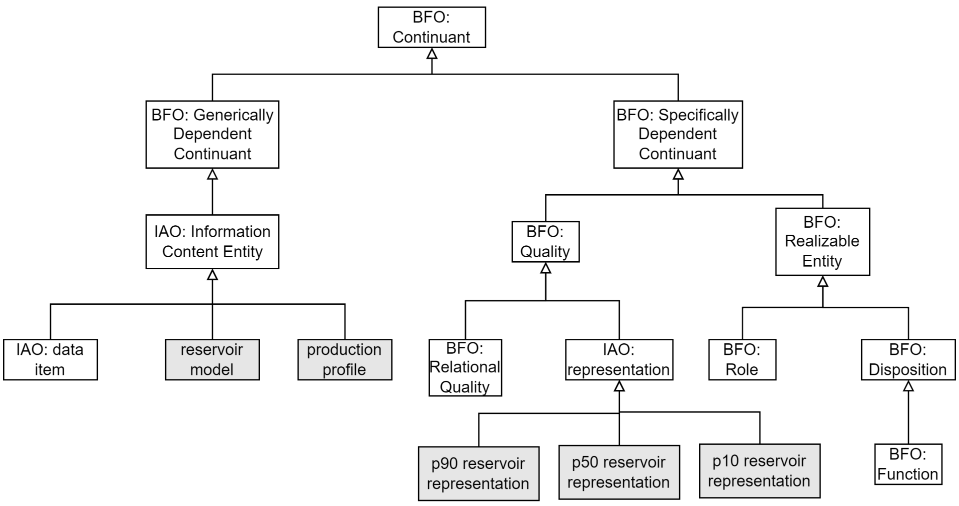


Figure 5.9: Entities of reservoir modelling

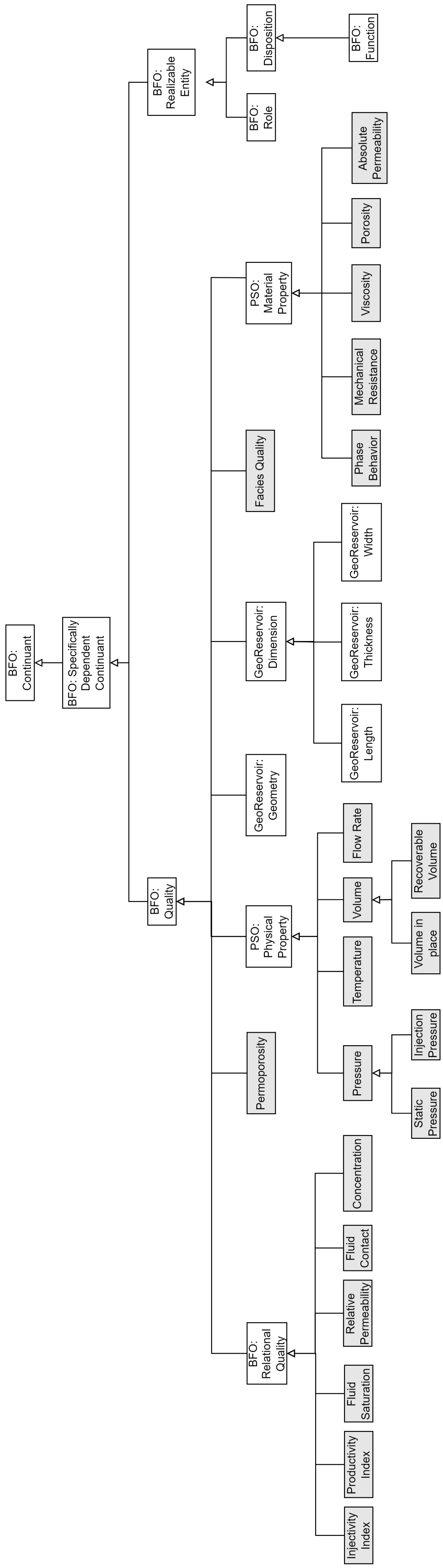


Figure 5.10: Characteristics in the Reservoir domain

In the next subsections, we provide natural language and formal definitions of all the classes that compose the ResRiskOnto Ontology.

5.4.1 Relations

Relations adopted in our ontology are those defined in the BFO and GeoCore, except to the case of the manifestation and causality relations.

BFO:participates_in = def. a relation between a continuant and a process, in which the continuant is somehow involved in the process.
Domain: Continuant
Range: Occurrent

The participation relation happens, in the case of risk experience, between the events that constitute the risk and the participants ("objects") for each of these events.

BFO:characteristic_ of = def. specifically dependent continuant (the characteristic) and any other entity (the bearer), in which the characteristic depends on the bearer for its existence.

This definition is adopted in ResRiskOnto to the case of characteristics that are exhibited in processes that occur under certain conditions — in our case, the characteristics that are expressed through the risk experience. In the case of reservoir characteristics that are fully exhibited in their bearers we use the more restricted relation **quality_of**.

manifested_by = def. a relation between a Realizable Entity is characteristic of some Independent Continuant and the Risk Event through which this characteristic is exhibited.
Domain: *BFO:Specifically Dependent Continuant*
Range: Risk Event

causes = def. the relation of causality between one Risk Event and its previous Risk Event, in the case the former occurs because of the latter's occurrence.
Domain: Threat Event
Range: Risk Event

The relation of causality expresses the chain of events that characterizes the risk experience. However, an evaluation of the risk corpus shows that it is rather confusing to specify which events are bonded by this relation. For practical purposes we allow three types of causalities: the one between two different *Loss Events*, the causality between two different *Threat Events* and finally between a *Threat Event* and a subsequent *Loss Event*.

5.4.2 Events in the Reservoir-related Risk Experience

1. **Risk Event = def.** a *BFO:Process* that either (i) has the potential of causing a loss or (ii) impact intentions of a subject in a negative way.
   Relation: *has participant* at least one *BFO:Independent Continuant* and *manifests* a *BFO:Realizable Entity* *characteristic of* the Independent Continuant.

2. **Loss Event = def.** a *Risk Event* that impact intentions of a subject in a negative way.
   Relation: *causes* only another *Loss Event*.

3. **Threat Event = def.** a *Risk Event* has the potential of causing a loss.
   Relation: *causes* only a *Loss Event* or another *Threat Event*.

4. **Source Event = def.** a *Threat Event* that has participants entities with characteristics that are controllable by human action.

5. **Uncertain Event = def.** a *Threat Event* that has participants entities with characteristics that are non-controllable by human action.

6. **Decline Intensification = def.** a *Loss Event* that expresses a declining Flow Rate behavior of the Produced Fluids more severe than expected considering a certain production strategy.
   Relation: *has participant* some Hydrocarbon that by its turn *has role* only Produced Fluid.

7. **Leakage = def.** a *Loss Event* in which a certain amount of *GeoCore:Earth Fluid* leaks from the Reservoir or Production structure.
   Relation: *has participant* some *GeoCore:Earth Fluid*.

8. **Exudation = def.** a *Loss Event* in which some *GeoCore:Earth Fluid* flow from the Reservoir through a fracture in the Rocks overlying the Reservoir.

9. **Oil Ratio Decrease = def.** a *Loss Event* that expresses a declining Ratio of Produced Oil in relation to other Produced Fluids from a Reservoir considering a certain production strategy.
   Relation: *has participant* some *GeoCore:Earth Fluid* that by its turn *has role* only Produced Fluid.

10. **GOR Increase = def.** a *Oil Ratio Decrease* that expresses an increasing Ratio of Produced Gas in relation to the Produced Oil from a Reservoir considering a certain amount of effort in production.
    Relation: *has participant* some Gas that by its turn *has role* only Produced Gas.

11. **WOR Increase = def.** a *Oil Ratio Decrease* that expresses an increasing Ratio of Produced Water in relation to the Produced Oil from a Reservoir considering a certain amount of effort in production.
    Relation: *has participant* some Water that by its turn *has role* only Produced Water.

12. **Recovery Reduction = def.** a *Loss Event* expressing smaller amounts of the total volume of recovered Oil from a Reservoir than expected considering a certain production effort.
    Relation: *has participant* some Hydrocarbon that by its turn *has role* only Produced Fluid.

13. **Underproductivity = def.** a *Loss Event* expressing lower Flow Rates of Produced Fluid from a Reservoir than expected considering a certain production strategy.
    Relation: *has participant* some Hydrocarbon that by its turn *has role* only Produced Fluid.

    There is some debate on whether underproductivity should be expressed as an Uncertain Event — manifesting a disposition of the reservoir of allowing low fluid flow through it — as opposed to a Loss Event. Our choice, in this case, was merely based on the corpus analysis. The expression "lower oil production than expected" is rather common as an expression of risk impact, and experts advocate that it represents a different impact than Reduced Recovery and Decline Intensification. To illustrate the difference, we could consider two scenarios. In the first scenario, flow rates are low but production is sustained over a long period of time — thus resulting in the same recoverable volume. In the second, the flow rate would be as expected in the beginning of the production (which is not a case of underproductivity, than declining vigorously after a certain period of time.

14. **Well Collapse = def.** a *Loss Event* in which the structure of pipes that compose a Well are severely deformed.
    Relation: *has participant* some Well and *manifests* the Mechanical Resistance of the Well.

15. **Depletion = def.** a *Source Event* in which the amount of Produced Fluids in relation to the Injected Fluids leads to a decrease of pressure in portions of the Reservoir.
    Here is another case of a definition based on the common words of the risk corpus. While depletion is the exhaustion of a reservoir and is observed by its pressure drop, in the case of the risk sentences what is expressed as depletion is the inadequacy of some production strategy — which by its turn results in the depletion. Although our definition is confusing in relation to cause-and-effect words, we opted to stay as close as possible to corpus usage of the word depletion (which is always an event caused because of factors controllable by human action).

16. **Overpressuring = def.** a *Source Event* in which the amount and pressure of Injected Fluids in relation to the Produced Fluids lead to an increase of pressure in portions of the Reservoir.

17. **Hydraulic Fracturing = def.** a *Source Event* in which the injection fluids with certain characteristics (other than temperature) into an Injection Well leads to the induction of fractures in the Rock surrounding the well.

18. **Thermal Fracturing = def.** a *Source Event* in which the temperature of injected fluids leads to the induction of fractures in the *GeoCore:Rock* surrounding the well.

19. **Well Damaging = def.** a *Source Event* in which the characteristics of drilling fluids lead to a damage in Well productivity or injectivity.

20. **Reveal Inadequacy = def.** a *Source Event* in which the parameters chosen in production and injection efforts reveal to be Inadequate.
    Relation: *manifests* Inadequacy.

21. **Chanelling = def.** an *Uncertain Event* that is the accentuated flow of some *GeoCore:Earth Fluid* through parts of a Reservoir with greater Permeability.

Relation: *has participant* some *GeoCore:Geological Object* (or its parts) or the *GeoCore:Rock* that *constitutes* the *GeoCore:Geological Object*, and *manifests* the Permeability disposition of the object.

22. **Compartmentation = def.** an *Uncertain Event* that is the retention of some *GeoCore:Earth Fluid* within the barriers of a Reservoir.
    Relation: *has participant* some *GeoCore:Geological Object* (or its parts) or the *GeoCore:Rock* that *constitutes* the *GeoCore:Geological Object*, and *manifests* the Seal disposition of the object.

23. **Underinjectivity = def.** an *Uncertain Event* expressing smaller amounts of Injected Fluid into a Reservoir than expected considering a certain injection effort.
    Contrary to Underproductivity, in this case we consider Underinjectivity as an Uncertain Event because it does not necessarily represents an impact on the intentions of our risk subject. Oil & gas development projects don't have as ultimate goal the injection of a certain amount of fluid. However, the underinjectivity may cause the production to be inferior than expected.

24. **Precipitation = def.** an *Uncertain Event* in which an amount of *GeoCore:Earth Material* initially solved in some *GeoCore:Earth Fluid* in the Reservoir deposits as a solid.
    Relation: *has participant* some *GeoCore:Earth Material*.

25. **Fracturing = def.** an *Uncertain Event* that is the realization of the Rupture Potential of an Amount of *GeoCore:Rock*.
    Relation: *has participant* some *GeoCore:Rock* and *manifests* the Rupture Potential of the Rock.

26. **Reactivation = def.** an *Uncertain Event* that is the realization of the Reactivation Potential of a Fracture.
    Relation: *has participant* some Fracture and *manifests* the Reactivation Potential of the Fracture.

27. **Pore Collapse = def.** an *Uncertain Event* that is the realization of the Collapse Potential of a Karst or Amount of *GeoCore:Rock*.
    Relation: *has participant* some *GeoCore:Rock* or Karst and *manifests* the Collapse Potential of their pores.

28. **Subsidence = def.** an *Uncertain Event* that is the sudden sinking or gradual downward settling of the ground's surface with little or no horizontal motion, as a result of Pore Collapse.

29. **Behave Unexpectedly = def.** an *Uncertain Event* in which some *BFO:Material Entity* of the Reservoir domain reveal different characteristics and behaviors than expected.
    Relation: *manifests* only Unpredictability.

5.4.3
Objects (and their roles) in the Reservoir-related Risk Experience

30. **Well = def.** an *BFO:Object* characterized by a hole drilled in the ground, properly equipped to reach the Reservoir depth.

31. **Production Well = def.** a *BFO:Role* that is possessed by a certain Well properly equipped to obtain Oil or Natural Gas from a Reservoir.

32. **Injection Well = def.** a *BFO:Role* that is possessed by a certain Well properly equipped to insert fluids in a Reservoir.

33. **Equipment = def.** an *BFO:Object* generically described as pumps, valves and other machinery or infrastructure used in the production of Oil and Natural Gas.

34. **Reservoir = def.** a *GeoReservoir:Sedimentary Geological Object* that possesses the Reservoir Role. **Relation:** *has role* Reservoir Role. Being a Reservoir is a temporary, externally-grounded condition of certain Geological Objects that are constituted by Hydrocarbon. A Geological Object with Hydrocarbon may, under certain economical conditions, cease to be a Reservoir — but would still be a Geological Object (take the recent interest over oil shale, for example). Taking into account that reservoir is the third most frequent word in our corpus (not considering the so-called stop words), we note that the perception that the reservoir community has about reservoirs is really special. Reservoirs are perceived as existentially independent entities, carrying identity and unity. For that reason we defined the Reservoir entity as the Geological Object that possesses the Reservoir Role.

35. **Reservoir Role = def.** a *BFO:Role* possessed by *GeoCore:Geological Objects* composed of a porous Amount of *GeoCore:Rock* or *GeoCore:Unconsolidated Earth Material* that store Water and Hydrocarbon in a certain condition that is of commercial interest.

36. **Caprock = def.** a *BFO:Role* possessed by *GeoCore:Geological Objects* that forms a barrier or seal above and around a Reservoir so that fluids cannot migrate beyond it.

37. **Water = def.** a naturally occurring *GeoCore:Earth Fluid* that is constituted by molecules containing one oxygen and two hydrogen atoms each, connected by covalent bonds.

38. **Hydrocarbon = def.** a naturally occurring *GeoCore:Earth Fluid* that is generated by some Geological Process and is an organic chemical compound constituted mainly by molecules composed by carbon and hydrogen atoms, that can eventually contain oxygen, nitrogen and sulfur atoms. **Relation:** *generated by* some *GeoCore:Geological Process*

39. **Gas = def.** a mixture of *Hydrocarbon* that is in gaseous form, either associated or not to Oil when contained into the Reservoir.

40. **Oil = def.** a liquid mixture of *Hydrocarbon* compounds.

41. **Injected Fluid = def.** a *BFO:Role* that is possessed by some Amount of *GeoCore:Earth Fluid* to be inserted in a Reservoir.

42. **Injected Gas = def.** the *Injected Fluid* that is composed mainly by Natural Gas.

43. **Injected Water = def.** the *Injected Fluid* that is composed mainly by Water.

44. **Injected Steam = def.** the *Injected Fluid* that is composed mainly by Steam.

45. **Produced Fluid = def.** a *BFO:Role* that is possessed by some Amount of *GeoCore:Earth Fluid* to be extracted from within a Reservoir.

46. **Produced Gas = def.** the *Produced Fluid* that is composed mainly by Natural Gas.

47. **Produced Water = def.** the *Produced Fluid* that is composed mainly by Water.

48. **Produced Oil = def.** the *Produced Fluid* that is composed mainly by Oil.

49. **Clay = def.** the *GeoReservoir:Sediment* constituted by silt-sized grains.

50. **Sand = def.** the *GeoReservoir:Sediment* that is constituted by detrital grains finer than gravel and coarser than silt.

51. **Aquifer = def.** an *BFO:Fiat Object Part* of a Reservoir constituted by an Amount of *GeoCore:Rock* that is fully saturated with Water.
Relation: *part of* some Reservoir.

52. **Fracture = def.** an *BFO:Fiat Object Part* of a *GeoCore:Geological Object* that is the carrier of some defined Fracture Structure, expressing internal arrangement qualities that concretize such Fracture Structure.
Relation: *part of* some *GeoCore:Geological Object.*

53. **Fault = def.** an *Fracture Object* of which the Fracture Structure is planar — thus being a Fault Structure.
Relation: *part of* some *GeoCore:Geological Object.*

54. **Karst = def.** an *BFO:Fiat Object Part* of a *GeoCore:Geological Object* that is the carrier of some defined Karst Structure, expressing internal arrangement qualities that concretize such Karst Structure.
Relation: *part of* some *GeoCore:Geological Object.*

55. **Fracture Structure = def.** an *GeoCore:Geological Structure* that is concretized by a topological mechanical discontinuity in one or several connected *GeoCore:Geological Objects* (GARCIA et al., 2020b).
Relation: *generically depends on* some Fracture.

56. **Fault Structure = def.** an *Fracture Structure* that is approximately planar and is concretized by a displacement between two *GeoCore:Geological Objects* or two *BFO:Fiat Object Parts* of a *GeoCore:Geological Object* (GARCIA et al., 2020b).
Relation: *generically depends on* some Fault.

57. **Karst Structure = def.** an *GeoCore:Geological Structure* that is generated by underground drainage systems and realized by sinkholes and caves, and is generated by a *GeoCore:Geological Process* of dissolution of soluble Amounts of *GeoCore:Rock.*
Relation: *generically depends on* some Karst.

58. **Well Aggregate = def.** an *BFO:Object Aggregate* that has member parts Wells with different configurations, altogether drilled with the objective of extracting Hydrocarbon from a Reservoir. Here is another definition based on words that are frequently used in the risk corpus. Experts may refer to the ensemble of Wells that compose the production strategy of a reservoir.

59. **Well Pattern = def.** an *BFO:Generically Dependent Continuant* that describes the arrangement of Wells and Pipes in relation to a given Reservoir. The Well Pattern refer to the understanding of domain experts that a Well Aggregate might carry a pattern of arrangement in relation to a Reservoir — like the 5-spot pattern for onshore fields or, in the case of Brazilian Pre-Salt, the alignment of production wells with the top of the reservoir formation. A Well Aggregate may carry a Well Pattern.

60. **Top of Reservoir = def.** the upper *GeoCore:Geological Boundary* of a Reservoir.

61. **Internal Reservoir Surfaces = def.** the internal *GeoCore:Geological Boundary* located at the top or bottom of the different Depositional Units within a Reservoir.

5.4.4
Characteristics manifested during the Reservoir-related Risk Experience

62. **Inadequacy = def.** an *BFO:Realizable Entity* that specifically-depend on the entities that participate in a Risk Event, of not performing optimally in terms of the strategy to produce a reservoir's hydrocarbon (whatever the best strategy may be).
Relation: *manifested by* some Risk Event.

63. **Unpredictability = def.** an *BFO:Realizable Entity* that specifically-depend on the entities that participate in a Risk Event, of not behaving as expected.
Relation: *manifested by* some Risk Event.

64. **Collapse Potential = def.** a *BFO:Disposition* of porous Amount of *GeoCore:Rock* or Karsts and is realized by a drastic reduction in its pores or void spaces.
Relation: *manifested by* some Pore Collapse.

65. **Reactivation Potential = def.** a *BFO:Disposition* that inheres in a Fracture and is realized by a change in its Permeability Disposition under certain conditions.
Relation: *manifested by* some Reactivation.

66. **Rupture Potential = def.** a *BFO:Disposition* of an Amount of *Geo-Core:Rock* that is realized by the emergence of Fractures within it under certain conditions
Relation: *manifested by* some Fracturing.

67. **Permeability = def.** a *BFO:Disposition* of *GeoCore:Earth Material* and *GeoCore:Geological Objects* (or its parts) and realized by its capability to allow fluids to pass through it (GARCIA et al., 2020b).

68. **Seal = def.** a *BFO:Disposition* that inheres in a *GeoCore:Geological Object* (or its parts) that is its capacity to form a barrier, containing or isolating fluids from adjacent porous Geological Objects.

69. **Mechanical Resistance = def.** a *BFO:Disposition* of a *BFO:Material Entity* that is its capacity of suffering external forces without deforming.

5.4.5 Qualities in the Reservoir domain

70. **Material Property = def.** a *BFO:Quality* that can be measured to identify the physical characteristics of a *BFO:Material Entity* (CHEONG; BUTSCHER, 2019).
    Relation: *quality of* only *BFO:Material Entity*.

71. **Porosity = def.** a *PSO:Material Property* of porous *GeoCore:Earth Materials* that is the ratio of the volume of the empty spaces and the total volume of the material.

72. **Absolute Permeability = def.** a *PSO:Material Property* of a porous *GeoCore:Earth Material* that is the interconnectivity between its void spaces.

73. **Permoporosity = def.** a *PSO:Material Property* that is the amount and interconnectivity of pores (or void spaces) within it. Reservoir professionals are interested in the characteristics that make up for the commercial interest of the Hydrocarbon stored in the reservoir. In the case of the reservoir rocks, the porosity and permeability are usually expressed as "permoporosity" — representing at the same time the amount of hydrocarbon stored per rock volume and the ability of the rock to allow fluid flow. It is a term generically used to represent the quality of the reservoir's rock — probably because of the correlation between porosity and permeability that exists for most of the sandstones.

74. **Phase Behavior = def.** a *PSO:Material Property* of an *GeoCore:Earth Material* and describes the complex interaction between physically distinct, separable portions of matter called phases that are in contact with each other.

75. **Viscosity = def.** a *PSO:Material Property* of a *GeoCore:Earth Fluid* and is the measure of its resistance to deformation at a given rate. that inheres in an *BFO:Material Entities*, defined by a scalar that measures the action of forces applied over the *BFO:Continuant Fiat Boundary* of such Entities.

76. **Physical Property = def.** a *BFO:Quality* that determines the physical state of a *BFO:Material Entity*, and can be measured as quantitative values based on some measurement units (CHEONG; BUTSCHER, 2019).
    Relation: *quality of* only *BFO:Material Entity.*

77. **Flow Rate = def.** a *PSO:Physical Property* of a fluid *BFO:Material Entity* that is the Volume of Fluid in movement that passes per unit time across a given section, expressed by a scalar.

78. **Pressure = def.** a *PSO:Physical Property* of *BFO:Material Entities*, defined by a scalar that measures the action of forces applied over the *BFO:Continuant Fiat Boundary* of such Entities.

79. **Injection Pressure = def.** the *Pressure* applied to Injected Fluids in the interior of a Injector Well.

80. **Static Pressure = def.** a *Pressure* that inheres in a Reservoir when its fluids are in stationary condition.

81. **Temperature = def.** a *PSO:Physical Property* of *BFO:Material Entities* and is a physical quantity that expresses hot and cold.

82. **Volume = def.** a *PSO:Physical Property* of *BFO:Material Entities*, and is a scalar expressing the three-dimensional space occupied by this entity.

83. **In Place Volume = def.** a *Volume* of an *GeoCore:Earth Fluid* within a Reservoir before production.

84. **Facies Quality = def.** a *BFO:Quality* within a *GeoCore:Geological Object* that realizes a given *GeoReservoir:Facies.*

85. **Clay Content = def.** a *BFO:Relational Quality* that is the proportion of Clay in a Sandstone.
    Relation: *quality of* only Clay and *GeoReservoir:Sedimentary Rock.*

86. **Concentration = def.** a *BFO:Relational Quality* that inheres in a mixture of two or more *GeoCore:Earth Materials*, with at least one

solvent and at least one solute, that measures the abundance of a constituent divided by the total volume of a mixture.
Relation: *quality of* min 2 *GeoCore:Earth Material*.

87. **Fluid Contact = def.** a *BFO:Relational Quality* that is formed between two fluid phases within a Reservoir that are adjacent and segregated by different densities, where they are externally connected to each other.
Relation: *quality of* exactly 2 *GeoCore:Earth Fluid*.

88. **Injectivity Index = def.** a *BFO:Relational Quality* of an Injection Well, its adjacent Amount of *GeoCore:Rock* and the *GeoCore:Earth Fluid* in a Reservoir, and expresses the ability of the rock to receive fluids (in the role of Injected Fluid) from the wellbore.
Relation: *quality of* only Injection Well or *GeoCore:Rock* or *GeoCore:Unconsolidated Earth Material* or *GeoCore:Earth Fluid*.

89. **Productivity Index = def.** a *BFO:Relational Quality* of a Production Well, its adjacent Amount of *GeoCore:Rock* and the *GeoCore:Earth Fluid* in a Reservoir, and expresses the ability of the rock to deliver fluids (in the role of Produced Fluid) to the wellbore.
Relation: *quality of* only Production Well or *GeoCore:Rock* or *GeoCore:Unconsolidated Earth Material* or *GeoCore:Earth Fluid*.

90. **Relative Permeability = def.** a *BFO:Relational Quality* expressed by a dimensionless measure of the effective permeability of a porous *GeoCore:Earth Material* to a given phase of a multiphase *GeoCore:Earth Fluid*.
Relation: *quality of* only *GeoCore:Earth Fluid* or *GeoCore:Rock* or *GeoCore:Unconsolidated Earth Material*.

91. **Fluid Saturation = def.** a *BFO:Relational Quality* expressed by a dimensionless measure of the fraction of the pore volume of a porous *GeoCore:Earth Material* occupied by a given *GeoCore:Earth Fluid*.
Relation: *quality of* only *GeoCore:Earth Fluid* or *GeoCore:Rock* or *GeoCore:Unconsolidated Earth Material*.

5.4.6 Entities of Reservoir Modelling

92. **Reservoir Model = def.** a *IAO:Information Content Entity* that is about the material entities that generated reservoir data, and is generated by combining such data with a reservoir representation.

93. **Production Profile = def.** a *IAO:Information Content Entity* generated by the reservoir model and is about the simulated production profile of the fluids according to the reservoir model.

94. **Representation = def.** a *BFO:Quality* which is about or is intended to be about a PORTION OF REALITY (POR) (CEUSTERS; SMITH, 2015).

95. **P90 Reservoir Scenario = def.** a *Representation* that is about collected data from the Reservoir interpreted in a way that the combination of unknown reservoir variables have 90% of cumulative probability of occurrence for a given objective function in an uncertainty analysis.

96. **P50 Reservoir Scenario = def.** a *Representation* that is about collected data from the Reservoir interpreted in a way that the combination of unknown reservoir variables have 50% of cumulative probability of occurrence for a given objective function in an uncertainty analysis.

97. **P10 Reservoir Scenario = def.** a *Representation* that is about collected data from the Reservoir interpreted in a way that the combination of unknown reservoir variables have 10% of cumulative probability of occurrence for a given objective function in an uncertainty analysis.

    A reservoir model is a simplified representation or interpretation of a fragment of the reality of the reservoir, obtained by rock and fluid samples, and indirect data (e.g. seismic data and well profiles). This scarce amount of data is interpreted according to the modeller's knowledge about the Geological Processes and geological characteristics that could generate a reality in which the reservoir is realized.

5.4.7 Ontology Implementation

We implemented our ontology in OWL language using the Protégé tool (MUSEN, 2015), a free and publicly available knowledge model editor. BFO and GeoCore have OWL implementations, allowing a direct integration with ResRiskOnto. This is made by directly importing the GeoCore Ontology (which in its turn was implemented importing the BFO Ontology) and then creating the classes and axioms of ResRiskOnto according to the formalization presented in the previous subsections.

OWL is a Semantic Web language designed to represent knowledge, and ontologies implemented in this language are interoperable. It is also the W3C standard for representing ontologies, and one of the most common languages

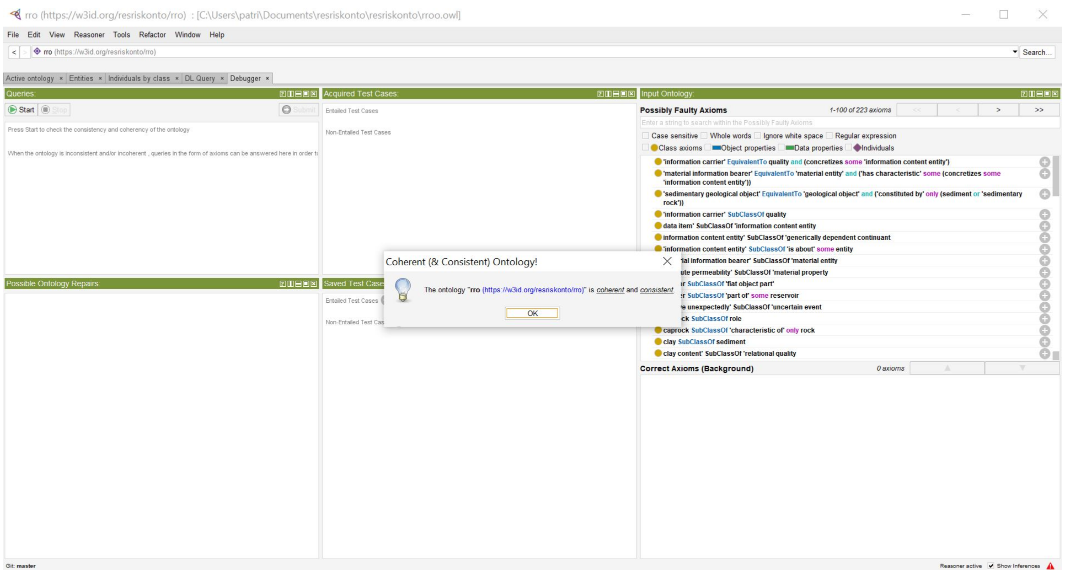


Figure 5.11: ResRiskOnto consistency checking using the OntoDebug plugin

for ontology sharing. ResRiskOnto is freely available[2], and its terms persist in the ontology's permanent URI (Uniform Resource Identifier) [3].

One of the advantages of the Protégé tool is that it offers extensions for consistency checking of the ontology. In our case, we used the Protégé plugin OntoDebug[4], that implements built-in algorithms for coherence and consistency checking. Figure 5.11 shows that ResRiskOnto is a consistent and coherent ontology.

The next Chapter presents the experiments that were conducted to validate the application of the proposed ontology in the interpretation of Petrobras' collection of reservoir-related risk sentences.

[2]https://github.com/patriciaferreiradasilva/resriskonto
[3]https://w3id.org/resriskonto/rro
[4]https://protegewiki.stanford.edu/wiki/OntoDebug

6
Validating the Ontology

In this section we present the application of the ontology in the risk corpus. In the first subsection, we describe the processes of document selection and annotation. In the last subsection, we describe the NLP tasks conducted to validate the ontology, and show the results obtained with such tasks.

6.1
Document selection and annotation

In (SILVA et al., 2021) we describe a subset containing 340 target sentences describing those risks associated with fouling and geomechanical aspects of the petroleum reservoir. This subset was used to validate the process of applying semantic analysis in the ontology building process, and provided 54 sentences, randomly selected to the annotation process.

From the remaining set that constitute our risk corpus, we selected another 498 sentences to subject to the annotation process. Those sentences were randomly selected, respecting the criteria of describing risks that had their probability and impact evaluated in the original documents. The total of 552 texts that were annotated represent roughly 20% of the size of the original corpus.

Figure 6.1 shows the distribution of the 20 most frequent words in the total set of risk sentences and in the selected sentences.

The annotation process was conducted as the means to enable the natural language processing tasks to validate the ontology. Because of the lack of standardization regarding events in reservoir-related risks in oil & gas projects, an annotation with multiple domain experts would demand an effort in training annotators that is prohibitive for the purposes of this work. For this reason, the annotation of the documents was conducted only by the author of this work. The distribution of *tokens* in annotated documents is in Figure 6.2.

Annotating is the activity of assigning labels to a text (or fragments of a text), which can be done manually or automatically. It is an expensive and a time-consuming task, since it involves the collection and preparation of texts, the definition of categories (labels), and the combination of the work of domain experts with well-controlled methods and tools.

To ensure annotation quality, a curation process should also be conducted. A curator is responsible to manage the annotation process and monitor

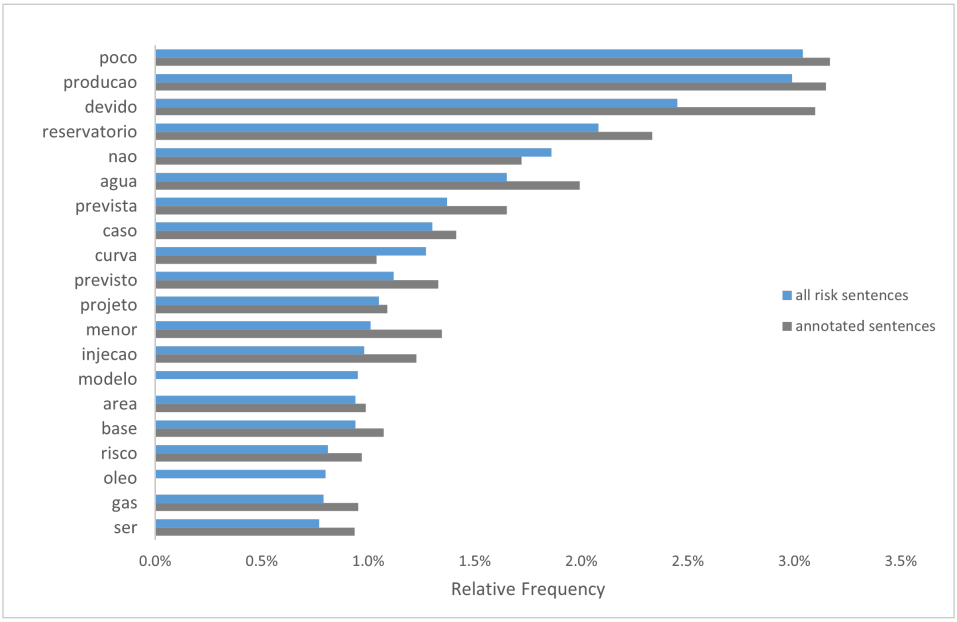


Figure 6.1: Relative frequency of most frequent words in risk sentences and selected sentences for annotation.

parameters that indicate the quality of the annotated text (such as agreement among multiple annotators).

In our annotation process, we used the ERAS tool, that will be presented in the next section.

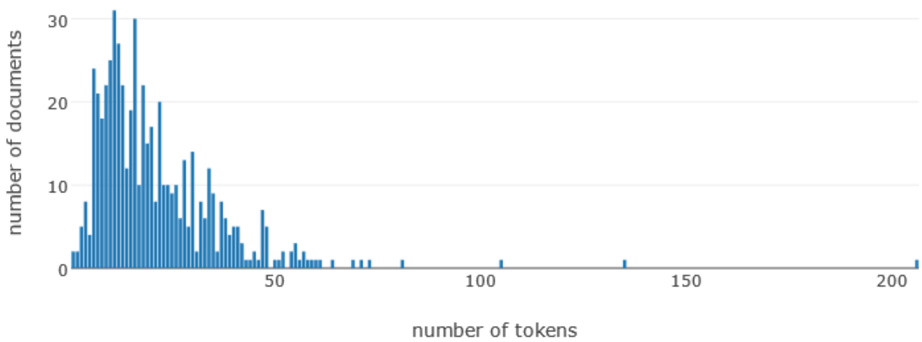


Figure 6.2: Distribution of number of tokens in annotated documents

6.2 ERAS (*Entities and Relations Annotation System*)

ERAS is a novel-based text annotation tool developed to facilitate and manage the process of text annotation (GROSMAN et al., 2020). ERAS is a freely available tool that presents the advantage of offering features to support annotation quality control, while most annotation tools require a commercial license.

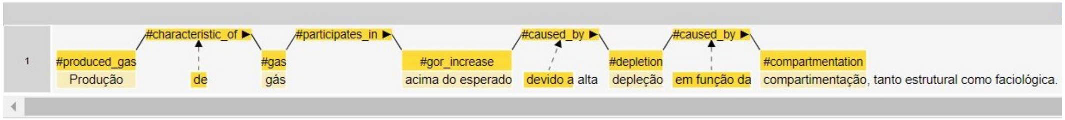


Figure 6.3: Annotation example in ERAS tool.

Considering a set of words W and a set of classes (of the ontology) C, the annotation is the process that generates an instance $e \in E$, where e is the a assignment of some class $c \in C$ to a word $w \in W$ or to a sequence of words $S \subset W$.

One example of annotated text in ERAS is the one present at Figure 6.3, where given the subset with the first part of the sentence W = {Produção, de, gás, acima, do, esperado} and the tags C = {#produced_gas, #gas, #gor_increase}, we have the set of annotations $E = \{e_1, e_2, e_3\}$:

e_1 = (⟨Produção⟩, #produced_gas);
e_2 = (⟨gás⟩, #gas); and
e_1 = (⟨acima,do,previsto⟩, #gor_increase).

Analogously, given a set E of entities and a set P of properties, the annotation process creates an instance $r \in R$, where r is an assignment of some property $p \in P$ to a pair of entities $(e_i, e_j) \in E$, and R is a set of relations.

Considering P = {(characteristic_of, (#produced_gas,#gas)), (participates_in, (#gas,#gor_increase))}, we see in Figure 6.3 the set of relations $R = \{r_1, r_2\}$, where:

r_1 = ((e_1, e_2),characteristic_of); and
r_2 = ((e_2, e_3),participates_in).

In NLP tasks, these features are useful in Named Entity Recognition (NER) and Relation Extraction (RE) tasks. Besides the annotation of entities and relations, ERAS also supports the association of some portion of text to a relation (connector annotation), an information that can be used in the feature engineering to improve the results of RE tasks.

Connector annotation creates instances of type $k \in K$, where K is a set of connectors and k is the assignment of a relation $r \in R$ to a word $w \in W$ or to a sequence of words $S \subset W$. In Figure 6.3, we observe $K = \{k_1\}$, where:

$k_1 = (\langle \text{de} \rangle, r_1)$.

Besides annotation, ERAS also supports: (i) the management of textual datasets; (ii) the use of tokenizers and POS taggers to text formatting and characterization; (iii) ontology management, uploaded in the OWL format, via the selection of a subset of entities and relations that will be available for the annotator; and (iv) management of the process by a curator, who can compare annotations form multiple users and monitor parameters (*e.g.* self-agreement and inner-annotator agreement).

In the next section, we explain some choices during the annotation process, showing a few examples of how they indicate the ontology suitability to the proposed task.

6.3 Annotation

For simplification purposes, we selected for annotation the classes of ontology representing domain and application level, excluding top-level classes that represent general entities (*e.g.* BFO:*Continuant* or BFO:*Occurrent.*). Some exceptions are the geology domain classes specified in GeoCore, as *GeoCore:Geological Object* or *GeoCore:Geological Process*

For the same reason, relations inherited from top-level and domain-level ontologies were excluded from the annotation process (*e.g.* **BFO:part_of**). The relations included in the annotation process are those described in Subsection 5.4.1. In Figure 6.4 we observe a case in which it was possible to identify clearly the relations between risk events, the objects that participate in them, and their manifested qualities.

The lack of standardization in the risk sentences become very clear during the annotation process. Because domain experts didn't have a well defined set of concepts while constructing risk sentences, we noticed sentences in which the impacts of the risk were not stated explicitly (as in Figure 6.5), or even sentences in which it was not possible to identify the words that could correspond to a risk event (as in Figure 6.6).

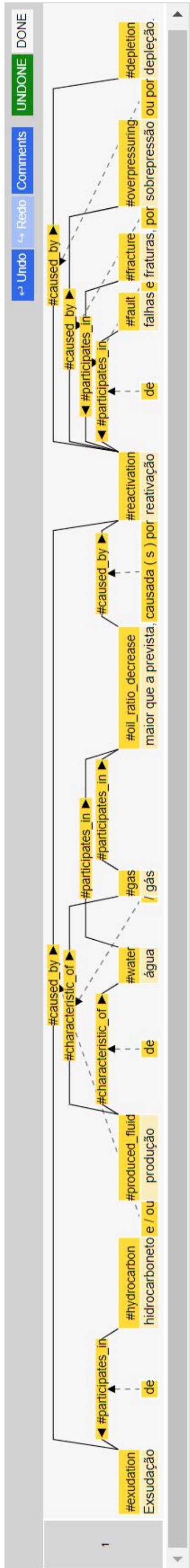


Figure 6.4: Example of annotation with emphasis on the annotated relations.

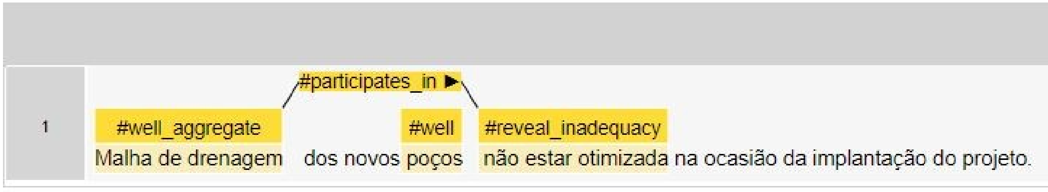


Figure 6.5: Risk sentence with Source Event identified.

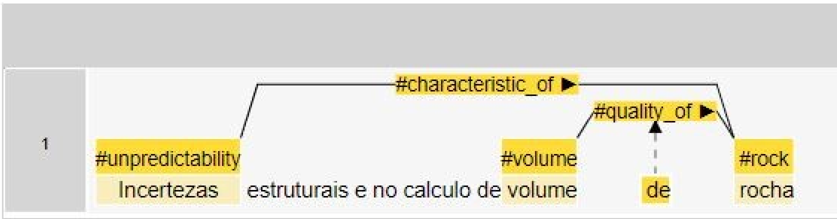


Figure 6.6: Risk sentence without the identification of events.

Another semantic choice adopted by the annotator was to consider an equivalency between the expressions "fluid production" and "produced fluid", as can de observed in Figure 6.3. The same is valid to "fluid injection" and "injection fluid", and the possible combinations with fluid instances (*e.g.* "oil", "gas", "water"). This choice is based in a perception that such expression intends to refer to the origin or the role a fluid has in a risk experience — and not to the production process itself.

Out of the 97 classes defined in ResRiskOnto, 86 were identified during text annotation. Classes not identified in the texts were mainly those that represent a concept generalization (*e.g. Risk Event* or *Physical Property*), and those that represented Generically Dependent Continuants that are carried by Objects (*e.g. Fault Structure*). Figure 6.7 present the distribution of tags in annotated sentences.

In the next section, we describe the Natural Language Processing tasks

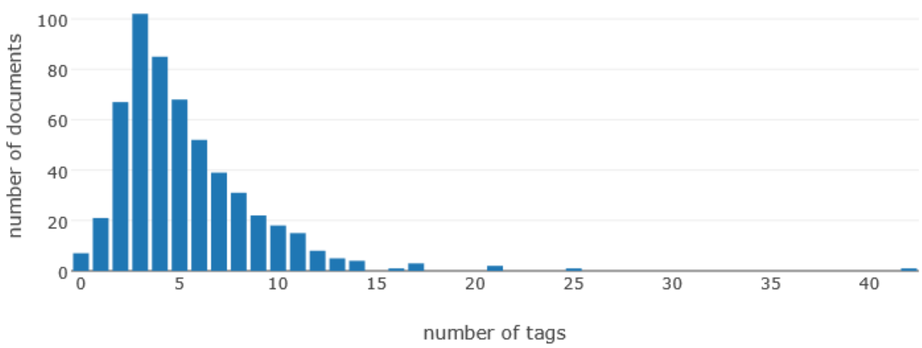


Figure 6.7: Distribution of number of tags in annotated documents

used to validate the ontology application in the interpretation of our risk sentences.

6.4 Named Entity Recognition experiments

A NER model evaluates, for each *token*, the corresponding entity to be automatically recognized. It is important to notice that the ERAS tool splits each annotated class in two tags, indicating whether it represents the beginning ("B-") or the middle of an entity ("I-"). For this reason, our annotated documents generated a total of 159 classes for NER.

Two techniques were conducted in NER tasks: Conditional Random Forest (CRF) algorithms available in the Sklearn package (PEDREGOSA et al., 2011), and Bidirectional Long-Short Term Memory strategies (HOCHREITER; SCHMIDHUBER, 1997).

In both cases, the metrics to evaluate the quality of the models were the traditional *Accuracy*, and *F1-Score* [1]:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{6-1}$$

$$Precision = \frac{TP}{TP+FP} \tag{6-2}$$

$$Recall = \frac{TP}{TP+FN} \tag{6-3}$$

$$F1-Score = \frac{2*Precision*Recall}{Precision+Recall} = \frac{2*TP}{2*TP+FP+FN} \tag{6-4}$$

Where TP stands for the True Positives, TN the True Negatives, FP for the False Positives and FN for the False Negatives in the validation set of the model.

All the experiments were conducted using resources of the SDumont supercomputer[2]. The author of this work acknowledges the National Laboratory for Scientific Computing (LNCC/MCTI, Brazil) for providing the HPC resources that have contributed to the results reported in the remainder of the section. The experiments were allocated in a computational node composed of 2 processors Intel Xeon Cascade Lake Gold 6252, and 4 Graphics Processing Units (GPU) NVIDIA Volta V100, with RAM memory of 384Gb.

[1]https://en.wikipedia.org/wiki/Precision_and_recall
[2]http://sdumont.lncc.br

6.4.1 NER task with CRF model

Conditional Random Field models are a class of statistical modelling methods, that fall into the sequence modelling family. Whereas a classifier predicts a label for a single sample without considering "neighbouring" samples, a CRF can take context into account. To do so, the predictions are modelled as a graphical model, which represents the presence of dependencies between the predictions [3].

CRFs are commonly applied in natural language processing tasks, since the prediction is dependent only on its immediate neighbours (context). It is that a CRF is an undirected graphical model whose nodes can be divided into exactly two disjoint sets $\boldsymbol{X}$ (observations) and $\boldsymbol{Y}$ (output variables), where the conditional distribution $p(\boldsymbol{Y}|\boldsymbol{X})$ is then modeled.

For the feature engineering step, we used the spaCy natural language processing package (HONNIBAL; MONTANI, 2017). For each *token*, the following features were extracted:

Features(X_i):

- Token(i): the *token* X_i;
- Simple Token (SM_i): lowercased *token* X_i;
- Token Lemma (TK_i): the canonical form of the *token* X_i;
- POS (POS_i): the *part-of-speech* tagging for the *token* X_i;
- Capitalized Token ($CAPS_I$): checks if the *token* X_i is in all capital letters;
- Title Token (T_i): checks if the *token* X_i is a title, *i.e.*, if its first character is in capital letter; igit Token (D_i): checks if the *token* X_i is all numeric;
- No-digit Token (ND_i): checks if the *token* X_i contains no numeric character;
- Beginning-of-sentence (BOS_i): checks if *token* X_i is the beginning of the sentence;
- End-of-sentence (EOS_i): checks if *token* X_i is the end of the sentence;

We also tested the sensitivity of the tags in the model to the probability and impact information that quantify a risk experience. For this reason, in the case of the 498 sentences selected that respected the criteria of describing risks that had their probability and impact evaluated, we created an extended set of features, that included:

Extended Features:

[3] https://en.wikipedia.org/wiki/Conditional_random_field

- Probability ($PROB_i$): the probability associated with the sentence in which the *token* X_i is present;
- Impact ($IMPACT_i$): the impact associated with the sentence in which the *token* X_i is present;

For each token, we also extracted $PREV(X_i, N)$ and $NEXT(X_i, N)$ that correspond, respectively, to the features of *token* in the previous N and next N position in relation to X_i.

The experiments were conducted with $PREV(X_i, N)$ and $NEXT(X_i, N)$ features with N ranging from 0 to 5. We conducted experiments with three datasets:

1. CRF-A with all the collection of 552 annotated documents, that did not include the Extended Features;
2. CRF-R with the collection of 498 that had probability and impact information, not including Extended Features for these two variables;
3. CRF-RE with the collection of 498 that had probability and impact information, with Extended Features for these two variables;

After creating the features for the annotated documents, we randomly divide the dataset in two sets, namely the TRAIN and TEST sets. The TRAIN set is used to train the models while the TEST set is used for model evaluation.

The size of TRAIN and TEST sets follow a proportion of 80%/20%. For the CRF-A experiment, we have TRAIN and TEST sets with 442 and 110 documents, respectively. CRF-R and CRF-RE experiments have each 398 and 100 documents in the TRAIN and TEST sets.

The two last experiments helped to evaluate if the features of *impact* and *probability* could improve the model results. In all experiments, the CRF model took the parameters *lbfgs* for the training algorithm (Gradient descent using the L-BFGS method), 0.1 for both L1 and L2 regularization coefficient, and 200 maximum iterations for optimization algorithms.

Table 6.1 shows the overall performance of the CRF strategy for each experiment. It is important to notice that the features representing risk impact and probability had no effect in the model performance.

Another observable result is that the models performed better without considering the features of neighbouring *tokens*. This is probably due to the fact that the ontological classes were populated with the frequent words in the corpus, which could lead to a strong correlation between classes and *token* forms.

	Accuracy						F1-Score					
	N=0	N=1	N=2	N=3	N=4	N=5	N=0	N=1	N=2	N=3	N=4	N=5
CRF-A	0.8091	0.8018	0.7995	0.7918	0.7873	0.7886	0.7850	0.7795	0.7743	0.7660	0.7610	0.7616
CRF-R	0.7857	0.7938	0.7974	0.7979	0.7847	0.7872	0.7635	0.7702	0.7727	0.7727	0.7563	0.7587
CRF-RE	0.7857	0.7938	0.7974	0.7979	0.7847	0.7872	0.7635	0.7702	0.7727	0.7727	0.7563	0.7587

Table 6.1: Results for NER tasks for experiments with Conditional Random Forest models.

The results for the Named Entity Recognition task model CRF-A with N=0 are shown in Table A.1. The overall weighted *Accuracy* and $F1 - score$ of the model are of 0.783 and 0.785, respectively.

6.4.2
NER task with Bi-LSTM

Recurrent Neural Networks (RNN) are a type of artificial neural network in which the connections of the network's nodes form a directed graph. The output for an entry X_i in RNN network depends on the entry itself and on the state of the network from the previous entry. They are broadly applied in machine learning algorithms aimed at sequential data, such as time series and texts.

Long-Short Term Memory, or LSTM is a RNN that stores information through its sequence. It is composed by a *gates* that constitute methods for storing or forgetting information through the network. Bidirectional LSTM (or Bi-LSTM) combines two LSTM networks — one that takes the input in a forward direction, and a second one taking the input in a backward direction — providing information on the surroundings of each *token*.

Coupling the output of the Bi-LSTM with a linear chain CRF is an strategy that has proven to generate good results in NER texts. The LSTM-CRF model requires less orthographic information since it gets more contextual information out of the bidirectional LSTMs (LAMPLE et al., 2016).

The LSTM network takes as input sequences of fixed length with the words of the text represented as real-valued vectors in a predefined vector space (BENGIO; DUCHARME; VINCENT, 2000). This technique, called *word embeddings* (WE), is one of the main foundations of modern NLP tasks, enabling machine learning algorithms to achieve great generalization capabilities. *Word embeddings* provide meaningful representations of words, being able to capture syntactic and semantic features based on their context.

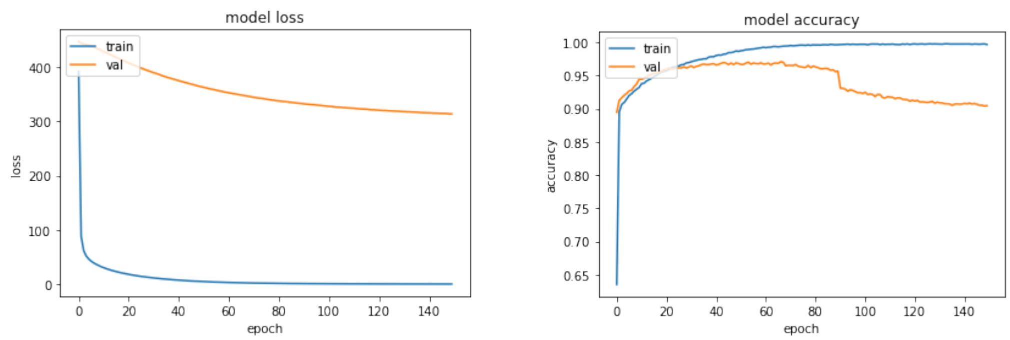
We used PetroVec, a pre-trained *word embedding* model for the specific domain of oil and gas in Portuguese (GOMES et al., 2021)[4]. PetroVec is trained using a large specialized oil and gas corpus in Portuguese, composed of

[4]https://petroles.puc-rio.ai/

an extensive collection of domain-related documents from leading institutions (CORDEIRO, 2020). From the available *word embedding* models, we chose a 100-dimension model trained using the oil & gas domain corpus along with a Petrobras dataset of reports.

Also in the case of the Bi-LSTM approach, we split the dataset in TRAIN and TEST sets in a proportion of 80%/20% (or 442 and 110 documents, respectively). The model divides the TRAIN set yet into two new sets, TRAIN and VALIDATION, in a 90%/10% proportion (398 and 44 documents, respectively). We conducted three experiments with *batch size* of 32 and 85, 150 and 200 *epochs* respectively.

Model optimization and performance learning curves for the experiment with 150 *epochs* is shown in Figure 6.8. With increasing epochs, the model tends to be overfitted to the TRAIN set without necessarily increasing performance in relation to the VALIDATION set.



(a) Loss (b) Accuracy

Figure 6.8: Model performance : (a) Loss over TRAIN and VALIDATION sets per *epoch*; (b) Accuracy in TRAIN and VALIDATION sets per *epoch*.

The results for the Named Entity Recognition task model Bi-LSTM-CRF with 85 *epochs* are shown in Table A.1. The overall weighted *Accuracy* and $F1 - score$ of the model are of 0.905 and 0.81, respectively.

While deep learning techniques perform well in large datasets, stochastic machine learning methods achieve better results in small datasets (ALOM et al., 2019). Our annotated dataset is relatively small and has very sparse classes through the sentences, which might be the reason why the Bi-LSTM-CRF approach showed results only slightly better when compared to the CRF approach.

7
Conclusion

In this work we proposed ResRiskOnto, an application ontology for reservoir-related risks in the oil & gas domain. ResRiskOnto is developed as an extension of the BFO an GeoCore ontologies, and uses the Common Ontology of Value and Risk as conceptual basis for risk descriptions. It also reuses concepts developed in the GeoReservoir Ontology and Information Artifacts Ontology.

Considering the increasing adoption of conceptual modelling in applications with non-structured data, ResRiskOnto characteristics have the advantage of facilitating ontology management and reuse for future studies in the petroleum reservoir domain.

In the next sections, we present the contributions of this work, as well as possibilities to be explored in future work.

7.1
Contributions

To build ResRiskOnto, we conducted an ontological analysis in relevant terms present in a corpus of approximately 2500 risk sentences, documented through over ten years of risk management activities in Petrobras. The result is an ontology composed of a predefined set of concepts that resonate with their area of expertise.

An important challenge to overcome in the definition and adoption of an ontology is reaching the agreement of community experts towards the proposed conceptualisation. The construction process makes ResRiskOnto an application ontology composed of the words dominated by reservoir professionals. We assume that the frequent usage of certain terms might represent an advantage for ontology adoption in future risk documentation.

The methodology adopted in the ontology's elaboration itself was documented in an article, constituting a relevant instrument for other conceptualisations that will be applied in natural language processing tasks over a body of texts.

The petroleum reservoir domain, an inherently risky activity, lacks standardization regarding the events and parameters to be considered during oil & gas project risk analysis. This is yet another contribution of ResRiskOnto, an ontology centered in the events that compose a risk experience, naming 24 specific types of risk events for the reservoir activity. *Risk Events* are divided in *Loss Events* (that impact intentions of a subject in a negative way) and *Threat Events* (that have the potential of causing a loss), in a clear cause-and-consequence fashion that resonates with well-established definitions of risk in the literature.

The ontology classes are adequately identified in a set of 552 risk assessment documents. This annotated corpus illustrates how risk sentences in natural language may not be complete regarding the expression of the uncertainties and consequences in a risk experience. An adequate conceptualisation can be useful in assuring the quality of future risk documentation.

To validate ResRiskOnto, the annotated sentences were subject to tasks of named entity extraction, that reached an overall weighted $F1-Score$ of 0.81. The trained models can be used to perform named entity extraction over the risk corpus.

Finally the competency questions, formulated to define the ontology's scope, can now be addressed:

1. What types of events compose a Risk Experience in the Petroleum Reservoir domain?
   R: ResRiskOnto defines Risk Events as the ones that either (i) Loss Events, that have the potential of causing a loss or (ii) Threat Events, that impact intentions of a subject in a negative way. Since Threat Events can have controllable non-controllable parameters, we further divide them into Source Events and Uncertainty Events. ResRiskOnto also details other 24 types of events based on the risk corpus.

2. Does the ontology properly describes at least 250 risk sentences?
   R: During the validation experiment, a total of 552 sentences were annotated — thus showing ResRiskOnto as a promising conceptual model to describe at least 250 sentences.

3. Is the performance evaluation of the NER tasks in the original corpus annotated with the ontology satisfactory?
   R: NER models showed weighted $Accuracy$ and $F1-score$ of 0.905 and 0.81, respectively. Considering the tags related to Risk Events only, the weighted $F1-Scores$ is 0.667, which is adequate to enable natural language processing tasks over reservoir risks documentation within an acceptable error range.

4. Do reservoir experts recognize the proposed risk events?
   R: ResRiskOnto is constructed with words dominated by reservoir professionals that occur frequently in risk descriptions. For this reason, we

believe that it can be easily adopted by domain experts. However, as mentioned, the community commitment over a set of concepts should be the object of future work.

7.2 Limitations and future research

ResRiskOnto provides a complex conceptualisation (29 classes of Risk Events, 97 total classes), one that may not be easily learned, even by domain experts. Besides, the ever changing characteristic of technologies may also render a certain risk event obsolete, while other types of risk events may surge - thus demanding continuous update of the ontology to guarantee its suitability. Such complexity and necessity for update constitute a barrier to the ontology adoption.

For this reason, having provided the means to standardize risks in the reservoir domain, we suggest future work to be focused in the needs of domain experts, using the ontology to organize and extract meaning from the existing risk corpus and to facilitate the elaboration of future risk documentation. In the remainder of this section are presented the possibilities of future applications with ResRiskOnto.

To organize and extract meaning of the existing risk documentation, possible applications are:

1. visualization of structured risk sentences;
2. the development of a SQA tool for reservoir risks;
3. evaluation of similarity between risk sentences (considering the ontology classes).

All these applications depend on a well-trained NER model, one able to adequately extract entities from risk documentation. To improve the existing NER model we recommend the application of data augmentation techniques, having as seed the annotated corpus and the ontology classes. The available NER model, trained with a dataset of 552 sentences, could be replaced by one trained from thousands of sentences. Artificial Intelligence models trained with large amounts of data are known to perform better in general, and the current score of 0.81, could be used as baseline for comparison.

In possession of a well-trained NER model, entities extracted from the existing risk corpus can be used in the proposed applications. In the case of a SQA tool, a relation extraction (RE) model should also be trained to complement the named entity recognition model. The NER-RE combination can be used to structure the risk corpus in a RDF format, one that can be queried by a SQA system.

Visualization applications demand in-depth conversations with specialists. Whether by highlighting entities or grouping similar sentences, it is important to know what type of information would a domain specialist search for in a standardized risk documentation.

To facilitate the elaboration of future risk documentation, ResRiskOnto's risk conceptualisation can be used to elaborate a "quality score" or "completeness score" for risk sentences in natural language. Having a score based on a clear standard such as ResRiskOnto can enable reservoir professionals to check and guarantee the quality of future risk documentation.

Together with domain specialists, the risk corpus could then be evaluated, generating a dataset for future natural language processing tasks. Sentence quality can be verified via real-time applications, in which reservoir specialists receive insights on how to elaborate adequate risk sentences during the activity of risk assessment.

One idea of such an application is a system that suggests sentences (or that autocompletes sentences while the user is writing). Another possibility is the real-time calculation of sentence quality, in similar fashion to some *password strength meter* available in popular websites.

Overall, we expect that the proposed risk standardization can improve reservoir-risk documentation in future oil & gas projects.

8
Bibliography

ALBUQUERQUE, F. C. et al. A methodology for traffic-related twitter messages interpretation. **Computers in Industry**, Elsevier, v. 78, p. 57–69, 2016. Cited in page 37.

ALOM, M. Z. et al. A state-of-the-art survey on deep learning theory and architectures. **Electronics**, Multidisciplinary Digital Publishing Institute, v. 8, n. 3, p. 292, 2019. Cited in page 91.

ALPAY, O. A. A practical approach to defining reservoir heterogeneity. **Journal of Petroleum Technology**, OnePetro, v. 24, n. 07, p. 841–848, 1972. Cited in page 29.

APOSTOLAKSI, G. Probability and risk assessment: The subjectivistic viewpoint and some suggestions. **Nuclear Safety**, v. 19, n. 3, p. 305–315, 1978. Cited in page 23.

ARP, R.; SMITH, B.; SPEAR, A. D. **Building ontologies with basic formal ontology**. [S.l.]: Mit Press, 2015. Cited 9 times in pages 9, 34, 35, 38, 39, 41, 45, 47, and 50.

ASNAR, Y.; GIORGINI, P.; MYLOPOULOS, J. Goal-driven risk assessment in requirements engineering. **Requirements Engineering**, Springer, v. 16, n. 2, p. 101–116, 2011. Cited in page 22.

AVEN, T. On how to define, understand and describe risk. **Reliability Engineering System Safety**, v. 95, n. 6, p. 623–631, 2010. ISSN 0951-8320. Disponível em: <https://www.sciencedirect.com/science/article/pii/S095183201000027X>. Cited in page 21.

AVEN, T. Risk management. In: ______. **Risks in Technological Systems**. London: Springer London, 2010. p. 175–198. ISBN 978-1-84882-641-0. Disponível em: <https://doi.org/10.1007/978-1-84882-641-0_12>. Cited 2 times in pages 17 and 23.

AVEN, T.; KROHN, B. S. A new perspective on how to understand, assess and manage risk and the unforeseen. **Reliability Engineering & System Safety**, Elsevier, v. 121, p. 1–10, 2014. Cited in page 22.

AVEN, T.; RENN, O. On risk defined as an event where the outcome is uncertain. **Journal of risk research**, Taylor & Francis, v. 12, n. 1, p. 1–11, 2009. Cited in page 21.

AVEN, T.; RENN, O. The role of quantitative risk assessments for characterizing risk and uncertainty and delineating appropriate risk management options, with special emphasis on terrorism risk. **Risk Analysis: An International Journal**, Wiley Online Library, v. 29, n. 4, p. 587–600, 2009. Cited in page 21.

AVEN, T.; RENN, O.; ROSA, E. A. On the ontological status of the concept of risk. **Safety Science**, Elsevier, v. 49, n. 8-9, p. 1074–1079, 2011. Cited in page 21.

BAKER, R. O.; YARRANTON, H. W.; JENSEN, J. **Practical reservoir engineering and characterization**. [S.l.]: Gulf Professional Publishing, 2015. Cited 6 times in pages 9, 11, 25, 27, 30, and 32.

BENGIO, Y.; DUCHARME, R.; VINCENT, P. A neural probabilistic language model. **Advances in Neural Information Processing Systems**, v. 13, 2000. Cited in page 90.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web. **Scientific american**, JSTOR, v. 284, n. 5, p. 34–43, 2001. Cited in page 36.

BOHOLM, Å.; CORVELLEC, H. A relational theory of risk. **Journal of risk research**, Taylor & Francis, v. 14, n. 2, p. 175–190, 2011. Cited in page 21.

BORST, W. N. Construction of engineering ontologies for knowledge sharing and reuse. 1999. Cited in page 34.

BRICKLEY, D.; GUHA, R. V.; LAYMAN, A. Resource description framework (rdf) schema specification. Technical report, W3C, 1999. W3C Proposed Recommendation. http://www. w3 . . . , 1999. Cited in page 36.

CEUSTERS, W. An information artifact ontology perspective on data collections and associated representational artifacts. In: **MIE**. [S.l.: s.n.], 2012. p. 68–72. Cited in page 44.

CEUSTERS, W.; SMITH, B. Aboutness: Towards foundations for the information artifact ontology. 2015. Cited 2 times in pages 44 and 79.

CHELMIS, C. et al. Toward an automatic metadata management framework for smart oil fields. **SPE Economics & Management**, OnePetro, v. 5, n. 01, p. 33–43, 2013. Cited in page 13.

CHEONG, H.; BUTSCHER, A. Physics-based simulation ontology: an ontology to support modelling and reuse of data for physics-based simulation. **Journal of Engineering Design**, Taylor & Francis, v. 30, n. 10-12, p. 655–687, 2019. Cited 2 times in pages 76 and 77.

CICCONETO, F. Georeservoir: an ontology for deep-marine depositional system description. 2021. Cited in page 43.

CORCHO, O. et al. Building legal ontologies with methontology and webode. In: **Law and the semantic web**. [S.l.]: Springer, 2005. p. 142–157. Cited in page 45.

CORDEIRO, F. C. Petrolês-como construir um corpus especializado em óleo e gás em português. **PUC-Rio, Rio de Janeiro, RJ-Brasil: PUC-Rio**, 2020. Cited in page 91.

COUNCIL, N. R. et al. Science and judgment in risk assessment. National Academies Press, 1994. Cited in page 23.

DACONTA, M. C.; OBRST, L. J.; SMITH, K. T. **The Semantic Web: a guide to the future of XML, Web services, and knowledge management**. [S.l.]: John Wiley & Sons, 2003. Cited 2 times in pages 9 and 35.

DIONNE, G. Risk management: History, definition, and critique. **Risk management and insurance review**, Wiley Online Library, v. 16, n. 2, p. 147–166, 2013. Cited in page 17.

FISCHHOFF, B.; WATSON, S. R.; HOPE, C. Defining risk. **Policy sciences**, Springer, v. 17, n. 2, p. 123–139, 1984. Cited in page 21.

FURTADO, P. H. T. **[en] AUTOMATIC INTERPRETATION OF EQUIPMENT OPERATION REPORTS**. Tese (Doutorado) — Pontifical Catholic University of Rio de Janeiro, 2017. Cited in page 37.

GANGEMI, A. et al. Sweetening ontologies with dolce. In: SPRINGER. **International Conference on Knowledge Engineering and Knowledge Management**. [S.l.], 2002. p. 166–181. Cited in page 35.

GARCIA, L. F. et al. The geocore ontology: A core ontology for general use in geology. **Computers & Geosciences**, Elsevier, v. 135, p. 104387, 2020. Cited 4 times in pages 9, 35, 41, and 59.

GARCIA, L. F. et al. What geologists talk about: Towards a frequency-based ontological analysis of petroleum domain terms. In: **ONTOBRAS**. [S.l.: s.n.], 2020. p. 190–203. Cited 4 times in pages 51, 58, 74, and 76.

GENESERETH, M. R.; NILSSON, N. J. **Logical foundations of artificial intelligence**. [S.l.]: Morgan Kaufmann, 2012. Cited in page 34.

GOMES, D. d. S. M. et al. Portuguese word embeddings for the oil and gas industry: Development and evaluation. **Computers in Industry**, Elsevier, v. 124, p. 103347, 2021. Cited in page 90.

GROSMAN, J. S. et al. Eras: Improving the quality control in the annotation process for natural language processing tasks. **Information Systems**, Elsevier, v. 93, p. 101553, 2020. Cited in page 83.

GRUBER, T. R. A translation approach to portable ontology specifications. **Knowledge acquisition**, Elsevier, v. 5, n. 2, p. 199–220, 1993. Cited 2 times in pages 34 and 36.

GUARINO, N. **Formal ontology in information systems: Proceedings of the first international conference (FOIS'98), June 6-8, Trento, Italy**. [S.l.]: IOS press, 1998. v. 46. Cited 2 times in pages 9 and 36.

GUARINO, N.; OBERLE, D.; STAAB, S. What is an ontology? In: **Handbook on ontologies**. [S.l.]: Springer, 2009. p. 1–17. Cited 2 times in pages 34 and 35.

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. Cited in page 46.

GUARINO, N.; WELTY, C. Identity and subsumption. In: **The Semantics of Relationships**. [S.l.]: Springer, 2002. p. 111–126. Cited in page 46.

GUIZZARDI, G. **Ontological foundations for structural conceptual models**. Tese (Doutorado) — University of Twente, out. 2005. Cited in page 35.

GUIZZARDI, G. Ontological foundations for structural conceptual models. 2005. Cited in page 42.

HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. **Neural computation**, MIT Press, v. 9, n. 8, p. 1735–1780, 1997. Cited in page 87.

HÖFFNER, K. et al. Survey on challenges of question answering in the semantic web. **Semantic Web**, IOS Press, v. 8, n. 6, p. 895–920, 2017. Cited in page 37.

HONNIBAL, M.; MONTANI, I. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. To appear. 2017. Cited in page 88.

ISO Central Secretary. **Industrial automation systems and integration — Integration of life-cycle data for process plants including oil and gas production facilities — Part 2: Data model**. Geneva, CH, 2003. Cited in page 38.

ISO Central Secretary. **Risk Management - Guidelines**. Geneva, CH, 2009. Cited in page 21.

ISO Central Secretary. **Risk Management - Guidelines**. Geneva, CH, 2018. Disponível em: <https://www.iso.org/obp/ui/#iso:std:iso:31000:ed-2:v1:en>. Cited 3 times in pages 9, 17, and 18.

ITTOO, A.; BOSCH, A. van den et al. Text analytics in industry: Challenges, desiderata and trends. **Computers in Industry**, Elsevier, v. 78, p. 96–107, 2016. Cited 2 times in pages 13 and 37.

KAPLAN, S.; GARRICK, B. J. On the quantitative definition of risk. **Risk Analysis**, v. 1, n. 1, p. 11–27, 1981. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1539-6924.1981.tb01350.x>. Cited in page 22.

KHAZRAEE, E.; LIN, X. Demystifying ontology. In: **UDC Seminar 2011**. [S.l.: s.n.], 2011. p. 41–54. Cited in page 35.

LAMPLE, G. et al. Neural architectures for named entity recognition. **arXiv preprint arXiv:1603.01360**, 2016. Cited in page 90.

LOOSEMORE, M. et al. **Risk management in projects**. [S.l.]: Routledge, 2012. Cited in page 22.

LUND, M. S.; SOLHAUG, B.; STØLEN, K. **Model-driven risk analysis: the CORAS approach**. [S.l.]: Springer Science & Business Media, 2010. Cited in page 22.

MA, Y. Z.; POINTE, P. R. L. et al. **Uncertainty Analysis and Reservoir Modeling: Developing and Managing Assets in an Uncertain World, AAPG Memoir 96**. [S.l.]: AAPG, 2011. v. 96. Cited 3 times in pages 31, 32, and 33.

MORAD, S. et al. The impact of diagenesis on the heterogeneity of sandstone reservoirs: A review of the role of depositional fades and sequence stratigraphy. **AAPG Bulletin**, v. 94, p. 1267–1309, 08 2010. Cited 2 times in pages 9 and 29.

MORAES, G. **Gestão de Riscos - Volume 2 - 1a Edição**. Gerenciamento Verde Editora, 2013. ISBN 9788599331378. Disponível em: <https://books.google.com.br/books?id=Rheo3dIGZBoC>. Cited in page 17.

MOTTA, O. M. et al. Megaprojects front-end planning: The case of brazilian organizations of engineering and construction. **American Journal of Industrial and Business Management**, Scientific Research Publishing, v. 4, n. 08, p. 401, 2014. Cited in page 14.

MUSEN, M. A. The protégé project: a look back and a look forward. **AI matters**, ACM New York, NY, USA, v. 1, n. 4, p. 4–12, 2015. Cited in page 79.

NILES, I.; PEASE, A. Towards a standard upper ontology. In: **Proceedings of the international conference on Formal Ontology in Information Systems-Volume 2001**. [S.l.: s.n.], 2001. p. 2–9. Cited in page 35.

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 and . . . , 2001. Cited 3 times in pages 34, 36, and 45.

OBERLE, D. Semantic management of middleware. In: **Proceedings of the 1st international doctoral symposium on Middleware**. [S.l.: s.n.], 2004. p. 299–303. Cited in page 35.

PATT, A. G.; SCHRAG, D. P. Using specific language to describe risk and probability. **Climatic Change**, p. 17–30, 2003. Cited in page 20.

PEDREGOSA, F. et al. Scikit-learn: Machine learning in python. **the Journal of machine Learning research**, JMLR. org, v. 12, p. 2825–2830, 2011. Cited in page 87.

PETROBRAS. **Glossário Petrobras**. 2007. <https://www.agenciapetrobras.com.br/Glossario>. Accessed: 2021-07-27. Cited 2 times in pages 25 and 51.

PETROBRAS, A. 2022. Disponível em: <https://www.agenciapetrobras.com.br/Multimedia/ListarConteudoExclusivo?p_tipo=1>. Cited 3 times in pages 9, 25, and 28.

RENN, O. Three decades of risk research: accomplishments and new challenges. **Journal of Risk Research**, Routledge, v. 1, n. 1, p. 49–71, 1998. Disponível em: <https://doi.org/10.1080/136698798377321>. Cited 2 times in pages 18 and 22.

ROSA, E. A. Metatheoretical foundations for post-normal risk. **Journal of risk research**, Taylor & Francis, v. 1, n. 1, p. 15–44, 1998. Cited in page 21.

SALES, T. P. et al. The common ontology of value and risk. In: SPRINGER. **International conference on conceptual modeling**. [S.l.], 2018. p. 121–135. Cited 7 times in pages 9, 22, 36, 42, 52, 54, and 58.

SANT'ANNA, D. T. et al. Generating knowledge graphs from unstructured texts: Experiences in the e-commerce field for question answering. In: **ASLD@ ISWC**. [S.l.: s.n.], 2020. p. 56–71. Cited in page 37.

Front-End-Loading (FEL) Process Supporting Optimum Field Development Decision Making, All Days de **SPE Kuwait Oil and Gas Show and Conference**, (SPE Kuwait Oil and Gas Show and Conference, All Days). SPE-167655-MS. Disponível em: <https://doi.org/10.2118/167655-MS>. Cited in page 13.

SCHLUMBERGER. **Glossário Schlumberger**. 2019. <https://glossary.oilfield.slb.com/en/>. Accessed: 2021-07-27. Cited 2 times in pages 25 and 51.

SIENA, A.; MORANDINI, M.; SUSI, A. Modelling risks in open source software component selection. In: SPRINGER. **International Conference on Conceptual Modeling**. [S.l.], 2014. p. 335–348. Cited in page 22.

SILVA, P. F. et al. How do specialists express risks: an applied ontology for the oil & gas domain. 2021. Cited 2 times in pages 46 and 81.

SOUSA, A. G. d. **An approach to answering natural language questions in Portuguese from ontologies and knowledge bases**. Tese (Doutorado) — PUC-Rio, 2019. Cited in page 37.

STUDER, R.; BENJAMINS, V. R.; FENSEL, D. Knowledge engineering: Principles and methods. **Data & knowledge engineering**, Elsevier, v. 25, n. 1-2, p. 161–197, 1998. Cited in page 34.

SUÁREZ-FIGUEROA, M. C. et al. Introduction: Ontology engineering in a networked world. In: **Ontology engineering in a networked world**. [S.l.]: Springer, 2012. p. 1–6. Cited 2 times in pages 45 and 49.

TESSAROLLO, A.; RADEMAKER, A. Inclusion of lithological terms (rocks and minerals) in the open wordnet for english. In: **Proceedings of the LREC 2020 Workshop on Multimodal Wordnets (MMW2020)**. [S.l.: s.n.], 2020. p. 33–38. Cited 2 times in pages 25 and 51.

VISHNYAKOV, V. et al. **Primer on enhanced oil recovery**. [S.l.]: Gulf Professional Publishing, 2019. Cited 2 times in pages 9 and 29.

WEBER, K. How heterogeneity affects oil recovery. **Reservoir characterisation**, Academic Press, p. 487–544, 1986. Cited in page 29.

WHEATON, R. **Fundamentals of applied reservoir engineering: appraisal, economics and optimization**. [S.l.]: Gulf Professional Publishing, 2016. Cited in page 25.

ZITHA, P. et al. Increasing hydrocarbon recovery factors. **Society of Petroleum Engineers**, p. 1–9, 2011. Cited 2 times in pages 9 and 26.

A
Results for the natural language processing tasks

In this section we present the results for the NLP tasks conducted during the validation process of ResRiskOnto.

A.1
NER Tasks with CRF model

Table A.1 contains the result of the Named Entity Recognition task with Conditional Random Forest, in the experiment CRF-A described in Section 6.4.1.

	precision	recall	f1-score	support
outsider	0.842	0.938	0.888	1195
b__reveal__inadequacy	0.455	0.417	0.435	12
i__reveal__inadequacy	0.474	0.643	0.545	14
b__underproductivity	0.621	0.600	0.610	30
i__underproductivity	0.722	0.716	0.719	116
b__underinjectivity	1.000	0.167	0.286	12
b__precipitation	1.000	1.000	1.000	3
b__uncertain__event	0.000	0.000	0.000	2
i__uncertain__event	0.000	0.000	0.000	1
b__well__collapse	0.857	0.857	0.857	7
i__well__collapse	0.750	1.000	0.857	12
b__reactivation	0.857	1.000	0.923	6
b__gor__increase	0.000	0.000	0.000	2
b__behave__unexpectedly	0.667	0.419	0.514	43
i__behave__unexpectedly	0.615	0.548	0.580	73
b__wor__increase	1.000	0.571	0.727	14
i__wor__increase	1.000	0.346	0.514	26
b__exudation	1.000	1.000	1.000	1
b__oil__ratio__decrease	0.333	1.000	0.500	1
i__oil__ratio__decrease	0.750	1.000	0.857	3
b__overpressuring	1.000	1.000	1.000	2
b__depletion	0.875	0.778	0.824	9
b__chanelling	0.857	0.545	0.667	11
b__compartmentation	0.778	0.875	0.824	8
i__compartmentation	0.000	0.000	0.000	0

Table A.1: Results for NER task with CRF-A experiment.

Table A.1 continuation

	precision	**recall**	**f1-score**	**support**
b__recovery__reduction	0.625	0.625	0.625	8
i__recovery__reduction	0.743	0.839	0.788	31
b__decline__intensification	0.000	0.000	0.000	0
i__depletion	1.000	0.500	0.667	6
i__gor__increase	0.000	0.000	0.000	4
b__leakage	0.000	0.000	0.000	1
i__leakage	0.000	0.000	0.000	2
b__pore__collapse	1.000	0.500	0.667	2
b__subsidence	1.000	1.000	1.000	1
i__pore__collapse	0.000	0.000	0.000	1
i__underinjectivity	0.444	0.118	0.186	34
b__hydraulic__fracturing	0.000	0.000	0.000	1
i__hydraulic__fracturing	0.000	0.000	0.000	0
b__well__damaging	0.000	0.000	0.000	1
b__source__event	0.000	0.000	0.000	2
i__source__event	0.000	0.000	0.000	1
b__fracturing	0.000	0.000	0.000	0
i__fracturing	0.000	0.000	0.000	0
b__thermal__fracturing	0.000	0.000	0.000	1
i__thermal__fracturing	1.000	0.333	0.500	3
i__well__damaging	0.000	0.000	0.000	2
i__overpressuring	1.000	0.500	0.667	4
i__chanelling	0.000	0.000	0.000	7
i__decline__intensification	0.000	0.000	0.000	0
b__well__aggregate	0.667	0.667	0.667	3
i__well__aggregate	0.667	0.800	0.727	5
b__well	0.804	0.882	0.841	51
b__facies	0.667	0.667	0.667	3
b__water	0.963	0.963	0.963	27
b__earth__fluid	0.750	1.000	0.857	3
b__rock	1.000	0.625	0.769	8
b__fault	1.000	1.000	1.000	12
b__geological__object	0.750	0.429	0.545	7
i__rock	1.000	1.000	1.000	1
b__geological__contact	0.000	0.000	0.000	0
b__reservoir	0.906	0.967	0.935	30
b__gas	1.000	1.000	1.000	5
b__hydrocarbon	1.000	1.000	1.000	3
b__fracture	1.000	1.000	1.000	3
i__well	0.500	1.000	0.667	9
b__depositional__unit	0.714	0.833	0.769	6
b__equipment	1.000	0.800	0.889	5
b__top__of__reservoir	1.000	1.000	1.000	3
b__oil	1.000	0.778	0.875	9
i__geological__object	0.333	0.500	0.400	2
b__geological__object__position	0.000	0.000	0.000	0

Table A.1 continuation

	precision	**recall**	**f1-score**	**support**
i__facies	0.000	0.000	0.000	0
b__facies__association	0.000	0.000	0.000	0
i__facies__association	0.000	0.000	0.000	0
b__earth__material	1.000	0.667	0.800	3
i__depositional__unit	1.000	1.000	1.000	3
b__well__location	0.000	0.000	0.000	5
b__geological__boundary	0.000	0.000	0.000	2
b__karst	0.000	0.000	0.000	0
i__karst	0.000	0.000	0.000	0
b__well__pattern	0.000	0.000	0.000	1
i__well__location	0.000	0.000	0.000	8
i__gas	0.000	0.000	0.000	0
i__top__of__reservoir	0.000	0.000	0.000	2
i__equipment	0.000	0.000	0.000	3
i__geological__boundary	0.000	0.000	0.000	0
b__aquifer	1.000	1.000	1.000	2
i__well__pattern	0.000	0.000	0.000	0
b__unconsolidated__earth__material	0.000	0.000	0.000	0
i__earth__material	0.000	0.000	0.000	0
i__geological__object__position	0.000	0.000	0.000	0
b__sand	0.000	0.000	0.000	0
b__quality	0.000	0.000	0.000	1
i__quality	0.000	0.000	0.000	1
b__volume	1.000	0.875	0.933	8
b__productivity__index	1.000	0.444	0.615	9
b__in__place__volume	1.000	1.000	1.000	8
b__permoporosity	0.857	1.000	0.923	6
b__dimension	0.667	0.333	0.444	6
b__concentration	0.000	0.000	0.000	2
i__in__place__volume	0.944	1.000	0.971	17
b__fluid__saturation	1.000	1.000	1.000	1
b__injectivity__index	0.500	1.000	0.667	2
b__absolute__permeability	1.000	1.000	1.000	2
b__viscosity	0.000	0.000	0.000	0
b__geometry	0.000	0.000	0.000	2
i__geometry	0.000	0.000	0.000	0
i__permoporosity	1.000	1.000	1.000	2
b__fluid__contact	0.000	0.000	0.000	2
b__porosity	0.000	0.000	0.000	0
b__flow__rate	1.000	0.500	0.667	2
b__pressure	0.250	1.000	0.400	1
b__static__pressure	0.000	0.000	0.000	1
i__static__pressure	0.000	0.000	0.000	1
i__pressure	0.000	0.000	0.000	0
b__clay__content	0.000	0.000	0.000	0
i__productivity__index	0.000	0.000	0.000	2

Table A.1 continuation

	precision	**recall**	**f1-score**	**support**
b__temperature	0.000	0.000	0.000	0
b__facies__quality	0.000	0.000	0.000	2
i__facies__quality	0.000	0.000	0.000	2
i__injectivity__index	0.000	0.000	0.000	0
b__relative__permeability	1.000	1.000	1.000	1
i__relative__permeability	1.000	1.000	1.000	1
b__relational__quality	0.000	0.000	0.000	0
b__injected__fluid	0.500	0.500	0.500	2
b__injected__water	0.636	0.778	0.700	9
b__produced__gas	0.000	0.000	0.000	2
b__produced__water	0.500	1.000	0.667	4
b__production__well	1.000	1.000	1.000	11
b__produced__fluid	0.000	0.000	0.000	1
b__injection__well	1.000	1.000	1.000	7
b__injection__pressure	0.667	1.000	0.800	2
i__injection__pressure	0.667	1.000	0.800	4
b__caprock	1.000	1.000	1.000	1
b__produced__oil	0.000	0.000	0.000	7
b__injected__steam	0.000	0.000	0.000	0
b__injected__gas	0.000	0.000	0.000	0
b__inadequacy	0.000	0.000	0.000	5
i__inadequacy	0.000	0.000	0.000	5
b__unpredictability	0.692	0.750	0.720	12
b__rupture__potential	0.000	0.000	0.000	0
i__rupture__potential	0.000	0.000	0.000	0
b__permeability	0.500	0.500	0.500	2
i__permeability	1.000	0.500	0.667	2
i__unpredictability	0.000	0.000	0.000	5
b__seal	0.000	0.000	0.000	4
b__collapse__potential	0.000	0.000	0.000	0
i__seal	0.000	0.000	0.000	0
b__reservoir__model	1.000	0.923	0.960	13
i__reservoir__model	0.933	0.933	0.933	15
b__data__item	0.889	1.000	0.941	8
b__representation	0.857	1.000	0.923	18
i__representation	0.947	1.000	0.973	18
b__production__profile	0.500	0.500	0.500	2
i__production__profile	0.500	1.000	0.667	2
b__p50__reservoir__scenario	0.800	0.800	0.800	5
i__p50__reservoir__scenario	1.000	0.500	0.667	4
i__data__item	1.000	1.000	1.000	7
b__p10__reservoir__scenario	1.000	1.000	1.000	1
b__p90__reservoir__scenario	1.000	1.000	1.000	1
micro avg	0.809	0.809	0.809	2200
macro avg	0.461	0.445	0.437	2200
weighted avg	0.783	0.809	0.785	2200

	precision	recall	f1-score	support
outsider	0.874	0.92	0.897	1337
i__underinjectivity	0.6	0.214	0.316	14
i__chanelling	0	0	0	0
b__behave__unexpectedly	0.478	0.458	0.468	48
b__exudation	0	0	0	0
i__behave__unexpectedly	0.627	0.684	0.654	76
i__well__collapse	1	0.5	0.667	4
b__leakage	0	0	0	1
b__well__collapse	1	0.667	0.8	3
i__uncertain__event	1	0.75	0.857	4
b__chanelling	1	0.857	0.923	7
i__well__damaging	0	0	0	0
b__underinjectivity	0.5	0.2	0.286	5
b__well__damaging	0	0	0	0
i__pore__collapse	0	0	0	0
b__underproductivity	0.783	0.857	0.818	21
b__oil__ratio__decrease	0	0	0	1
i__reveal__inadequacy	0.429	0.36	0.391	25
b__fracturing	1	1	1	2
b__thermal__fracturing	0	0	0	1
i__underproductivity	0.84	0.851	0.846	74
i__overpressuring	1	1	1	3
i__source__event	0	0	0	1
b__compartmentation	0.8	0.8	0.8	10
i__hydraulic__fracturing	0	0	0	0

Table A.2: Results for NER task with Bi-LSTM-CRF (150 *epochs*.)

A.2
NER Tasks with Bi-LSTM-CRF model

Table A.2 contains the result of the Named Entity Recognition task with a Bidirectional Long-Short Term Memory algorithm combined with a Conditional Random Forest Layer, with *epochs*, as described in Section 6.4.2.

Table A.2 continuation	**precision**	**recall**	**f1-score**	**support**
b__hydraulic__fracturing	1	1	1	1
i__wor__increase	0.833	1	0.909	5
i__depletion	1	0.3	0.462	10
i__gor__increase	0.5	0.25	0.333	8
b__gor__increase	0.5	1	0.667	1
b__overpressuring	0.667	1	0.8	2
i__fracturing	0	0	0	4
i__compartmentation	0	0	0	2
b__depletion	1	0.5	0.667	4
i__decline__intensification	0	0	0	0
b__recovery__reduction	0.846	0.846	0.846	13
b__reveal__inadequacy	0.476	0.556	0.513	18
b__wor__increase	0.75	1	0.857	3
b__source__event	0.5	0.5	0.5	2
i__thermal__fracturing	0.5	1	0.667	1
b__pore__collapse	0	0	0	0
b__decline__intensification	0	0	0	0
b__precipitation	1	1	1	4
b__reactivation	1	1	1	1
i__leakage	0	0	0	0
i__oil__ratio__decrease	0	0	0	1
i__recovery__reduction	0.875	0.913	0.894	46
b__uncertain__event	1	0.667	0.8	3
b__subsidence	0	0	0	0
i__geological__boundary	0	0	0	0
b__gas	0.889	1	0.941	8
i__seal	0	0	0	0
i__geological__object__position	0	0	0	0
i__facies	0	0	0	4
b__geological__object	0	0	0	6
b__fault	1	1	1	10
i__earth__material	0	0	0	0
i__well	0.462	0.5	0.48	12
i__equipment	0	0	0	0
b__geological__contact	0	0	0	0
i__geological__object	0	0	0	3
i__gas	0	0	0	0
i__karst	0	0	0	0
i__well__aggregate	1	0.714	0.833	7
b__earth__material	0.5	0.5	0.5	2
b__well__location	0	0	0	2
b__facies	0.7	0.778	0.737	9
b__top__of__reservoir	1	1	1	2
b__oil	0.727	0.889	0.8	9
b__sand	0	0	0	0
b__aquifer	1	1	1	3

Table A.2 continuation	precision	recall	f1-score	support
b__well	0.763	0.714	0.738	63
b__facies__association	1	0.333	0.5	3
b__geological__boundary	0	0	0	0
b__fracture	0.6	0.75	0.667	4
b__geological__object__position	0	0	0	2
b__water	1	0.9	0.947	20
b__equipment	0	0	0	4
b__fluid__contact	1	1	1	2
i__rock	0	0	0	1
b__earth__fluid	0.75	0.6	0.667	5
b__karst	1	1	1	1
b__hydrocarbon	1	0.4	0.571	5
b__rock	0.75	1	0.857	6
b__depositional__unit	0.875	0.875	0.875	8
i__depositional__unit	1	0.25	0.4	4
b__well__aggregate	1	0.8	0.889	5
b__unconsolidated__earth__material	0	0	0	0
i__top__of__reservoir	0	0	0	1
i__facies__association	0	0	0	2
i__well__location	0	0	0	3
i__well__pattern	0	0	0	3
b__well__pattern	0	0	0	2
b__reservoir	0.8	1	0.889	28
b__in__place__volume	0.875	1	0.933	7
b__absolute__permeability	0	0	0	0
i__productivity__index	0	0	0	0
b__permoporosity	1	0.917	0.957	12
i__permeability	0.667	1	0.8	2
b__volume	0.667	1	0.8	6
b__concentration	1	0.333	0.5	3
b__temperature	0	0	0	0
i__geometry	0	0	0	0
i__static__pressure	0	0	0	1
i__in__place__volume	1	0.75	0.857	20
b__relational__quality	0	0	0	0
b__static__pressure	0	0	0	1
i__injectivity__index	0	0	0	0
b__dimension	0.857	0.857	0.857	7
b__fluid__saturation	1	1	1	3
i__relative__permeability	0	0	0	0
i__quality	0	0	0	0
i__injection__pressure	0.75	1	0.857	6
i__permoporosity	1	0.333	0.5	3
b__geometry	0	0	0	0
b__clay__content	0	0	0	1
b__viscosity	0	0	0	0

Table A.2 continuation	precision	recall	f1-score	support
b__quality	0	0	0	0
i__pressure	0	0	0	2
i__facies__quality	0	0	0	2
b__productivity__index	1	0.429	0.6	7
b__flow__rate	0.8	0.8	0.8	5
b__injectivity__index	1	0.5	0.667	4
b__relative__permeability	0	0	0	0
b__porosity	0	0	0	0
b__pressure	0.667	0.8	0.727	5
b__facies__quality	0	0	0	2
b__injection__well	1	1	1	8
b__produced__fluid	0	0	0	1
b__injection__pressure	0.75	1	0.857	3
b__injected__fluid	0.5	0.5	0.5	4
b__injected__gas	1	0.5	0.667	2
b__produced__oil	1	1	1	2
b__caprock	1	1	1	1
b__injected__steam	0	0	0	0
b__produced__gas	1	1	1	1
b__production__well	1	1	1	6
b__injected__water	0.2	0.333	0.25	3
b__produced__water	1	1	1	3
b__seal	0	0	0	3
b__permeability	1	1	1	2
i__inadequacy	0.333	0.5	0.4	2
b__rupture__potential	0	0	0	0
i__rupture__potential	0	0	0	0
b__collapse__potential	0	0	0	1
b__unpredictability	0.6	0.75	0.667	8
b__inadequacy	1	0.333	0.5	6
i__unpredictability	0	0	0	4
b__reservoir__model	0.824	0.933	0.875	15
i__data__item	0	0	0	3
b__p10__reservoir__scenario	0	0	0	0
b__representation	0.917	0.917	0.917	12
i__representation	0.917	1	0.957	11
b__data__item	0.4	0.2	0.267	10
b__production__profile	0.8	0.5	0.615	8
i__p50__reservoir__scenario	0.9	1	0.947	9
i__production__profile	0.667	0.545	0.6	11
b__p90__reservoir__scenario	1	1	1	1
i__reservoir__model	0.818	0.6	0.692	15
b__p50__reservoir__scenario	1	0.917	0.957	12
micro avg	0.829	0.822	0.825	2281
macro avg	0.469	0.431	0.434	2281
weighted avg	0.811	0.822	0.81	2281